

Министерство образования и науки Российской Федерации

Федеральное государственное бюджетное образовательное учреждение  
высшего профессионального образования  
«Оренбургский государственный университет»

# **МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ: ИССЛЕДОВАНИЕ СОЦИАЛЬНЫХ, ЭКОНОМИЧЕСКИХ И ЭКОЛОГИЧЕСКИХ ПРОЦЕССОВ (РЕГИОНАЛЬНЫЙ АСПЕКТ)**

Рекомендовано Ученым советом федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Оренбургский государственный университет» в качестве учебного пособия для студентов, обучающихся по программам высшего профессионального образования по экономическим направлениям подготовки

Оренбург  
2012

УДК 330.4 (075.8)

ББК 65в631я73

М 34

Рецензент - доктор экономических наук, профессор Е.М. Дусаева

Авторы: О.И. Бантикова, В.И. Васянина, Ю.А. Жемчужникова, А.Г. Реннер, Е.Н. Седова, О.И. Стебунова, Л.М. Туктамышева, О.С. Чудинова

М34 Математическое моделирование: исследование социальных, экономических и экологических процессов (региональный аспект): учебное пособие/ О.И. Бантикова, В.И. Васянина, Ю.А. Жемчужникова, А.Г. Реннер, Е.Н. Седова, О.И. Стебунова, Л.М. Туктамышева, О.С. Чудинова /под ред. А.Г. Реннера; Оренбургский гос. ун-т. – Оренбург: ОГУ, 2012. – 366с. – 1 электрон. опт. диск (CD-ROM).  
ISBN

В учебном пособии, предназначенном для студентов старших курсов, проводящих самостоятельные исследования, предложены подходы к анализу таких острых социально-экономических проблем как безработица, занятость, оценка эколого-экономических рисков, демографическая и экономическая безопасности, ценообразование на рынке жилья, инвестиционная привлекательность, оценка последствий от вступления России в ВТО, требующих целенаправленного использования разнообразных математических методов.

Разделы учебного пособия сформированы по принципу однородности используемого математического инструментария: непараметрических методов статистики, методов многомерного статистического анализа, моделей бинарного и множественного выбора, моделей регрессии для панельных данных, методов анализа многомерных временных рядов. В заключении каждого раздела представлены вопросы и задания для самостоятельной работы, а также тематика для проведения исследовательских работ. Приложения к учебному пособию, а также исходная информационная база для исследовательских работ содержатся на электронном носителе.

Учебное пособие предназначено для студентов различных специальностей и направлений, изучающих дисциплины, связанные с математическим моделированием социальных, экономических и экологических процессов, а также для аспирантов, интересующихся данной проблемой.

УДК 330.4 (075.8)

ББК 65в631я73

ISBN

© Реннер А.Г., 2012

© ОГУ, 2012

## Содержание

Введение.....	8
1 Непараметрические методы статистики в анализе социально-экономических процессов.....	10
1.1 Обзор непараметрических методов проверки однородности распределения совокупностей.....	10
1.2 Методы анализа данных типа времени жизни.....	18
1.3 Исследование влияния социально-демографических и профессионально-квалификационных характеристик безработного на продолжительность поиска работы.....	23
Вопросы и задания по теоретическому материалу раздела 1.....	50
Задание к исследовательской работе.....	54
Список использованных источников к разделу 1.....	55
2 Методы многомерного статистического анализа в социально-экономических исследованиях.....	57
2.1 Обзор методов многомерной классификации.....	58
2.2 Методы снижения размерности признакового пространства.....	64
2.3 Методы построения интегральных показателей.....	68
2.4 Территориальное зонирование города Оренбурга на основе многомерного статистического анализа факторов, определяющих цену жилья на вторичном рынке.....	75
2.5 Многомерная классификация муниципальных образований Оренбургской области по состоянию показателей, характеризующих экономическую и демографическую безопасность как основных составляющих национальной безопасности.....	84
2.6 Классификация муниципальных образований Оренбургской области по показателям, характеризующим инвестиционную привлекательность региона....	94

2.7 Многомерная классификация административно-территориальных образований по показателям, характеризующим антропогенную нагрузку и эколого-экономический риск.....	101
2.8 Моделирование интегральных показателей, характеризующих социально-экономические процессы региона и сравнительный анализ муниципальных образований Оренбургской области по синтетическим категориям.....	116
2.8.1 Сравнительный анализ муниципальных образований Оренбургской области по уровню демографической безопасности региона.....	116
2.8.2 Сравнительный анализ муниципальных образований Оренбургской области по уровню экономической безопасности региона.....	119
2.8.3 Сравнительный анализ муниципальных образований Оренбургской области по уровню инвестиционной привлекательности региона.....	125
Вопросы и задания по теоретическому материалу раздела 2 .....	141
Задание к исследовательской работе .....	144
Список использованных источников к разделу 2.....	144
3 Моделирование социально-экономических явлений и процессов со специфическими зависимыми переменными .....	149
3.1 Модели бинарного выбора.....	150
3.2 Модели множественного выбора .....	159
3.3 Системы одновременных моделей бинарного выбора (двумерные пробит-модели).....	164
3.4 Ранжирование категорий безработных по степени их востребованности на рынке труда на основе моделей бинарного выбора .....	171
3.5 Построение интегрального показателя, характеризующего экономическую безопасность региона, на основе моделей бинарного выбора.....	174
3.6 Моделирование и прогнозирование эколого-экономического риска на основе моделей бинарного и множественного выбора .....	179
3.7 Построение интегральных показателей антропогенной нагрузки и эколого-экономического риска на основе моделей множественного выбора .....	191

3.8 Моделирование взаимного влияния эколого-экономического риска и показателей социально-экономического развития муниципалитетов на основе двумерных пробит-моделей.....	205
Вопросы и задания по теоретическому материалу раздела 3 .....	218
Задания к исследовательской работе .....	221
Список использованных источников к разделу 3.....	223
4 Моделирование социально-экономических процессов с учетом пространственно-временной неоднородности данных .....	225
4.1 Модели линейной регрессии для панельных данных .....	226
4.1.1 Модель линейной регрессии с фиксированными эффектами .....	228
4.1.2 Модель линейной регрессии со случайными эффектами.....	233
4.2 Моделирование числа трудоустроенных безработных с учетом пространственно-временной неоднородности данных .....	236
4.3 Модели оценки стоимости квартир на вторичном рынке жилья с учетом пространственно-временной неоднородности .....	244
4.4 Построение интегрального индикатора, характеризующего уровень демографической безопасности с учетом неоднородности данных во времени ..	253
4.5 Моделирование зависимостей между показателями, характеризующими инвестиционный потенциал, инвестиционную активность и инвестиционный риск с учетом пространственной и временной неоднородности данных .....	261
4.6 Модели эколого-экономического риска с учетом пространственной и временной неоднородности данных.....	267
Вопросы и задания по теоретическому материалу раздела 4 .....	271
Задания к самостоятельной и исследовательской работе .....	273
Список использованных источников к разделу 4.....	275
5 Моделирование взаимосвязей социально-экономических показателей на основе многомерных временных рядов .....	278
5.1 Особенности изучения взаимосвязей показателей, измеренных во времени.	
Методы нетрадиционного корреляционного анализа факторов .....	279

5.2 Моделирование и прогнозирование показателей на основе многомерных временных рядов .....	284
5.3 Метод Зельнера при оценивании коэффициентов группы регрессионных уравнений .....	289
5.4 Анализ взаимосвязей между показателями, характеризующими социально-экономические процессы, измеренными во времени.....	292
5.4.1 Отбор факторов, оказывающих влияние на показатели объемов производства сельского хозяйства и промышленности региона .....	292
5.4.2 Отбор факторов, оказывающих влияние на демографические и миграционные процессы в регионе.....	297
5.4.3 Анализ взаимосвязей между показателями, характеризующими инвестиционную активность, инвестиционный потенциал и инвестиционный риск .....	302
5.5 Моделирование и прогнозирование динамики структуры показателей, характеризующих социально-экономические процессы.....	306
5.5.1 Моделирование динамики структуры общего прироста населения .....	306
5.5.2 Моделирование динамики структуры внешней миграции по основным отраслям экономики.....	312
5.5.3 Моделирование динамики отраслевой структуры инвестиций .....	320
5.6 Моделирование и прогнозирование социально-экономических процессов и систем на основе систем регрессионных уравнений.....	325
5.6.1 Построение модели объемов производства промышленности и сельского хозяйства в виде группы регрессионных уравнений.....	325
5.6.2 Сценарное прогнозирование объемов производства промышленности и сельского хозяйства на основе группы регрессионных уравнений.....	329
5.6.3 Моделирование показателей рынка труда на основе систем одновременных регрессионных уравнений.....	333
5.6.4 Моделирование и прогнозирование основных индикаторов демографической безопасности на основе системы одновременных уравнений.....	348
Вопросы и задания по теоретическому материалу раздела 5 .....	357

Задание к исследовательской работе .....	359
Список использованных источников к разделу 5.....	362

## Введение

Цель учебного пособия состоит в выработке навыков проведения исследований по актуальным проблемам экономики, в социальной сфере, в области эколого-экономических рисков. В учебном пособии представлены математический инструментарий и результаты моделирования на основе продвинутого программного обеспечения таких острых социально-экономических проблем, как безработица, демографическая и экономическая безопасность, эколого-экономическое состояние, и, безусловно, актуальные для экономики регионов проблемы вступления России в ВТО, инвестиционной привлекательности, ценообразования на рынке жилья, а также вопросы внутренней и внешней миграции населения.

Для формирования обоснованной экономической и социальной политики региона необходимо располагать количественными оценками показателей характеризующих уровень безработицы, инвестиционную привлекательность, эколого-экономические риски, состояние рынка труда, производство, демографическую безопасность, внешнюю миграцию.

Решение перечисленных выше задач требует использования разнообразного математического аппарата. При построении математических моделей необходимо учитывать различия в типах данных, неоднородность данных во времени, в пространстве и, в связи с этим, особенности анализа соответствующих моделей. Так, анализ процессов, характеризующихся данными типа «время жизни», цензурированными выборками требует использования непараметрических статистических методов. При исследовании процессов, характеризующихся качественными признаками, возникает потребность в моделях бинарного или множественного выбора. Неоднородность объектов по набору характеризующих их свойств, необходимость построения латентных показателей требуют привлечения аппарата многомерных статических методов: классификации, компонентного анализа, шкалирования. Неустойчивость во времени па-



раметров моделей, построенных по пространственным выборкам, требует использования и развития моделей на панельных данных. Выявление характера взаимосвязей между измеренными во времени показателями требует применения методов нетрадиционного корреляционного анализа. Наконец, многообразие связей показателей, характеризующих сложные процессы, влечет использование аппарата моделей в форме систем регрессионных уравнений, в том числе, двумерных пробит-моделей.

Структура учебного пособия, состоящего из пяти глав, объединенных общностью используемого математического инструментария такова. В первых параграфах каждой главы дается описание основного математического аппарата. Степень детализации приводимого математического инструментария зависит от его новизны для слушателей знакомых, как правило, с курсами «теория вероятностей и математическая статистика», «анализ данных», «эконометрика», «теория случайных процессов». В последующих параграфах главы излагаются результаты исследования по обозначенной выше тематике, опирающиеся на изложенный инструментарий. В конце каждой главы приводятся вопросы как по математическому инструментарю и его использованию, так и по исследуемой тематике и результатам моделирования. В заключении приводятся задание и тематика возможных исследовательских работ. Приложения к учебному пособию содержат результаты расчетов, полученные с помощью прикладного программного обеспечения: Statistica, Stata, Eviews, SPSS.

Учебное пособие предназначено для студентов старших курсов, изучающих дисциплины «Моделирование социальных процессов», «Эконометрическое моделирование», «Моделирование демографических и миграционных процессов», «Моделирование эколого-экономических систем», проводящих исследования в рамках научно-исследовательской работы студентов, производственной практики и выполняющих выпускную квалификационную работу.

# **1 Непараметрические методы статистики в анализе социально-экономических процессов**

Раздел состоит из трех подразделов. Подраздел 1.1 посвящен обзору непараметрических методов проверки однородности распределения двух и более совокупностей. Основное внимание в нем уделено решению задачи проверки однородности распределения двух и более независимых совокупностей при отсутствии цензурированных наблюдений. Анализу данных типа времени жизни посвящен подраздел 1.2. Здесь рассмотрены непараметрический метод оценивания функции надежности (метод множительной оценки Каплана–Мейера), основные критерии проверки однородности распределения нескольких совокупностей, специальная регрессионная модель (модель пропорциональных интенсивностей), учитывающая особенности данных типа времени жизни. Третий подраздел посвящен исследованию влияния социально-демографических и профессионально-квалификационных характеристик безработного на продолжительность поиска работы с помощью непараметрических методов. В конце раздела приведены вопросы и задания по теоретическому материалу, задание к исследовательской работе и список использованных источников.

## **1.1 Обзор непараметрических методов проверки однородности распределения совокупностей**

В прикладных исследованиях часто возникает необходимость выяснить, различаются ли генеральные совокупности, из которых взяты выборки. Например, отличается ли потребительское поведение мужчин и женщин; влияет ли способ упаковки деталей на их потребительские свойства через год после хранения; эффективна ли применяемая методика лечения и т.д. Решение таких задач можно свести к проверке гипотез об однородности распределения сово-

купностей. При этом выбор критерия определяется, в первую очередь, тем, зависимы или нет рассматриваемые случайные величины. Кроме того, необходимо принимать во внимание, выполняется ли требование к закону распределения рассматриваемых совокупностей. Многие методы прикладного статистического анализа основаны, например, на предположении нормального закона распределения случайных величин. На практике это предположение не всегда выполняется, вследствие чего возникает необходимость обращаться к непараметрическим методам статистики. Непараметрическими называют методы, при использовании которых нет необходимости предполагать, что функция распределения результатов наблюдений принадлежит какому-либо определенному параметрическому семейству.

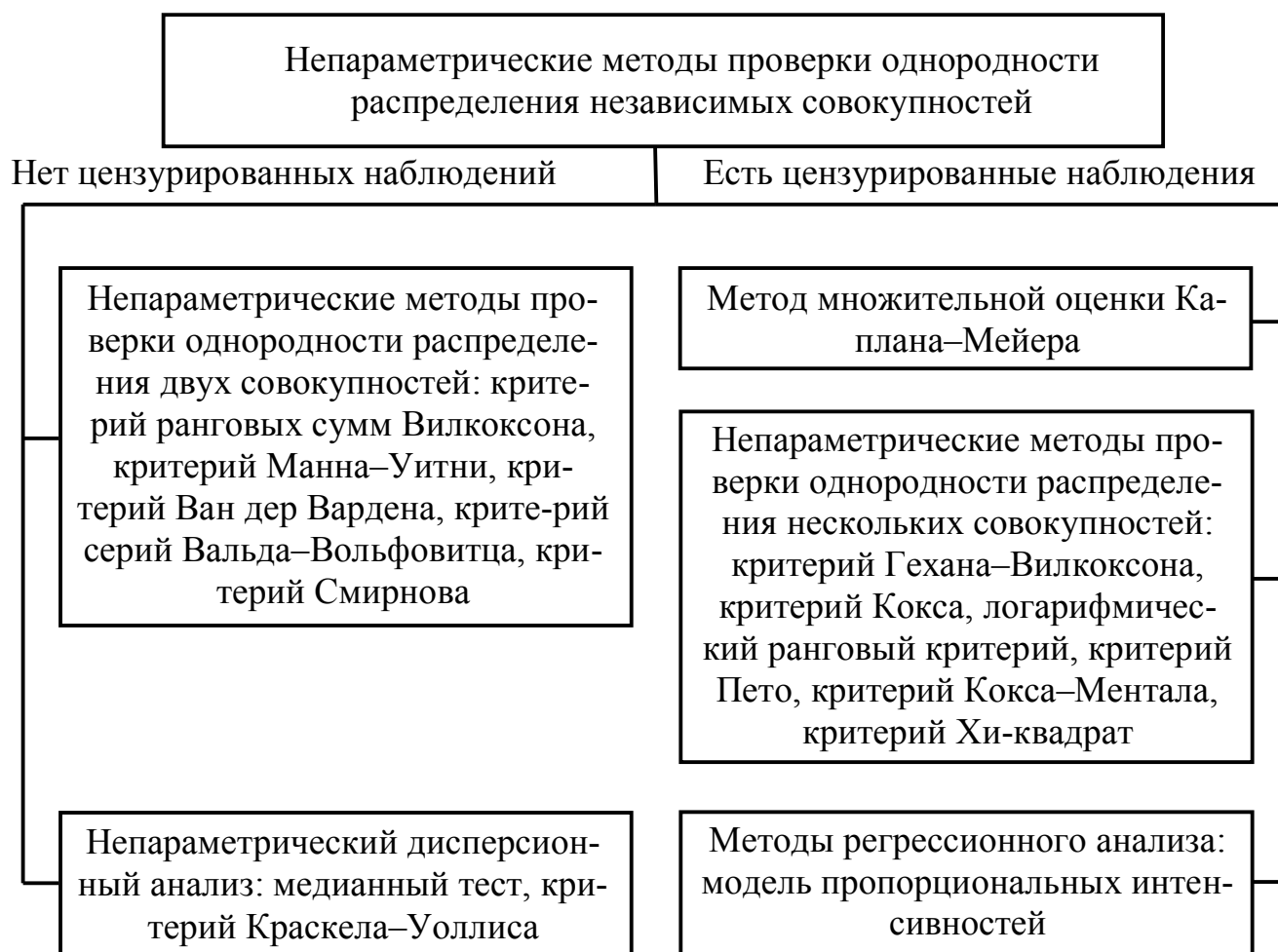


Рисунок 1.1 – Классификация непараметрических методов проверки однородности распределения независимых совокупностей

На рисунке 1.1 представлена классификация непараметрических методов проверки однородности распределения двух и более независимых совокупностей. В зависимости от того, есть в выборочной совокупности цензурированные наблюдения или нет, выделены две группы непараметрических методов. Первая группа методов применяется, когда выборочная совокупность состоит из полных данных. Вторая группа методов используется, когда среди исходных данных встречаются цензурированные. Использование второй группы методов обусловлено тем, что исходная информация может относиться к данным, развивающимся во времени. В различных областях научного знания класс задач, связанных с обработкой таких данных, называют по-разному: «анализ данных типа времени жизни», «анализ продолжительности», «анализ истории явлений», «анализ надежности» и другие варианты. В настоящее время общепринятым становится термин «анализ транзитных данных», введенный Т. Ланкастером. Такое название объясняется тем, что «транзитные данные» относятся не только к продолжительности события, но и к тому, что произойдет, когда оно закончится [1, 2].

Ставится задача на основе выборочных данных  $x_1, x_2, \dots, x_{n_1}$  и  $y_1, y_2, \dots, y_{n_2}$ , извлеченных из двух независимых совокупностей  $X$  и  $Y$ , проверить однородность распределения случайных величин  $X$  и  $Y$ .

Две совокупности  $X$  и  $Y$  однородны, если справедлива гипотеза:

$$H_0 : F_X(x) = F_Y(x). \quad (1.1)$$

Если гипотеза (1.1) верна, то выборки можно объединить и считать извлеченными из одной генеральной совокупности.

В некоторых случаях целесообразно проверять не совпадение функций распределения, а совпадение некоторых характеристик случайных величин  $X$  и  $Y$  – математических ожиданий, медиан, дисперсий и др. Однородность математических ожиданий означает, что справедлива гипотеза:

$$H_0 : M(X) = M(Y), \quad (1.2)$$

где  $M(X)$  – математическое ожидание случайной величины  $X$ ;

$M(Y)$  – математическое ожидание случайной величины  $Y$ .

Если гипотеза (1.1) верна, то верна и гипотеза (1.2), но из справедливости гипотезы (1.2) не следует справедливость гипотезы (1.1). Если же гипотеза (1.2) отвергается, то отвергается и гипотеза (1.1).

Для проверки гипотезы (1.2) используется  $t$ -критерий Стьюдента:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{(n_1 - 1)S_x^2 + (n_2 - 1)S_y^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}, \quad (1.3)$$

где  $\bar{x}$  и  $\bar{y}$  – выборочные оценки математических ожиданий случайных величин  $X$  и  $Y$ ;

$S_x^2$  и  $S_y^2$  – выборочные несмещенные оценки дисперсий случайных величин  $X$  и  $Y$ .

Использование критерия (1.3) возможно при соблюдении следующих двух условий:

1) нормальный закон распределения случайных величин  $X$  и  $Y$ :  
 $X \in N(a_x, \sigma_x)$ ,  $Y \in N(a_y, \sigma_y)$ ;

2) равенство дисперсий случайных величин  $X$  и  $Y$ :  
 $D(X) = \sigma_x^2 = D(Y) = \sigma_y^2$ .

Статистика (1.3) при справедливости нулевой гипотезы имеет распределение Стьюдента с числом степеней свободы  $\nu = n_1 + n_2 - 2$ .

Если оба условия выполнены, то проверка гипотезы (1.1) сводится к проверке гипотезы (1.2).

Если хотя бы одно из условий не выполнено, то нет оснований считать, что  $t$ -статистика (1.3), используемая для проверки гипотезы (1.2), при ее справедливости имеет распределение Стьюдента. В такой ситуации использование  $t$ -критерия возможно только при больших и примерно равных объемах выборок [3].

Априорное предположение о принадлежности функций распределения  $F_X(x)$  и  $F_Y(x)$  к какому-либо определенному параметрическому семейству обычно нельзя достаточно надежно обосновать. Поэтому для исследования следует использовать методы, пригодные при любом виде  $F_X(x)$  и  $F_Y(x)$ , т.е. непараметрические методы.

Для каждого параметрического критерия имеется, по крайней мере, одна непараметрическая альтернатива. Для проверки гипотезы (1.1) разработано много непараметрических методов – критерий ранговых сумм Вилкоксона, Манна–Уитни  $U$ -тест, критерий серий Вальда–Вольфовица, критерий Смирнова и другие [4]-[7]. При выборе критерия проверки однородности необходимо учитывать его мощность. Критерий ранговых сумм Вилкоксона, Манна-Уитни  $U$ -тест и ряд других ориентированы на проверку нулевой гипотезы (1.1) при альтернативной гипотезе сдвига:

$$H_1 : F_X(x) = F_Y(x + r), \quad r \neq 0, \quad (1.4)$$

то есть эти критерии являются непараметрическими альтернативами  $t$ -критерия Стьюдента.

Критерий Вилкоксона является одним из самых распространенных непараметрических ранговых критериев для проверки однородности двух независимых совокупностей. Несмотря на несостоятельность критерия в случае общей альтернативы и неестественность альтернативы сдвига в большинстве эконометрических исследований, критерий Вилкоксона прост в применении и часто позволяет обнаружить различие двух совокупностей. Следует отметить, что статистика Вилкоксона и статистика Манна-Уитна линейно связаны, поэтому на практике можно использовать любой из этих критериев [3, 4].

Для проверки гипотезы однородности (1.1) в случае альтернативы общего вида используется, например, критерий Смирнова [3]. Он является состоятельным при альтернативной гипотезе вида:

$$H_1 : F_X(x) \neq F_Y(x). \quad (1.5)$$

Критерий Вальда-Вольфовица проверяет гипотезу о том, что две независимые выборки извлечены из двух совокупностей, которые чем-то существенно различаются между собой, иными словами, различаются не только средними, но также формой распределения.

При больших объемах выборок для проверки гипотезы об однородности распределений можно пользоваться нормальными аппроксимациями статистик критериев [4, 8].

$$\text{Так, статистика Вальда-Вольфовица } Z = \frac{\left| N - \left( \frac{2n_1n_2}{n_1 + n_2} + 1 \right) \right| - \frac{1}{2}}{\sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}},$$

где  $N$  – число серий, при справедливости нулевой гипотезы и  $n_1 > 20$  и/или  $n_2 > 20$  имеет (приблизительно) стандартное нормальное распределение.

$$\text{Статистика Манна-Уитни } Z = \frac{W - \frac{1}{2}n_1n_2}{\sqrt{\frac{1}{12}n_1n_2(n_1 + n_2 + 1)}},$$

$$\text{где } W = \min \left\{ n_1n_2 + \frac{n_1(n_1 + 1)}{2} - R_1; n_1n_2 + \frac{n_2(n_2 + 1)}{2} - R_2 \right\};$$

$R_1$  – сумма рангов первой выборки;

$R_2$  – сумма рангов второй выборки, при справедливости нулевой гипотезы и  $n_1 > 8$  и  $n_2 > 8$  имеет (приблизительно) стандартное нормальное распределение.

Критерий Колмогорова–Смирнова представляет собой максимальную разницу между эмпирическими функциями распределения, построенными по выборкам  $D = \max |F_X^*(x) - F_Y^*(y)|$ . Если  $n_1 > 40$  и  $n_2 > 40$ , то при  $\alpha = 0,05$

$$D_{кр} = 1,36 \cdot \sqrt{\frac{n_1n_2}{n_1 + n_2}}.$$

Для проверки однородности распределения более двух независимых совокупностей используется дисперсионный анализ. Непараметрическими альтернативы однофакторного дисперсионного анализа являются критерий ANOVA Краскела-Уоллиса и медианный тест [4, 8].

Критерий Краскела-Уоллиса проверяет однородность распределения  $k$  случайных величин при альтернативной гипотезе сдвига. Критерий Краскела-

Уоллиса  $H = \frac{12}{n(n+1)} \left( \sum_{i=1}^k \frac{R_i^2}{n_i} \right) - 3(n+1)$ , где  $n = \sum_{i=1}^k n_i$ ,  $R_i$  – сумма рангов  $i$ -ой

выборки,  $i = \overline{1, \dots, k}$ , при справедливости нулевой гипотезы и  $n_i \geq 5$  и  $k \geq 4$  имеет приблизительно распределение «Хи-квадрат» с числом степеней свободы  $\nu = k - 1$ .

Медианный тест обладает меньшей мощностью и основан на подсчете числа наблюдений каждой выборки, которые попадают выше или ниже общей медианы выборок, и вычисляет затем значение статистики «Хи-квадрат» для таблицы сопряженности  $2 \times k$ , где  $k$  – число рассматриваемых случайных величин.

Непараметрические методы проверки однородности распределения двух зависимых совокупностей применяются, когда исходные данные представляют собой повторные парные наблюдения или попарно связанные выборки, т.е. когда для каждого объекта  $O_i$  ( $i = \overline{1, n}$ ) имеется два значения  $x_i$  и  $y_i$ , говорят «до обработки» и «после обработки». Например, регистрируются два значения – производительность станка до капитального ремонта и производительность станка после капитального ремонта. Ставится задача выяснить, есть ли эффект «обработки», т.е. изменилась ли производительность станков после ремонта. Решение такой задачи также сводится к проверке гипотезы об однородности распределения совокупностей  $X$  и  $Y$ . Для этого могут быть использованы, например, критерий знаковых рангов Вилкоксона и критерий знаков Фишера [4]. Оба критерия ориентированы на проверку нулевой гипотезы при альтернативной гипотезе сдвига.



Рассмотрим нормальные аппроксимации критериев. Критерий знаковых

рангов Вилкоксона  $Z = \frac{T^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$ , где  $T^+$  – это сумма положительных

знаковых рангов разностей  $y_i - x_i$ , при справедливости нулевой гипотезы и объеме выборки  $n > 25$  имеет (приблизительно) стандартное нормальное распределение [4, 8].

Критерий знаков Фишера  $Z = \frac{B - \frac{n}{2}}{\sqrt{\frac{n}{4}}}$ , где  $B$  – число положительных раз-

ностей  $y_i - x_i$ , при справедливости нулевой гипотезы и объеме выборки  $n > 100$  имеет (приблизительно) стандартное нормальное распределение.

Следует отметить, что если с помощью рассмотренных критериев гипотеза об однородности двух зависимых совокупностей  $X$  и  $Y$  отвергается, т.е. совокупности отличаются, по крайней мере, сдвигом  $\theta$ , то можно рассчитать непараметрические точечные и интервальные оценки  $\theta$ . Кроме того, рассмотренные критерий знаковых рангов Вилкоксона и критерий знаков Фишера могут использоваться и при проверке гипотез о значении медианы случайной величины, построения ее точечной и интервальной оценок [4].

Для проверки однородности распределения  $k > 2$  зависимых совокупностей следует использовать непараметрические альтернативы двухфакторного дисперсионного анализа, например, критерий Фридмана:

$$F = \frac{12}{kn(n+1)} \sum_{i=1}^n \left( \sum_{j=1}^k R_{ij} \right)^2 - 3k(n+1), \text{ где } R_{ij} \text{ – ранг } i\text{-го объекта по } j\text{-му признаку.}$$

Критерий Фридмана при справедливости нулевой гипотезы аппроксимируется распределением «Хи-квадрат» с числом степеней свободы  $n - 1$ .

Существуют также непараметрические альтернативы проверки гипотез о равенстве дисперсий совокупностей: критерий Ансари-Бредли, критерий Клотца, критерий Сиджела–Тьюки и другие [3, 4].

## 1.2 Методы анализа данных типа времени жизни

В анализе данных типа времени жизни особый интерес представляют группа или группы объектов, для каждого из которых определено точечное событие, называемое отказом [9]. Отказ происходит после некоторого интервала времени (наработки до отказа) для каждого объекта только один раз. Для точного определения наработки до отказа необходимо установить начало отсчета времени, выбрать масштаб для измерения времени и определить понятие отказа.

Для описания распределения случайной величины  $T$  – продолжительность события до отказа могут использоваться следующие функции:

- функция распределения  $F_T(t) = P(T < t)$ ;
- функция надежности (выживания)  $S_T(t) = P(T \geq t) = 1 - F_T(t)$ ;
- функция плотности вероятности  $f_T(t)$ , удовлетворяющая следующим

условиям:  $F_T(t) = \int_{-\infty}^t f_T(\tau) d\tau$ ;  $f_T(t) \geq 0$ ;  $\int_{-\infty}^{+\infty} f_T(\tau) d\tau = 1$ ;

- функция риска (функция интенсивности отказа)

$$\lambda_T(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f_T(t)}{S_T(t)}.$$

Сравнение распределений независимых случайных величин предлагается свести к сравнению их функций надежности. Непараметрическим методом оценивания функции надежности является метод множительной оценки Каплана–Мейера, не требующий предварительной группировки данных в интервальный вариационный ряд, учитывающий наличие цензурирования и одинаковых по длительностям наблюдений в выборке [9, 10]. Оценка функции надежности методом множительной оценки Каплана–Мейера имеет вид:

$$\hat{S}_T(t_j) = \prod_{i=1}^j \frac{(n_i - h_i)}{n_i} = \prod_{i=1}^j (1 - \hat{\lambda}(t_i)), \quad (1.6)$$

где  $t_1 < t_2 < \dots < t_k$  – расположенные в порядке возрастания значения продолжительности поиска работы, соответствующие нецензурированным наблюдениям;

$h_j$  – количество наблюдений, для которых продолжительность события составила  $t_j$ ;

$$n_j = \sum_{i \geq j}^k (h_i + m_i) - \text{количество наблюдений, для которых продолжительность события составила } t_j \text{ и более;}$$

$m_j$  – количество наблюдений, цензурированных в момент между  $t_j$  и

$t_{j+1}$ ;

$\hat{\lambda}(t_j)$  – оценка функции риска.

Для сравнения функций надежности, описывающих распределение рассматриваемых случайных величин, могут использоваться следующие непараметрические критерии, учитывающие наличие в выборочных данных цензурированных наблюдений: критерий Гехана–Вилкоксона, критерий Кокса, логарифмический ранговый критерий, критерий Пето, критерий Кокса–Ментала, критерий Хи-квадрат [2, 8]. При большом объеме выборочных данных для проверки однородности распределения совокупностей достаточно использовать критерий Хи-квадрат или его нормальную аппроксимацию.

Более сложные сравнения распределений независимых случайных величин можно выполнить, используя расширенные модели, в которых влияние объясняющих переменных определяется значениями неизвестных параметров [9, 10]. Решение этой задачи возможно путем построения специальных регрессионных моделей, учитывающих особенности данных типа времени жизни. Для описания влияния объясняющих переменных на наработку до отказа могут использоваться следующие модели:

- модель пропорциональных интенсивностей;
- модель ускоренных испытаний;

- немультимпликативные модели, основанные на функции интенсивности;
- модель с перенесенным началом отсчета времени;
- модель ускорения наступления отказов.

Наиболее популярной и простой в интерпретации является модель пропорциональных интенсивностей. Она задается через функцию интенсивностей и для постоянного во времени вектора объясняющих переменных  $x = (x_1, x_2, \dots, x_k)^T$  имеет вид:

$$\lambda(t, x, \lambda_0) = \varphi(x)\lambda_0(t), \quad (1.7)$$

где  $\lambda_0(t)$  – функция интенсивности при стандартных условиях  $x = 0$  (при отсутствии влияния объясняющих переменных).

Множитель  $\varphi(x)$  не зависит от наработки до отказа  $t$  и оказывает мультипликативный эффект на результирующий риск прекращения события  $\lambda(t, x, \lambda_0)$ . Полагают, что  $\varphi(0) = 1$ .

Таким образом, основная идея модели пропорциональных интенсивностей заключается в том, что влияние на длительность события соответствует умножению функции риска на постоянный множитель, который определяется значениями экзогенных факторов [9, 10]. Предположение пропорциональности рисков означает, что для двух объектов с разными значениями объясняющих переменных отношение их функций интенсивностей не зависит от времени.

Функция  $\varphi(x)$  может быть параметризована и зависеть от вектора параметров  $\beta = (\beta_1, \beta_2, \dots, \beta_k)^T$ , т.е.  $\varphi(x) = \varphi(x, \beta)$ . Так как функция  $\varphi(x, \beta)$  должна удовлетворять условиям  $\varphi(x, \beta) \geq 0$  и  $\varphi(0, \beta) = 1$ , то естественно предположить в качестве  $\varphi(x, \beta)$  функцию вида:

$$\varphi(x, \beta) = e^{\beta^T x}. \quad (1.8)$$

Возможны и другие варианты спецификации функции  $\varphi(x, \beta)$ .

Для модели (1.7) с учетом функции  $\varphi(x)$  вида (1.8) имеется четкая интерпретация параметров  $\beta_i$ ,  $i = \overline{1, \dots, k}$ , поскольку из равенства:  $\partial \ln \lambda(t, x, \beta, \lambda_0) / \partial x_i = \partial \ln \varphi(x, \beta) / \partial x_i = \beta_i$ ,

следует, что коэффициент  $\beta_i$  можно интерпретировать как постоянный пропорциональный эффект объясняющей переменной  $x_i$  на условную вероятность завершения события [10].

Оценка вектора  $\beta$  может быть получена методом максимального правдоподобия, так называемым частично вероятностным подходом Кокса и сведением модели к линейной регрессии и дальнейшей ее оценкой методом наименьших квадратов. В случае, когда о функции интенсивностей  $\lambda_0(t)$  ничего не известно, оценку вектора  $\beta$  можно найти с помощью подхода, предложенного Коксом и не требующего спецификации формы функции базового риска [9, 10].

Рассмотрим сначала случай, когда цензурирование отсутствует. Пусть  $\tau_1 < \tau_2 < \dots < \tau_n$  упорядоченные моменты наработок до отказов у  $n$  объектов. Обозначим через  $T_j$  номер объекта, отказавшего в момент  $t_j$ , при этом  $T_j = i$  тогда и только тогда, когда  $t_i = \tau_j$ . Множество риска  $R(\tau_j)$  составляют объекты, действующие непосредственно перед  $j$ -м упорядоченным отказом:  $R(\tau_j) = \{i : t_i \geq \tau_j\}$ . При заданной полной истории до  $j$ -й упорядоченной наработки до отказа  $H_j = \{\tau_1, \tau_2, \dots, \tau_j, i_1, i_2, \dots, i_{j-1}\}$  условная вероятность того, что  $i$ -й объект отказал в момент  $\tau_j$  в том случае, если один из объектов из множества риска  $R(\tau_j)$  отказал в момент  $\tau_j$  определяется следующим образом:

$$\frac{\lambda_i(\tau_j, x^{(i)}, \beta, \lambda_0)}{\sum_{k \in R(\tau_j)} \lambda_k(\tau_j, x^{(k)}, \beta, \lambda_0)} = \frac{\varphi(x^{(i)}, \beta)}{\sum_{k \in R(\tau_j)} \varphi(x^{(k)}, \beta)} \quad (1.9)$$

и не зависит от базовой функции интенсивности  $\lambda_0(t)$  в силу предположения мультипликативности (1.7).

Хотя соотношение (1.9) выводится как условная вероятность события  $T_j = i$  при заданной полной истории  $H_j$ , в действительности она как функция не зависит от  $\tau_1, \tau_2, \dots, \tau_j$ . Следовательно, вероятность  $p_j(i/i_1, i_2, \dots, i_j)$  равна условному распределению  $T_j$  только, если  $T_1 = i_1, T_2 = i_2, \dots, T_{j-1} = i_{j-1}$ . Тогда совместное распределение  $p(i_1, i_2, \dots, i_n)$  может быть получено по обычному цепному правилу для условных вероятностей:

$$p(i_1, i_2, \dots, i_n) = \prod_{j=1}^n p_j(i_j / i_1, i_2, \dots, i_{j-1}) = \prod_{j=1}^n \frac{\varphi(x^{(i_j)}, \beta)}{\sum_{k \in R(\tau_j)} \varphi(x^{(k)}, \beta)}. \quad (1.10)$$

Рассмотрим случай, когда выборочные данные содержат цензурированные наблюдения. Пусть в выборке объемом  $n$  имеются  $d$  полных наблюдений, а остальные являются цензурированными. Обозначим  $\tau_1 < \tau_2 < \dots < \tau_d$  упорядоченные моменты наработок до отказов. Как и раньше,  $T_j = i$ , если объект  $i$  отказал в момент  $\tau_j$ . Множество риска  $R(\tau_j) = \{i : t_i \geq \tau_j\}$  состоит из полных наблюдений, продолжительность которых больше либо равна  $\tau_j$ , а также из наблюдений, цензурированных на интервале  $(\tau_j, \tau_{j+1})$ . Выражение (1.10) в случае цензурирования называют функцией правдоподобия вида:

$$l(\beta) = \prod_{j=1}^d \frac{\varphi(x^{(i_j)}, \beta)}{\sum_{k \in R(\tau_j)} \varphi(x^{(k)}, \beta)}. \quad (1.11)$$

Оценка вектора параметров  $\beta$  находится с помощью метода максимального правдоподобия с использованием логарифмической функции правдоподобия вида:

$$L(\beta) = \sum_{j=1}^d \left\{ \ln \varphi(x^{(i_j)}, \beta) - \ln \left[ \sum_{k \in R(\tau_j)} \varphi(x^{(k)}, \beta) \right] \right\}. \quad (1.12)$$

Следует отметить, что модель с непрерывным временем (1.7) и функция правдоподобия (1.11) не применимы в случае наличия в выборочных данных совпадающих значений наработок до отказа. Для преодоления этой трудности используется приближение, впервые предложенное Р. Пето [9]. Если некоторые значения наработок до отказа регистрируются как равные одному и тому же числу  $\tau$ , то каждому из них соответствует свой член вида (1.9) с одинаковым знаменателем, в который входят все объекты из множества риска  $R(\tau)$ . Это приближение используется во многих статистических пакетах и дает вполне адекватные результаты.

### **1.3 Исследование влияния социально-демографических и профессионально-квалификационных характеристик безработного на продолжительность поиска работы**

Статистика безработицы обеспечивает предоставление крайне необходимой обществу информации, используемой государственными органами управления при разработке социально-экономической политики. Несмотря на то, что основным источником информации, всесторонне характеризующим конъюнктуру рынка труда, являются результаты выборочных обследований населения по проблемам занятости, текущая статистика органов федеральной государственной службы занятости населения по-прежнему сохраняет свое значение для анализа рынка труда. На основе информации служб занятости определяется официальный уровень безработицы и разрабатываются программы содействия занятости населения. Кроме того, службы занятости населения являются основным источником информации для подробного изучения одной из важней-

ших характеристик безработицы, называемой индикатором социальной напряженности в обществе, – ее продолжительности. При анализе продолжительности безработицы не следует ограничиваться только построением таблиц распределения безработных по интервалам времени поиска работы, даже если такие таблицы построены по отдельным категориям безработных. Видимые различия в продолжительности поиска работы безработными, различающимися по полу, возрасту, образованию и другим характеристикам, необходимо доказать.

Исследование влияния качественных характеристик безработного на продолжительность поиска работы предлагается проводить в три этапа.

Этап 1. Анализ продолжительности поиска работы каждой категорией безработных.

Этап 2. Сравнение продолжительности поиска работы различными категориями безработных.

Этап 3. Оценка влияния факторов на продолжительность поиска работы.

На первом этапе проводится исследование закона распределения продолжительности поиска работы каждой категорией безработных, рассчитываются основные числовые характеристики распределения.

Второй этап исследования предлагается свести к проверке гипотез об однородности распределения случайных величин, каждая из которых представляет собой продолжительность поиска работы отдельной категорией безработных. Однородность распределения случайных величин будет означать, что качественный признак, лежащий в основе группировки безработных, не оказывает значимого влияния на продолжительность поиска работы. Поскольку вид закона распределения продолжительности поиска работы каждой категорией безработных заранее не известен, а априорное предположение о принадлежности функции распределения к какому-либо определенному параметрическому семейству обычно нельзя достаточно надежно обосновать, то для исследования необходимо использовать непараметрические методы статистики.

Поскольку качественные признаки могут взаимодействовать друг с другом, то с помощью непараметрических критериев можно проверять гипотезы



об однородности распределения совокупностей, сформированных в результате многофакторных группировок безработных.

После выявления факторов, оказывающих существенное влияние на продолжительность поиска работы, необходимо охарактеризовать их влияние, то есть определить каким категориям безработных легче, а каким труднее найти работу. Для этого можно использовать оценки основных числовых характеристик распределения рассматриваемых совокупностей.

Возможность использования методов анализа данных типа времени жизни к исследованию продолжительности регистрируемой безработицы обусловлена тем, что информация о продолжительности поиска работы может быть неполной, то есть когда реальная продолжительность поиска работы дольше длительности наблюдения за безработными. Такие данные возникают, если безработный в момент прекращения наблюдения еще находился на учете в Центре занятости или безработный покинул Центр занятости без трудоустройства (снят с учета по личному заявлению, в связи с переменой места жительства и др.). Для этих безработных известно лишь, что время поиска работы превышает период наблюдения, истинная же продолжительность периода безработицы неизвестна. Такие наблюдения называются цензурированными справа. Наличие цензурированных данных затрудняет оценку влияния факторов на время трудоустройства, однако исключение их из рассмотрения уменьшает объем выборки, что снижает точность оценок.

В исследовании продолжительности поиска работы началом отсчета времени является дата регистрации безработного в Центре занятости населения в целях поиска подходящей работы, продолжительность поиска работы измеряется в днях, под отказом понимается трудоустройство безработного при содействии Центра занятости населения или самостоятельно. Нарботка до отказа – это продолжительность поиска работы, т.е. количество дней, начиная с даты регистрации безработного в целях поиска подходящей работы и заканчивая датой его трудоустройства.

Для описания распределения случайной величины  $T$  – продолжительность поиска работы или время трудоустройства могут использоваться следующие функции:

– функция распределения  $F_T(t) = P(T < t)$ , равная вероятности того, что продолжительность поиска работы меньше  $t$ ;

– функция надежности (выживания)  $S_T(t) = P(T \geq t) = 1 - F_T(t)$ , равная вероятности того, что продолжительность поиска работы превысит  $t$ ;

– функция плотности вероятности  $f_T(t)$ , характеризующая вероятность трудоустройства в течение малого промежутка времени;

– функция риска (функция интенсивности отказа)

за)  $\lambda_T(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f_T(t)}{S_T(t)}$ , характеризующая «риск» трудо-

устройства в течение малого промежутка времени при предположении, что человек был безработным до начала этого периода.

Исследование влияния факторов на продолжительность поиска работы реализовано на основе данных Центра занятости населения города Оренбурга. Информационную базу для исследования составили данные о безработных гражданах, зарегистрированных в целях поиска подходящей работы в период с 1 января 2000 года по 31 декабря 2005 года. Выбор данного временного интервала обусловлен, во-первых, значительными изменениями, происходившими в законодательстве по безработице в 90-е гг., касающимися как условий регистрации граждан в качестве безработных, так и правил выплаты пособия по безработице. Во-вторых, несовершенством информационной базы данных на первых этапах автоматизации деятельности служб занятости.

Вся информация о безработном, обратившемся в Центр занятости, фиксируется в карточке персонального учета и вносится в базу данных, в которой накапливается информация по всем безработным. Эта информация содержит социально-демографические характеристики безработных: пол, возраст, уровень образования, семейный статус, число детей и др., а также профессиональ-

но-квалификационные характеристики: профессия, общий стаж работы, стаж работы за последний год, средний размер заработной платы за последний год, последнее место работы (наименование предприятия, отрасль и форма собственности, причина увольнения), желаемая профессия, желаемая заработная плата и др. В базе данных хранится информация о датах постановки на учет, снятия с учета, получения статуса безработного, возникновения права на получение пособия по безработице, приостановления и прекращения выплаты пособия в случае нарушения правил, установленных законодательством, начала и окончания переобучения и др. Для покидающих Центр занятости указывается причина снятия с учета: трудоустроился при содействии Центра занятости, трудоустроился самостоятельно, направлен на переобучение, нарушил условия регистрации, поменял место жительства и др.

Объем выборки составил 45981 человек. В рассмотрение не вошли безработные, направленные на переобучение. Для безработных, трудоустроенных при содействии Центра занятости или самостоятельно в течение рассматриваемого периода, определено значение случайной величины  $T$ , равное числу дней с момента регистрации до момента трудоустройства. Количество полных наблюдений за временем трудоустройства составило 34053 (74,06%). Для безработных, снятых с учета в течение рассматриваемого периода без трудоустройства, определено количество дней пребывания на учете, а для безработных, которые продолжали находиться на учете в Центре занятости после окончания наблюдения за ними, – количество дней с момента регистрации до 31 декабря 2005 года. Такие данные являются цензурированными, их количество в выборке составило 11928 (25,94%).

Для анализа были выбраны следующие характеристики безработного:

- пол: мужской, женский;
- возраст: до 29 лет, от 30 до 49 лет, от 50 лет и старше;
- уровень профессионального образования: высшее, среднее, начальное, нет профессионального образования;
- профессия: рабочий, служащий;

- наличие опыта работы: нет опыта работы, есть опыт работы;
- специальность (для безработных с высшим профессиональным образованием): гуманитарно-социальная, педагогическая, специальность экономики и управления, специальность сельского и рыбного хозяйства, инженерная специальность, специальность других направлений подготовки.

В соответствии с этими характеристиками сформированы качественные признаки с несколькими уровнями возможных значений, группирующими выборочные данные на однородные классы. Кроме того, сформированы показатели, позволяющие идентифицировать граждан, официально зарегистрированных в качестве безработных, и безработных, которым выплачивалось пособие по безработице. Введена переменная, указывающая на трудоустройство безработного при содействии Центра занятости или самостоятельное трудоустройство.

Исследование влияния факторов на продолжительность поиска работы проведено с помощью двух групп методов: с помощью непараметрических методов анализа полных данных и с помощью методов анализа данных типа времени жизни. Для обработки данных использовался пакет прикладных программ Statistica 6.0, реализующий необходимые процедуры анализа и обладающий широкими графическими возможностями [8].

Результаты проверки гипотез об однородности распределения случайных величин  $X$  – продолжительность поиска работы мужчиной и  $Y$  – продолжительность поиска работы женщиной с помощью критериев Смирнова и Манна–Уитни представлены на рисунках 1.2 и 1.3.

Критерий Смирнова									
Группирующий признак: пол									
Группа 1 - мужчины Группа 2 - женщины									
Перем.	Макс. отр. Разн.	Макс. пол. Разн.	p-уровень	Среднее Группа 1	Среднее Группа 2	Ст. откл. Группа 1	Ст. откл. Группа 2	Числ набл. Группа 1	Число набл. Группа 2
T	-0,197094	0,000146	p < .001	31,57496	58,73307	64,56413	93,84405	16022	18031

Рисунок 1.2 – Результаты проверки гипотезы об отсутствии влияния пола безработного на продолжительность поиска работы с помощью критерия Смирнова

ва

Критерий Манна-Уитни									
Группирующая переменная: пол									
Группа 1 - мужчины Группа 2 - женщины									
Перем.	Сум. ранг. Группа 1	Сум. ранг. Группа 2	U	Z	р-уровень	Z скорр.	р-уровень	Число набл. Группа 1	Число набл. Группа 2
T	238049933	341770498	109689680	-38,3863	0,00	-38,4447	0,00	16022	18031

Рисунок 1.3 – Результаты проверки гипотезы об отсутствии влияния пола безработного на продолжительность поиска работы с помощью критерия Манна-Уитни

Результаты проверки гипотез об однородности распределения случайных величин  $X$  – продолжительность поиска работы безработным в возрасте до 29 лет,  $Y$  – продолжительность поиска работы безработным в возрасте от 30 до 49 лет и  $Z$  – продолжительность поиска работы безработным в возрасте от 50 лет и старше с помощью медианного теста и критерия ANOVA Краскела-Уоллиса представлены на рисунках 1.4, 1.5.

Завис. переменная: T	Независимая (группирующая) переменная: возраст		
	Код	Число наблюдений	Сумма рангов
До 29 лет	1	16152	261963759
От 30 до 49 лет	2	12888	221857919
От 50 лет и старше	3	4971	94569389

Рисунок 1.4 – Результаты проверки гипотезы об отсутствии влияния возраста безработного на продолжительность поиска работы с помощью критерия Краскела-Уоллиса

Завис. переменная: T	Медианный тест, общая медиана = 8,00			
	До 29 лет	От 30 до 49 лет	От 50 лет и старше	Общий
<= Медиана: набл.	8491,00	6516,00	2206,000	17213,00
ожидаемые	8174,54	6522,63	2515,828	
набл.-ожд.	316,46	-6,63	-309,828	
> Медиана: набл.	7661,00	6372,00	2765,000	16798,00
ожидаемые	7977,46	6365,37	2455,172	
набл.-ожд.	-316,46	6,63	309,828	
Сумма: набл.	16152,00	12888,00	4971,000	34011,00

Рисунок 1.5 – Результаты проверки гипотезы об отсутствии влияния возраста безработного на продолжительность поиска работы с помощью медианного теста

На основе отчетов, полученных в пакете Statistica, сформирована таблица 1.1, содержащая результаты исследования влияния на продолжительность по-

иска работы пола, наличия опыта работы, возраста, уровня профессионального образования, профессии и специальности безработного.

Таблица 1.1 – Результаты исследования влияния факторов на продолжительность поиска работы с помощью непараметрических методов анализа полных данных

Наименование критерия	Значение статистики	Уровень значимости	Вывод
$H_0$ : пол безработного не оказывает значимое влияние на продолжительность поиска работы			
Критерий Смирнова	не выдается	$p < 0,001$	$H_0$ отвергается
Критерий Манна–Уитни	$z_{набл} = -38,39$	$p = 0,00$	$H_0$ отвергается
$H_0$ : наличие опыта не оказывает значимое влияние на продолжительность поиска работы			
Критерий Смирнова	не выдается	$p < 0,001$	$H_0$ отвергается
Критерий Манна–Уитни	$z_{набл} = 32,05$	$p = 0,00$	$H_0$ отвергается
$H_0$ : возраст безработного не оказывает значимое влияние на продолжительность поиска работы			
Критерий Краскела–Уоллиса	$H_{набл} = 320,69$	$p = 0,000$	$H_0$ отвергается
Медианный тест	$\chi^2_{набл} = 102,07$	$p = 0,000$	$H_0$ отвергается
$H_0$ : уровень профессионального образования безработного не оказывает значимое влияние на продолжительность поиска работы			
Критерий Краскела–Уоллиса	$H_{набл} = 5467,09$	$p = 0,000$	$H_0$ отвергается
Медианный тест	$\chi^2_{набл} = 4657,43$	$p = 0,000$	$H_0$ отвергается
$H_0$ : профессия безработного не оказывает значимое влияние на продолжительность поиска работы			
Критерий Смирнова	не выдается	$p < 0,001$	$H_0$ отвергается
Критерий Манна–Уитни	$z_{набл} = -68,56$	$p = 0,00$	$H_0$ отвергается
$H_0$ : специальность безработного не оказывает значимое влияние на продолжительность поиска работы			
Критерий Краскела–Уоллиса	$H_{набл} = 94,62$	$p = 0,000$	$H_0$ отвергается
Медианный тест	$\chi^2_{набл} = 57,09$	$p = 0,000$	$H_0$ отвергается

Так как нулевые гипотезы отвергаются, можно сделать вывод о том, что все рассматриваемые качественные признаки оказывают значимое влияние на продолжительность поиска работы. Для характеристики этого влияния рассчитаны основные параметры распределения продолжительности поиска работы разными категориями безработных. Результаты представлены в таблице 1.2.

Таблица 1.2 – Основные числовые характеристики распределения продолжительности поиска работы разными категориями безработных (в днях)

Категория безработных	Средняя продолжительность	Медиана	Нижняя квартиль	Верхняя квартиль	Стандартное отклонение
Мужчины	32	7	4	24	64,56
Женщины	59	13	5	79	93,84
Есть опыт работы	51	10	5	64	87,11
Нет опыта работы	18	6	4	11	37,29
Возраст до 29 лет	33	8	5	34	58,84
Возраст от 30 до 49 лет	53	8	5	64	92,28
Возраст 50 лет и старше	68	13	5	90	107,79
Высшее образование	74	37	8	99	95,92
Среднее образование	70	29	7	94	99,18
Начальное образование	43	9	5	43	79,65
Нет образования	23	6	4	11	57,67
Рабочие профессии	28	6	4	14	64,07
Служащие	76	36	7	101	99,56
Гуманитарно-социальные специальности	75	45	11	114	84,76
Педагогические специальности	66	29	7	85	94,75
Специальности экономики и управления	81	42	9	113	100,97
Специальности сельского и рыбного хозяйства	74	32	7	90	105,84
Инженерные специальности	70	31	6	92	97,09
Другие специальности	73	34	7	100	94,76

На основе сравнения основных характеристик распределения продолжительности поиска работы мужчинами и женщинами сделан вывод, что мужчины трудоустраиваются быстрее, чем женщины. Средняя продолжительность поиска работы для мужчин составила 32 дня, для женщин – 59 дней. Согласно оценкам медианы 50% мужчин трудоустраиваются в течение первой недели пребывания на учете в Центре занятости, 50% женщин трудоустраиваются в течение первых двух недель (13 дней). Оценки нижней и верхней квартилей свидетельствуют о том, что 25% мужчин и женщин трудоустраиваются в течение соответственно первых четырех и пяти дней пребывания на учете, 75% мужчин трудоустраиваются в течение 24 дней, 75% женщин – в течение 79 дней.

Безработные без опыта работы трудоустраиваются быстрее, чем безработные с опытом работы. Три четверти безработных второй группы трудоустраиваются в течение 11 дней со дня регистрации, такая же доля безработных первой группы – в течение 64 дней. Вакансии, предлагаемые безработным второй группы, рассчитаны на труд невысокой квалификации, который не может претендовать на большое вознаграждение. Люди ранее не работавшие вынуждены трудоустраиваться на низкооплачиваемую работу, чтобы быстрее заработать стаж и в дальнейшем претендовать на лучшую работу с более высоким заработком. Средняя желаемая заработная плата в первой и второй группах безработных составила соответственно 3387 рублей и 2495 рублей, среднее число отклоненных безработным вакансий – 2,6 и 1,4, среднее число отказов работодателей в трудоустройстве безработных – 1,4 и 0,43.

Сравнение средних значений продолжительности поиска работы и оценок квартилей в разных возрастных группах показало, что быстрее всех трудоустраиваются молодые люди, 75% которых покидают Центр занятости населения в течение 34 дней. Такая же доля людей среднего возраста трудоустраивается в течение 62 дней, людей старшего возраста – в течение трех месяцев (90 дней). Как правило, чем старше человек, тем опытнее. Оценка парного коэффициента корреляции между возрастом безработного и его стажем составила 0,93. Как доказано ранее, наличие опыта работы влияет на время трудоустройства. С точки зрения имеющегося опыта работы каждой возрастной группе безработных можно дать следующую характеристику:

– к первой возрастной группе относятся выпускники различных учебных заведений, не имеющие опыта работы, а также безработные с небольшим стажем работы. Безработных данной группы можно охарактеризовать как молодых, энергичных, восприимчивых к новациям людей, не претендующих на высокий заработок. Средний размер желаемой заработной платы в данной группе составил 2868 рублей, среднее число отклоненных безработным вакансий – 1,79, среднее число отказов работодателей – 1,01;



– ко второй возрастной группе в основном относятся люди с опытом работы, средний общий стаж в данной группе составил 17 лет. Средний размер желаемой заработной платы у людей среднего возраста выше, чем у молодых людей и равен 3618 рублям. Безработные данной группы более требовательны к предлагаемой работе, чем молодежь, среднее число отклоненных безработным вакансий составило 2,95, среднее число отказов работодателей – 1,48;

– к третьей возрастной группе относятся в основном люди предпенсионного возраста с большим опытом работы, средний общий стаж в данной группе составил 30 лет. Средний размер желаемой заработной платы в данной группе составил 3653 рублей, что немного выше, чем во второй группе. Среднее число отклоненных безработным вакансий составило 3,13, среднее число отказов работодателей – 1,52. Несмотря на высокую квалификацию, безработных данной группы менее охотно берут на работу из-за близости пенсионного возраста.

Исследование влияния уровня профессионального образования на продолжительность поиска работы показало, что быстрее всех трудоустраиваются безработные, не имеющие профессионального образования, затем – безработные с начальным профессиональным образованием. Специалисты с высшим профессиональным образованием дольше остальных пребывают в поиске работы. Средний размер желаемой заработной платы составил: для безработных с высшим профессиональным образованием – 3993 рубля, средним профессиональным образованием – 3171 рубль, начальным профессиональным образованием – 3088 рублей, для не имеющих профессионального образования – 2897 рублей. Для безработных с высшим образованием было характерно наибольшее по сравнению с другими категориями безработных среднее число отказов работодателей в трудоустройстве – 2,34 и наименьшее среднее число отклоненных вакансий – 2,11. Для сравнения значения первого показателя для остальных категорий безработных составили соответственно 2,19, 1,47 и 0,94; значения второго показателя – 2,51, 2,74 и 2,35.

Безработные с рабочими профессиями трудоустраиваются быстрее, чем служащие. Так 75% рабочих трудоустраиваются в течение 14 дней, безработные же второй группы трудоустраиваются значительно дольше: 50% служащих трудоустраиваются в течение 36 дней, 75% служащих – в течение 101 дня. Средняя желаемая заработная плата в каждой группе безработных составила 3100 и 3612 рублей соответственно. Среднее число отклоненных вакансий составило 2,65 для рабочих и 2,44 для служащих. Большое количество вакансий, отклоненных безработными, говорит о том, что условия труда предлагаемой работы часто не удовлетворяют желаниям безработных. Сравнивая среднее число отказов работодателей в каждой группе 0,79 и 2,21, можно сделать вывод, что безработных, имеющих рабочие профессии, намного охотнее берут на работу, чем служащих, которым часто по тем или иным причинам отказывают в трудоустройстве.

В результате исследования влияния специальности безработного с высшим профессиональным образованием на продолжительность поиска работы, можно сделать вывод, что безработные с педагогическими, инженерными специальностями, специальностями сельского и рыбного хозяйства трудоустраиваются быстрее, чем безработные с гуманитарно-социальными специальностями и специальностями экономики и управления.

Из общего числа безработных, составивших выборочную совокупность, официально признанными в качестве безработных являются лишь 42%. С помощью критериев Смирнова и Манна–Уитни проведено сравнение продолжительности поиска работы безработными, имеющими официальный статус и не имеющими такового. Результаты проверки соответствующей гипотезы представлены на рисунках 1.6, 1.7 и свидетельствуют о значимости различий в распределении продолжительности поиска работы двумя категориями безработных в зависимости от наличия официального статуса безработного.

Основные числовые характеристики распределения представлены в таблице 1.3. На основе значений числовых характеристик распределения продолжительности поиска работы можно сделать вывод, что официально зарегист-

рированные безработные трудоустраиваются значительно дольше, чем безработные, не имеющие официальный статус. Это связано с тем, что согласно одному из положений Порядка регистрации безработных граждан безработными признаются граждане, не трудоустроенные в течение 10 дней со дня их регистрации в целях поиска подходящей работы.

Критерий Смирнова									
Группирующая переменная: статус безработного									
Группа 1 - имеют статус безработного Группа 2 - не имеют статус безработного									
Перем.	Макс. отр. Разн.	Макс. пол. Разн.	р-уровень	Среднее Группа 1	Среднее Группа 2	Ст. откл. Группа 1	Ст. откл. Группа 2	Число набл. Группа 1	Число набл. Группа 2
T	-0,000248	0,688809	p < .001	98,49810	14,90020	110,8058	31,29274	12650	21403

Рисунок 1.6 – Результаты проверки гипотезы об отсутствии влияния наличия официального статуса безработного на продолжительность поиска работы с помощью критерия Смирнова

Критерий Манна-Уитни									
Группирующая переменная: статус безработного									
Группа 1 - имеют статус безработного Группа 2 - не имеют статус безработного									
Перем.	Сум. ранг. Группа 1	Сум. ранг. Группа 2	U	Z	р-уровень	Z скорр.	р-уровень	Число набл. Группа 1	Число набл. Группа 2
T	323048665	256771767	27716861	122,8193	0,00	123,0062	0,00	12650	21403

Рисунок 1.7 – Результаты проверки гипотезы об отсутствии влияния наличия официального статуса безработного на продолжительность поиска работы с помощью критерия Манна-Уитни

Таблица 1.3 – Основные числовые характеристики распределения продолжительности поиска работы разными категориями безработных в зависимости от наличия официального статуса безработного (в днях)

Категория безработных	Средняя продолжительность	Медиана	Нижняя квартиль	Верхняя квартиль	Стандартное отклонение
Все безработные	45	8	5	52	82,51
Безработные, имеющие официальный статус	98	60	14	142	110,81
Безработные, не имеющие официальный статус	15	5	4	8	31,29

Практически все официально зарегистрированные безработные (99,9%) получали пособие по безработице. Поскольку пособие не могут получать граждане, не признанные в установленном порядке безработными, то следует ожидать, что продолжительность поиска работы безработными, получающими по-

собие, будет дольше, чем продолжительность поиска работы безработными, не получающими пособие. Это предположение подтверждается результатами проверки соответствующей гипотезы, представленными на рисунках 1.8, 1.9, и оценками основных числовых характеристик распределения продолжительности поиска работы разными категориями безработных в зависимости от выплаты пособия по безработице (таблица 1.4).

Критерий Смирнова									
Группирующая переменная: выплата пособия по безработице									
Группа 1 - получают пособие Группа 2 - не получают пособие									
Перем.	Макс. отр. Разн.	Макс. пол. Разн.	p-уровень	Среднее Группа 1	Среднее Группа 2	Ст. откл. Группа 1	Ст. откл. Группа 2	Число набл. Группа 1	Число набл. Группа 2
T	-0,000248	0,688514	p < .001	98,43005	14,93262	110,7726	31,44196	12652	21401

Рисунок 1.8 – Результаты проверки гипотезы об отсутствии влияния выплаты пособия по безработице на продолжительность поиска работы с помощью критерия Смирнова

Критерий Манна-Уитни									
Группирующая переменная: выплата пособия по безработице									
Группа 1 - получают пособие Группа 2 - не получают пособие									
Перем.	Сум. ранг. Группа 1	Сум. ранг. Группа 2	U	Z	p-уровень	Z скорр.	p-уровень	Число набл. Группа 1	Число набл. Группа 2
T	323047794	256772638	27760537	122,7755	0,00	122,9623	0,00	12652	21401

Рисунок 1.9 – Результаты проверки гипотезы об отсутствии влияния выплаты пособия по безработице на продолжительность поиска работы с помощью критерия Манна-Уитни

Таблица 1.4 – Основные числовые характеристики распределения продолжительности поиска работы разными категориями безработных в зависимости от выплаты пособия по безработице (в днях)

Категория безработных	Средняя продолжительность	Медиана	Нижняя квартиль	Верхняя квартиль	Стандартное отклонение
Безработные, получающие пособие по безработице	98	60	14	142	110,81
Безработные, не получающие пособие по безработице	15	5	4	8	31,44

Средняя продолжительность поиска работы безработными, получающими пособие по безработице, в 6,5 раз больше, чем средняя продолжительность поиска работы безработными, не получающими пособие по безработице. Различия в оценках медианы и верхней квартили еще более значительны.

Во всех рассмотренных выше случаях в основе группировки безработных лежал один признак, принимающий два или более значений. В такой ситуации к одной группе могут относиться безработные, различающиеся по другим не менее важным признакам. Так, безработные одного и того же пола могут отличаться между собой по возрасту и (или) образованию, профессии и др. В связи с этим выявленные различия в продолжительности поиска работы мужчинами и женщинами могут быть обусловлены, например, преобладанием среди мужчин безработных рабочих профессий, а среди женщин – безработных-служащих. Группирование по нескольким признакам позволит более детально изучить различия в продолжительности поиска работы, вызванные социально-демографическими и профессионально-квалификационными характеристикам безработных.

С помощью критерия Краскела–Уоллиса и медианного теста доказано различие во времени трудоустройства безработных, сгруппированных по полу, возрасту и уровню профессионального образования; по полу, возрасту и профессии; по полу, возрасту и специальности (для безработных с высшим профессиональным образованием).

Таким образом, с помощью непараметрических методов анализа полных доказано и охарактеризовано влияние на продолжительность поиска работы таких качественных признаков безработных как пол, наличие опыта работы, возраст, уровень профессионального образования, профессия, специальность, а также наличия официального статуса безработного и выплаты пособия по безработице [11-13]. При этом из рассмотрения были исключены около 26% безработных, для которых точная продолжительность поиска работы была не известна. Эти безработные стали учитываться при реализации второй группы методов – методов анализа данных типа времени жизни [13, 14].

На основе информации о продолжительности поиска работы безработными, зарегистрированными в целях поиска подходящей работы, и официально зарегистрированными безработными построены таблицы времен жизни, позволяющие представить графически оценки функций надежности и функций

риска для рассмотренных категорий безработных. Графики представлены на рисунках 1.10, 1.11.

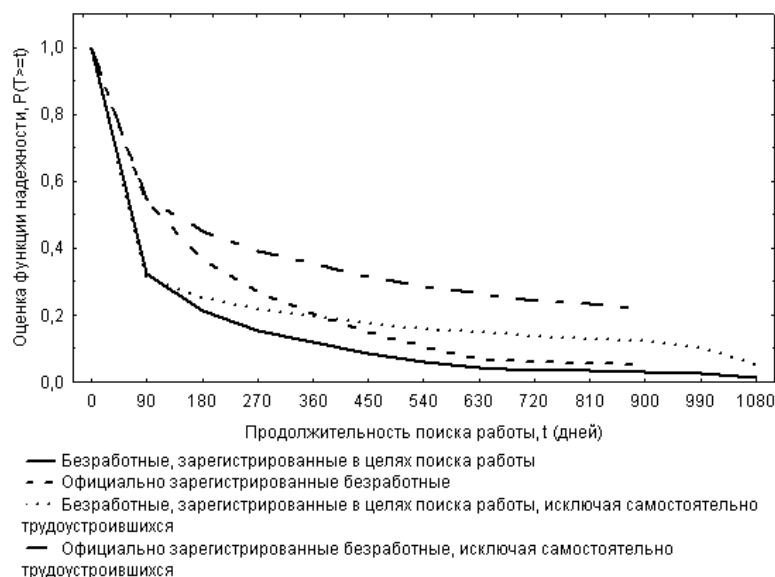


Рисунок 1.10 – Графики выборочных функций надежности для продолжительности поиска работы разными категориями безработных

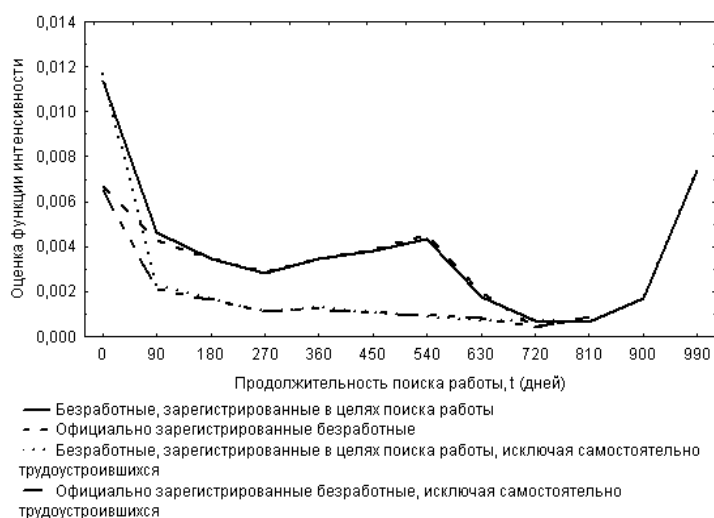


Рисунок 1.11 – Графики выборочных функций интенсивности для продолжительности поиска работы разными категориями безработных

График функции надежности для продолжительности поиска работы официально зарегистрированных безработных лежит выше графика функции надежности для продолжительности поиска работы всеми гражданами, обратившимися в Центр занятости за помощью в трудоустройстве. Это связано прежде всего с тем, что в первом случае не учитываются безработные, трудо-

устроенные в течение 10 дней с момента регистрации в целях поиска подходящей работы. Это же обстоятельство объясняет существенное различие интенсивности трудоустройства рассматриваемых категорий безработных в течение первого интервала времени (до 90 дней). В обоих случаях можно заметить, что исключая из рассмотрения самостоятельно трудоустроившихся безработных, продолжительность поиска работы увеличивается, а интенсивность трудоустройства снижается. Анализируя графики функций интенсивности, можно отметить, что наибольшая интенсивность трудоустройства наблюдается в течение первых недель пребывания на учете в Центре занятости. Затем интенсивность снижается. После 270 дней интенсивность начинает расти, достигая пика для безработных, трудоустроенных при содействии Центра занятости, спустя 1 год (360 дней) после постановки на учет и спустя 1,5 года (540 дней) для безработных, трудоустроенных при содействии Центра занятости или самостоятельно. Увеличение интенсивности трудоустройства в этот период можно объяснить тем, что согласно установленным Законом РФ «О занятости населения в Российской Федерации» условиям и срокам выплаты пособия по безработице, основная часть безработных теряет право на получение пособия после года пребывания на учете. Значения функций интенсивности после 2 лет пребывания на учете построены по небольшому числу наблюдений, поэтому не обладают высокой точностью.

Оценки функций надежности и функций интенсивности, построенные по таблицам времен жизни, зависят от группировки данных по продолжительности поиска работы в интервальный вариационный ряд. Несмотря на это, полученные графики указывают на то, насколько важно в исследовании правильно определиться с объектами наблюдения. Если рассматривать только официально зарегистрированных безработных, то при исследовании не будут учитываться категории граждан, пользующиеся наибольшим спросом на рынке труда, так как они трудоустраиваются за короткий промежуток времени – в течение 10 дней со дня регистрации в целях поиска подходящей работы, а также безработные, отказывающиеся от предлагаемых центром занятости вакансий и

ожидающие работу с лучшими условиями труда. При исключении из рассмотрения самостоятельно трудоустроившихся безработных (они составляют примерно 23% выборочных данных) не будет учитываться возможность найти работу с помощью родственников, друзей, средств массовой информации и других каналов трудоустройства. А ведь далеко не вся информация о вакантных местах попадает в банк вакансий Центра занятости. Туда, как правило, не поступают сведения о престижных рабочих местах, поскольку информация о них распространяется в основном по неформальным каналам. Кроме того, для безработного, зарегистрированного в Центре занятости населения в целях поиска подходящей работы, заранее неизвестно будет он трудоустроен при содействии Центра занятости, трудоустроится самостоятельно или покинет Центр занятости без трудоустройства. Перечисленными обстоятельствами и объясняется тот факт, что в исследовании участвуют не только официально зарегистрированные безработные, трудоустроенные при содействии Центра занятости населения, а все безработные, зарегистрированные в рассматриваемый период времени в Центре занятости в целях поиска подходящей работы.

Графики выборочных функций надежности для продолжительности поиска работы разными категориями безработных в зависимости от пола и возраста представлены на рисунках 1.12, 1.13.

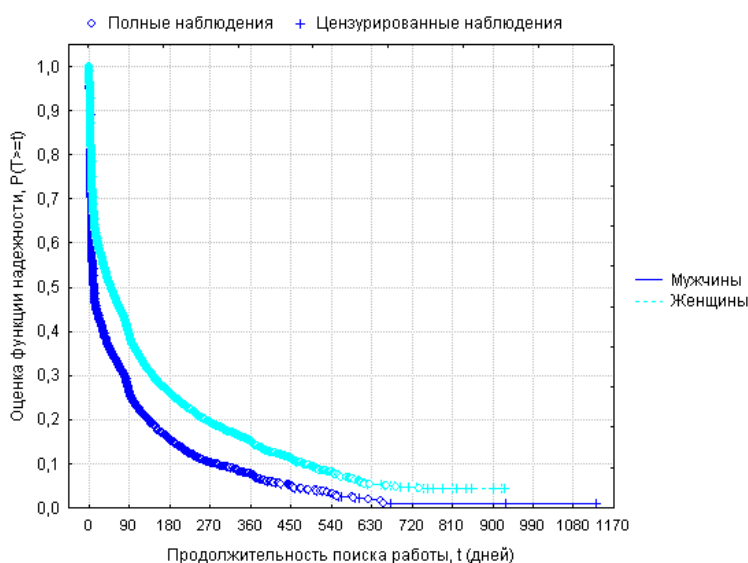


Рисунок 1.12 – Графики выборочных функций надежности для продолжительности поиска работы мужчинами и женщинами



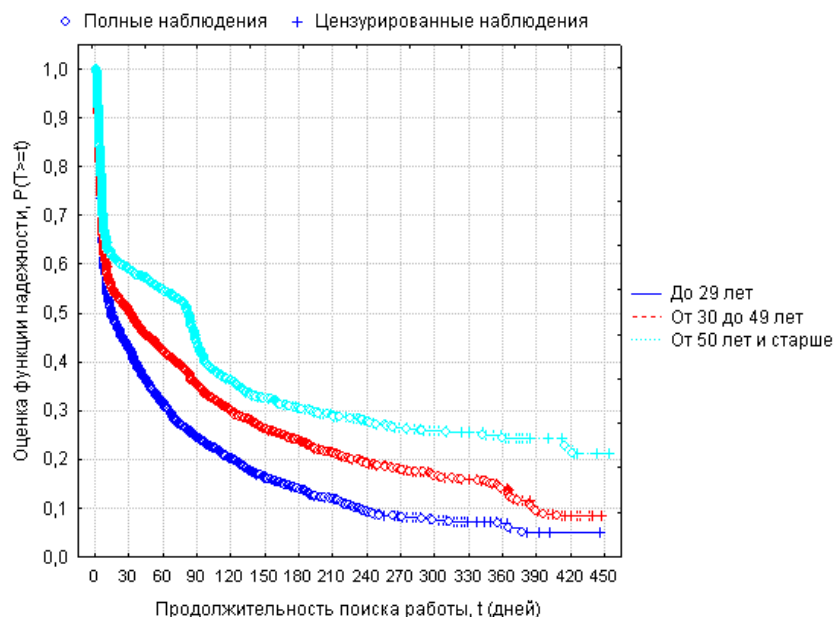


Рисунок 1.13 – Графики выборочных функций надежности для продолжительности поиска работы безработными разных возрастных групп

Для оценки функций надежности использовался метод множительной оценки Каплана–Мейера. Результаты сравнения функций надежности, описывающих распределение продолжительности поиска работы разными категориями безработных, с помощью критерия Хи-квадрат или его нормальной аппроксимации представлены в таблице 1.5.

Таблица 1.5 – Результаты исследования влияния факторов на продолжительность поиска работы с помощью непараметрических методов анализа данных типа времени жизни

Значение статистики	Уровень значимости	Вывод
1	2	3
$H_0$ : пол безработного не оказывает значимое влияние на продолжительность поиска работы		
$z_{набл} = -29,51$	$p = 0,0000$	$H_0$ отвергается
$H_0$ : наличие опыта не оказывает значимое влияние на продолжительность поиска работы		
$z_{набл} = 27,64$	$p = 0,0000$	$H_0$ отвергается
$H_0$ : возраст безработного не оказывает значимое влияние на продолжительность поиска работы		
$\chi^2_{набл} = 236,91$	$p = 0,0000$	$H_0$ отвергается

Продолжение таблицы 1.5

1	2	3
$H_0$ : уровень профессионального образования безработного не оказывает значимое влияние на продолжительность поиска работы		
$\chi^2_{набл} = 582,74$	$p = 0,0000$	$H_0$ отвергается
$H_0$ : профессия безработного не оказывает значимое влияние на продолжительность поиска работы		
$z_{набл} = -48,83$	$p = 0,0000$	$H_0$ отвергается
$H_0$ : специальность безработного не оказывает значимое влияние на продолжительность поиска работы		
$\chi^2_{набл} = 84,38$	$p = 0,0000$	$H_0$ отвергается

Результаты проверки гипотез, полученные на основе анализа данных типа времени жизни, подтверждают вывод, сформулированный на основе анализа полных данных, о том, что качественные характеристики безработного, среди которых пол, возраст, уровень профессионального образования, профессия, наличие опыта работы, специальность, оказывают значимое влияние на продолжительность поиска работы. Интерпретация влияния факторов сводится к следующему:

- мужчины трудоустраиваются быстрее, чем женщины;
- безработные в возрасте 50 лет и старше трудоустраиваются дольше, чем безработные моложе 50 лет, быстрее всех трудоустраиваются безработные в возрасте до 29 лет;
- безработные, не имеющие профессионального образования, трудоустраиваются быстрее, чем безработные с тем или иным уровнем профессионального образования. Дольше всех трудоустраиваются безработные с высшим и средним профессиональным образованием;
- безработные с рабочими профессиями трудоустраиваются быстрее, чем безработные служащие;
- безработные без опыта работы трудоустраиваются быстрее, чем безработные с опытом работы;
- безработные с педагогическими, инженерными специальностями, специальностями сельского и рыбного хозяйства трудоустраиваются быстрее,

чем безработные с гуманитарно-социальными специальностями и специальностями экономики и управления.

Поскольку характеристики безработных могут взаимодействовать друг с другом, то для оценки «чистого» влияния факторов проведен их совместный анализ путем оценки параметров моделей пропорциональных интенсивностей методом Кокса. Требование постоянства отношения функций рисков для двух групп безработных в течение всего периода наблюдения заменено более реалистичным требованием пропорциональности интенсивностей трудоустройства на более коротких интервалах продолжительности поиска работы: до 30 дней, от 30 до 90 дней, от 90 до 180 дней, от 180 до 270 дней, от 270 до 360 дней, более 360 дней. Для каждого из этих интервалов исследовано влияние объясняющих переменных на интенсивность трудоустройства. Для того, чтобы оценить эффект каждого уровня качественного признака, в моделях использовались двоичные переменные:

- женский пол – 1, мужской пол – 0;
- нет опыта работы – 1, есть опыт работы – 0.
- возраст от 30 до 49 лет – 1, другой возраст – 0;
- возраст 50 лет и старше – 1, другой возраст – 0;
- высшее профессиональное образование – 1, другое – 0;
- среднее профессиональное образование – 1, другое – 0;
- нет профессионального образования – 1, другое образование – 0;
- профессия: служащий – 1, рабочий – 0;
- гуманитарно-социальная специальность – 1, другая – 0;
- педагогическая специальность – 1, другая специальность – 0;
- специальность экономики и управления – 1, другая – 0;
- специальность сельского и рыбного хозяйства – 1, другая – 0;
- специальность, не относящаяся к перечисленным направлениям и инженерным специальностям – 1, другая специальность – 0.

В моделях, учитывающих пол безработного, возраст, уровень профессионального образования, профессию, наличие опыта работы, в «стандартную» категорию безработных, соответствующую отсутствию влияния объясняющих переменных, были отнесены мужчины в возрасте до 29 лет, имеющие начальное профессиональное образование, рабочую профессию и опыт работы. Результаты оценивания параметров модели пропорциональных интенсивностей, учитывающей пол безработного, возраст, уровень профессионального образования, профессию и наличие опыта работы, для продолжительности поиска работы до 30 дней представлены в таблице 1.6.

Таблица 1.6 – Результаты оценивания параметров модели пропорциональных интенсивностей, учитывающей пол безработного, возраст, уровень профессионального образования, профессию и наличие опыта работы, для продолжительности поиска работы до 30 дней

Объясняющая переменная ( $X_j$ )	Оценка $\beta_j$ эффекта объясняющей переменной на интенсивность трудоустройства ( $b_j$ )	Стандартная ошибка коэффициента ( $S_{b_j}$ )	Значение $t$ -статистики	Значимость
Женский пол	-0,10	0,02	-5,24	0,00
Возраст от 30 до 49 лет	0,11	0,02	5,53	0,00
Возраст от 50 лет и старше	0,17	0,03	5,66	0,00
Высшее проф. образование	-0,05	0,04	-1,40	0,16
Среднее проф. образование	0,01	0,03	0,38	0,70
Нет проф. образования	0,27	0,03	10,60	0,00
Служащий	-0,23	0,03	-8,28	0,00
Нет опыта работы	0,21	0,04	5,55	0,00

В рассматриваемой модели значимыми являются все коэффициенты кроме коэффициентов при переменных, отвечающих за высшее и среднее профессиональное образование. Это означает, что нет различия в интенсивности трудоустройства безработных, имеющих высшее, среднее и начальное профессиональное образование. При этом коэффициент при переменной, учитывающей отсутствие профессионального образования, значим и оценка его принимает положительное значение, что говорит о том, что у безработных без про-

фессионального образования «риск» трудоустройства выше, чем у безработных, имеющих тот или иной уровень профессионального образования. Влияние остальных объясняющих переменных можно интерпретировать следующим образом:

- интенсивность трудоустройства женщин ниже, чем интенсивность трудоустройства мужчин;
- интенсивность трудоустройства безработных старше 30 лет выше, чем интенсивность трудоустройства безработных моложе 30 лет;
- интенсивность трудоустройства служащих ниже, чем интенсивность трудоустройства рабочих;
- интенсивность трудоустройства безработных без опыта работы выше, чем интенсивность трудоустройства безработных с опытом работы.

Низкая интенсивность трудоустройства безработных в возрасте до 29 лет компенсируется высокой интенсивностью трудоустройства безработных, не имеющих опыта работы, поскольку почти все из них (98%) относятся к первой возрастной группе безработных. По значениям *t*-статистик, можно сделать вывод, что наибольшее влияние на интенсивность трудоустройства оказывают переменные, отвечающие за отсутствие профессионального образования и профессию служащего.

Во второй модели, построенной для продолжительности поиска работы от 30 до 90 дней, незначимыми являются коэффициенты при переменных, учитывающих отсутствие профессионального образования и профессию. Таким образом, для данного интервала нет различия в интенсивности трудоустройства рабочих и служащих, а также безработных, имеющих начальное профессиональное образование и не имеющих профессионального образования. Влияние объясняющих переменных по сравнению с предыдущей моделью изменилось: интенсивность трудоустройства женщин стала выше интенсивности трудоустройства мужчин; наибольшая интенсивность трудоустройства стала наблюдаться у безработных в возрасте до 29 лет, наименьшая интенсивность трудоустройства у безработных в возрасте 50 лет и старше. Наличие высшего или

среднего профессионального способствует увеличению интенсивности трудоустройства, отсутствие опыта работы снижает «риск» трудоустройства.

Для продолжительности поиска работы от 90 до 180 дней незначимо влияние пола и наличия опыта работы. Как и в первой модели для безработных старше 30 лет интенсивность трудоустройства выше, чем у безработных в возрасте до 29 лет. В зависимости от образования наибольшей интенсивностью трудоустройства характеризуются безработные с начальным профессиональным образованием. Интенсивность трудоустройства служащих для этого интервала выше интенсивности трудоустройства рабочих.

В моделях для продолжительности трудоустройства от 180 до 270 дней и от 270 до 360 дней значимым является только коэффициент при переменной, принимающей значение единица для безработных в возрасте от 30 до 49 лет. Оценки коэффициентов положительные, следовательно, интенсивность трудоустройства безработных данной возрастной категории выше, чем у безработных моложе 30 или старше 50 лет.

В последней модели для безработных с продолжительностью трудоустройства более 360 дней значимыми являются коэффициенты при переменных, характеризующих пол, возраст от 30 до 49 лет и наличие опыта работы. По оценкам коэффициентов при этих переменных можно сделать следующие выводы: интенсивность трудоустройства мужчин выше интенсивности трудоустройства женщин, влияние возраста для рассматриваемого интервала аналогично предыдущим двум моделям, для безработных без опыта работы «риск» трудоустройства выше, чем для безработных с опытом работы, однако, безработные без опыта работы составляют лишь 0,4 % всех безработных с продолжительностью поиска работы более года.

Построены также модели, учитывающие пол, возраст, специальность безработного и наличие опыта работы. «Стандартную» категорию безработных составили мужчины в возрасте до 29 лет, инженеры по специальности, имеющие опыт работы. Значимыми являются модели для продолжительности поиска работы до 30 дней, от 30 до 90 дней, от 90 до 180 дней и более 360 дней. Ре-

результаты оценивания параметров моделей представлены в приложении Г. Все факторы оказывают значимое влияние только в первой модели (таблица 1.7).

По оценкам коэффициентов этой модели можно сделать следующие выводы:

- интенсивность трудоустройства мужчин выше интенсивности трудоустройства женщин;
- интенсивность трудоустройства безработных старше 30 лет выше, чем интенсивность трудоустройства безработных моложе 30 лет;
- интенсивность трудоустройства безработных без опыта работы выше интенсивности трудоустройства безработных с опытом работы;

по оценкам коэффициентов при объясняющих переменных, характеризующих специальность безработного, можно выделить две группы безработных, различающихся по интенсивности трудоустройства. Первую группу составляют безработные с гуманитарно-социальными специальностями и специальностями экономики и управления. Для них интенсивность трудоустройства ниже, чем для безработных второй группы, к которым относятся безработные с педагогическими, инженерными специальностями, специальностями сельского и рыбного хозяйства и специальностями других направлений.

Таблица 1.7 – Результаты оценивания параметров модели пропорциональных интенсивностей, учитывающей пол безработного, возраст, специальность и наличие опыта работы, для продолжительности поиска работы до 30 дней

Объясняющая переменная ( $X_j$ )	Оценка $\beta_j$ эффекта объясняющей переменной на интенсивность трудоустройства ( $b_j$ )	Стандартная ошибка коэффициента ( $S_{b_j}$ )	Значение $t$ -статистики	Значимость
1	2	3	4	5
Женский пол	-0,14	0,05	-3,12	0,00
Возраст от 30 до 49 лет	0,19	0,04	4,41	0,00
Возраст от 50 лет и старше	0,27	0,07	4,05	0,00
Нет опыта работы	0,27	0,07	4,00	0,00
Гуманитарно-социальная специальность	-0,17	0,07	-2,29	0,02

Продолжение таблицы 1.7

1	2	3	4	5
Педагогическая специальность	0,03	0,07	0,43	0,66
Специальность экономики и управления	-0,14	0,05	-2,70	0,01
Специальность сельского и рыбного хозяйства	-0,03	0,09	-0,34	0,73
Другая специальность	-0,07	0,10	-0,70	0,48

Для сравнения продолжительности поиска работы безработными в разные годы методом множительной оценки Каплана–Мейера найдены оценки функций надежности для продолжительности поиска работы в разные периоды времени. Графики выборочных функций надежности представлены на рисунке 1.14.

За рассматриваемый период времени с 2000 по 2005 годы наблюдалась тенденция увеличения как доли безработных, трудоустроенных в течение десяти дней с момента регистрации, так и доли безработных с длительным периодом поиска работы. Это объясняется, с одной стороны, расширением банка вакансий, с другой стороны, нарастанием структурного дисбаланса спроса и предложения на рынке труда, проявляющегося в существовании безработицы при достаточном количестве вакантных рабочих мест.

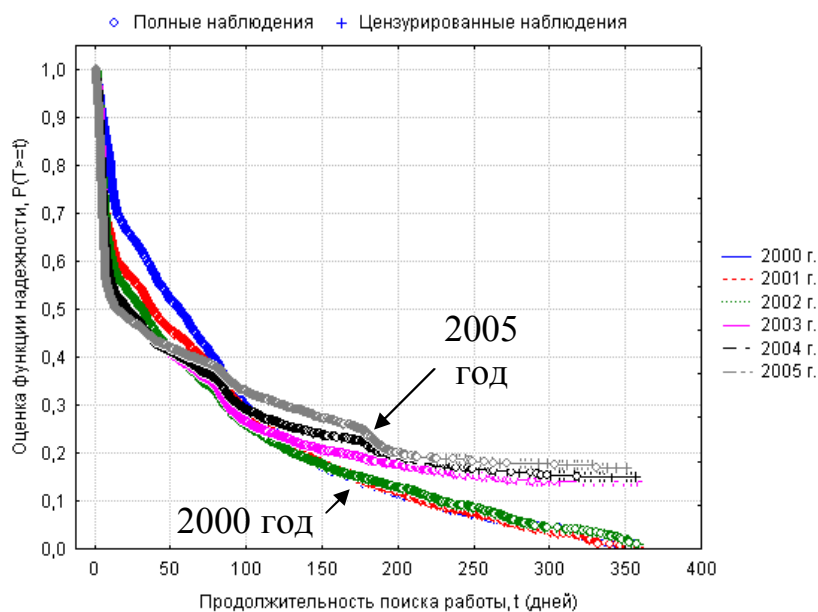


Рисунок 1.14 – Графики выборочных функций надежности для продолжительности поиска работы безработными в разные годы



Для исследования динамики влияния факторов на продолжительность регистрируемой безработицы проведено сравнение функций надежности для продолжительности поиска работы разными категориями безработных в разные годы. Были построены также графики выборочных функций надежности для продолжительности поиска работы различными категориями безработных для каждого года с 2000 по 2005 гг. В результате анализа полученных результатов сформулированы следующие выводы:

- пол безработного, возраст, уровень профессионального образования, профессия, наличие опыта работы и специальность оказывали значимое влияние на продолжительность поиска работы в каждом году наблюдения;
- влияние рассматриваемых факторов усиливалось с каждым годом.

Таким образом, в результате исследования влияния качественных характеристик безработного на продолжительность поиска работы, проведенного с помощью непараметрических статистических методов анализа полных данных и данных типа времени жизни, доказано, что пол безработного, возраст, уровень профессионального образования, профессия, наличие опыта работы, специальность (для безработных с высшим профессиональным образованием) оказывают значимое влияние на время трудоустройства. В результате сравнения основных числовых характеристик распределения продолжительности поиска работы разными категориями безработных и многофакторного регрессионного анализа охарактеризовано влияние каждого из этих факторов на время трудоустройства безработного. Кроме того, исследовано влияние на продолжительность поиска работы наличия официального статуса безработного и выплаты пособия по безработице, проведен анализ особенностей трудоустройства безработных в разные годы. Полученные результаты позволяют сделать вывод о наличии пространственной и временной неоднородности данных, которую необходимо учитывать при моделировании трудоустройства безработных [15, 16].

## Вопросы и задания по теоретическому материалу раздела 1

1) Какие методы называются непараметрическими? Приведите примеры статистических методов, относящихся к непараметрическим

2) Какие две совокупности называются однородными? Приведите классификацию непараметрических методов проверки однородности распределения совокупностей, учитывающую количество совокупностей, зависимость или независимость совокупностей, наличие цензурированных данных

3) Охарактеризуйте непараметрические методы проверки однородности распределения двух независимых совокупностей

4) Какие существуют непараметрические альтернативы однофакторного дисперсионного анализа. Дайте им характеристику

5) Объясните приложение непараметрических методов проверки однородности распределения к выяснению, влияет ли наличие опыта работы и образование безработного на продолжительность поиска работы

6) Прокомментируйте результаты проверки гипотезы об отсутствии влияния профессии безработного на продолжительность поиска работы с помощью критерия Смирнова и критерия Манна-Уитни, представленные на следующих рисунках:

Критерий Смирнова									
Группирующая переменная: профессия									
Группа 1 - рабочие Группа 2 - служащие									
Перем.	Макс. отр. Разн.	Макс. пол. Разн.	р-уровень	Среднее Группа 1	Среднее Группа 2	Ст. откл. Группа 1	Ст. откл. Группа 2	Число набл. Группа 1	Число набл. Группа 2
T	-0,395868	0,000154	p < .001	28,44186	75,68732	64,07637	99,95706	21430	12623

Критерий Манна-Уитни									
Группирующая переменная: профессия									
Группа 1 - рабочие Группа 2 - служащие									
Перем.	Сум. ранг. Группа 1	Сум. ранг. Группа 2	U	Z	р-уровень	Z скорр.	р-уровень	Число набл. Группа 1	Число набл. Группа 2
T	304819325	275001106	75186160	-68,5593	0,00	-68,6637	0,00	21430	12623

7) Прокомментируйте результаты проверки гипотезы об отсутствии влияния специальности безработного на продолжительность поиска работы с

помощью критерия Краскела-Уоллиса и медианного теста, представленные на следующих рисунках:

Завис. переменная: Т	Независимая (группирующая) переменная: специальность Критерий Краскела-Уоллиса: $H(5, N=10253) = 94,61914$ $p = ,0000$		
	Код	Число наблюдений	Сумма рангов
Гуманитарно-социальные спец-ти	1	1148	6228922
Педагогические спец-ти	2	1265	6058840
Спец-ти экономики и управления	3	4254	22905791
Спец-ти сельского и рыбного хоз-ва	4	423	2077435
Инженерные специальности	5	2852	13733516
Другие специальности	6	311	1562628

Завис. переменная: Т	Медианный тест, общая медиана = 36,0000 Независимая (группирующая переменная): специальность Хи-квадрат = 57,09447, $df = 5$ , $p = ,0000$						
	Гум.-соц.	Педагогич.	Экон. и упр.	Сел. и рыб. хоз-ва	Инженер.	Другие	Общий
<= Медиана: набл.	530,000	698,000	1985,000	227,000	1530,000	161,000	5131,00
ожидаемые	574,504	633,055	2128,867	211,6857	1427,252	155,6365	
набл.-ожд.	-44,504	64,945	-143,867	15,3143	102,748	5,3635	
> Медиана: набл.	618,000	567,000	2269,000	196,000	1322,000	150,000	5122,00
ожидаемые	573,496	631,945	2125,133	211,3143	1424,748	155,3635	
набл.-ожд.	44,504	-64,945	143,867	-15,3143	-102,748	-5,3635	
Сумма: набл.	1148,000	1265,000	4254,000	423,000	2852,000	311,000	10253,00

8) Охарактеризуйте непараметрические методы проверки однородности распределения двух зависимых совокупностей

9) Какие существуют непараметрические альтернативы двухфакторного дисперсионного анализа. Дайте им характеристику

10) Могут ли непараметрические методы проверки однородности распределения зависимых совокупностей использоваться для исследования влияния характеристик безработного на продолжительность поиска работы?

11) В чём особенность цензурированных данных или данных типа времени жизни?

12) Чем объясняется наличие цензурированных данных при анализе продолжительности регистрируемой безработицы?

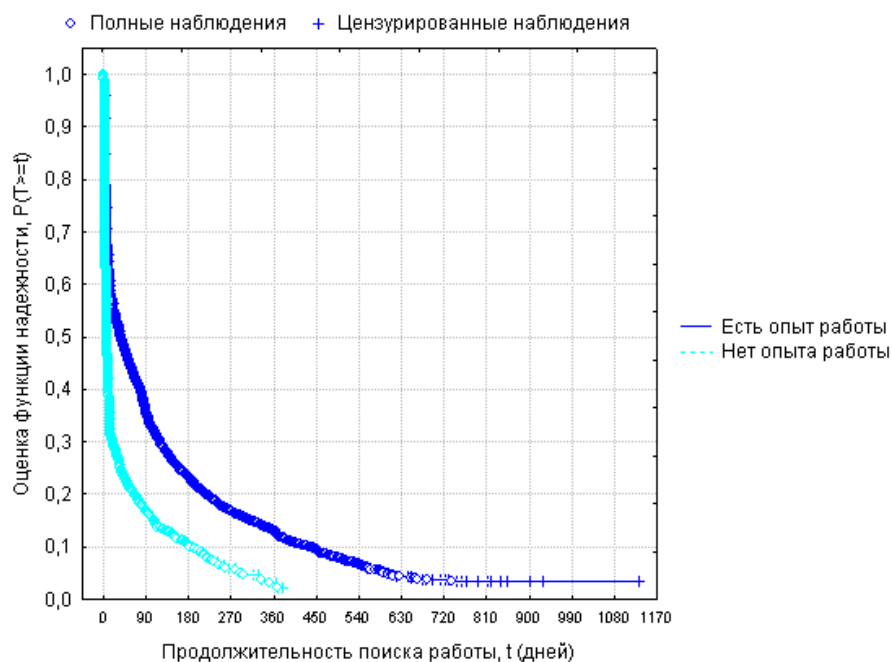
13) Какие функции используются для описания распределения случайной величины  $T$  – продолжительность события до отказа

14) Поясните смысл функции распределения, функции надежности, функции плотности вероятности, функции риска случайной величины  $T$  – продолжительность поиска работы

15) Для чего используется метод множительной оценки Каплана–Мейера и в чём его суть?

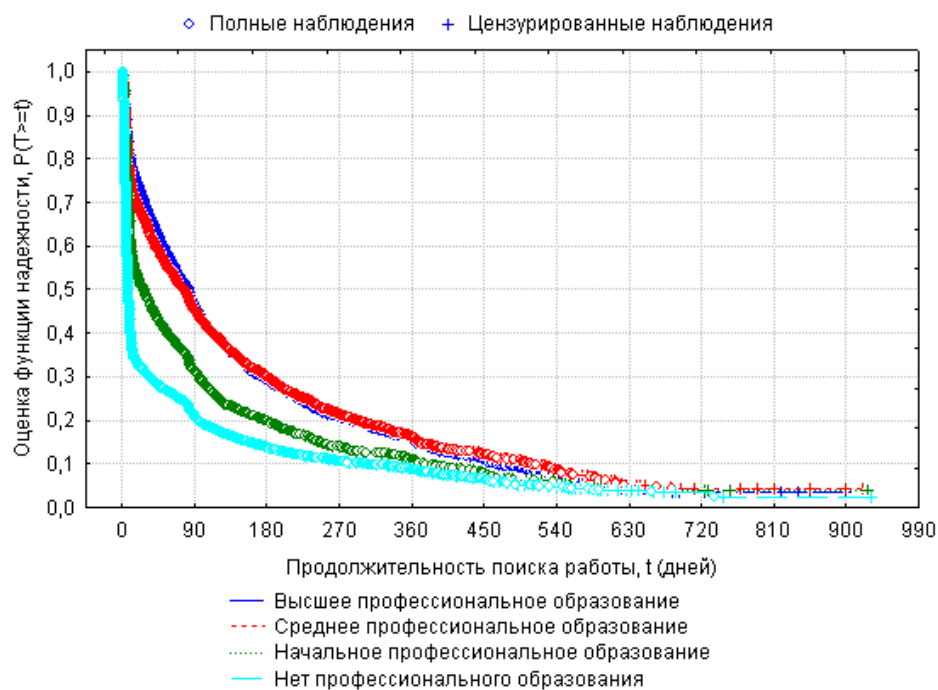
16) Охарактеризуйте непараметрические критерии сравнения функций надежности при наличии в выборочных данных цензурированных наблюдений

17) Графики выборочных функций надежности для продолжительности поиска работы безработными с опытом и без опыта работы представлены на следующем рисунке:



Охарактеризуйте влияние наличия опыта работы на продолжительности поиска работы

18) Графики выборочных функций надежности для продолжительности поиска работы безработными с разным уровнем профессионального образования представлены на следующем рисунке:



Охарактеризуйте влияние уровня профессионального образования на продолжительности поиска работы

- 19) Дайте характеристику модели пропорциональных интенсивностей
- 20) Какие существуют методы оценки параметров модели пропорциональных интенсивностей и в чём их суть?
- 21) Результаты оценивания параметров модели пропорциональных интенсивностей для продолжительности поиска работы в пакете Statistica представлены на следующем рисунке:

	Хи-квадрат = 86,2825 df = 9 p = ,00000			
	Козф.	Станд. ошибка	t-значение	p
Пол женский	-0,143191	0,045939	-3,11699	0,001829
Возраст от 30 до 50 лет	0,190045	0,043076	4,41189	0,000010
Возраст от 50 лет и старше	0,269807	0,066538	4,05496	0,000050
Нет опыта работы	0,268045	0,066999	4,00071	0,000063
Гуманитарно-социальная спец-ть	-0,169847	0,074277	-2,28666	0,022222
Педагогическая специальность	0,028983	0,066903	0,43320	0,664871
Спец-ть экономики и управления	-0,139006	0,051425	-2,70306	0,006874
Спец-ть сельского и рыбного хозяйства	-0,031960	0,094248	-0,33911	0,734530
Другая специальность	-0,071388	0,102332	-0,69761	0,485424

Запишите оценку модели пропорциональных интенсивностей и охарактеризуйте полученные результаты.

22) Сформулируйте постановку задачи исследования влияния социально-демографических и профессионально-квалификационных характеристик безработного на продолжительность поиска работы. Опишите алгоритм и математический инструментарий решения задачи

### **Задание к исследовательской работе**

Исходная информация – данные о продолжительности поиска работы безработными, зарегистрированными в Центре занятости населения и характеризующимися следующими признаками:

- пол: мужской, женский;
- возраст: до 29 лет, от 30 до 49 лет, от 50 лет и старше;
- уровень профессионального образования: высшее, среднее, начальное, нет профессионального образования;
- профессия: рабочий, служащий;
- наличие опыта работы: нет опыта работы, есть опыт работы.

Провести исследование влияния социально-демографических и профессионально-квалификационных характеристик безработного на продолжительность поиска работы.

Тестовые исходные данные приведены в файле bezrub.xls. По возможности собрать свою информационную базу, увеличив количество признаков – характеристик безработного. В исследовании не ограничиваться использованием рассмотренных в главе 1 методов. Расширение математического инструментария возможно за счет привлечения методов анализа таблиц сопряженности, методов анализа соответствий и других математико-статистических методов.

## Список использованных источников к разделу 1

- 1 Lancaster, T. The Econometric Analysis of Transition Data / T. Lancaster. - Cambridge, Cambridge University Press, 1990.
- 2 Юнкеров, В.И. Математико-статистическая обработка данных медицинских исследований / В.И. Юнкеров, С.Г. Григорьев. – СПб.: ВМедА, 2002. – 266 с.
- 3 Орлов, А.И. Эконометрика: учебник для вузов / А.И. Орлов. - 3-е изд., перераб. и доп. – М.: Издательство «Экзамен», 2004. – 567 с.
- 4 Холлендер, М. Непараметрические методы статистики / Пер. с англ. Д.С. Шмерлинга; Под ред. Ю.П. Адлера и Ю.Н. Тюрина; предисл. Ю.П. Адлера, Ю.Н. Тюрина и Д.С. Шмерлинга. – М.: Финансы и статистика, 1983. – 518 с.
- 5 Тихомиров, Н.П. Эконометрика: Учебник / Н.П. Тихомиров, Е.Ю. Дорохина – М.: Издательство «Экзамен», 2003. – 512 с.
- 6 Четыркин, Е.М. Вероятность и статистика / Е.М. Четыркин, И.Л. Калихман. – М.: Финансы и статистика, 1982. – 319 с.
- 7 Прикладная статистика. Основы эконометрики: учебник для вузов: В 2 т. 2-е изд., испр. – Т.1: Айвазян С.А., Мхитарян В.С. Теория вероятностей и прикладная статистика. – М.: ЮНИТИ-ДАНА, 2001. – 656 с.
- 8 Боровиков, В. STATISTICA. Искусство анализа данных на компьютере: Для профессионалов / В. Боровиков. - 2-е изд. (+CD). – СПб.: Питер, 2003. – 688 с.
- 9 Кокс, Д.Р. Анализ данных типа времени жизни / Д.Р. Кокс, Д. Оукс. – М.: Финансы и статистика, 1988. – 191 с.
- 10 Kiefer, N. Economic Duration Data and Hazard Functions, Journal of Economic Literature Vol. XXVI (June 1988). pp. 646–679.
- 11 Реннер, А.Г. Исследование влияния факторов на продолжительность пребывания выпускника вуза на учете в Центре занятости / А.Г. Реннер, О.С. Бравичева // Математико-статистический анализ социально-

экономических процессов: межвузовский сборник научных трудов. Выпуск 2. – М.: МЭСИ, 2004. – С. 115–117.

12 Бравичева, О.С. Использование непараметрических методов статистики для оценки детерминант продолжительности безработицы / О.С. Бравичева // Научные школы и результаты в российской статистике: материалы Международной научно-практической конференции. – СПб.: Знание, 2006. – С. 237–239.

13 Бравичева, О.С. Методы и инструментальные средства исследования влияния факторов на продолжительность поиска работы / О.С. Бравичева // Вестник ОГУ. – 2006. – №2. – С. 140–144.

14 Бравичева, О.С. Оценка детерминант продолжительности безработицы на основе анализа транзитных данных / О.С. Бравичева // Взаимодействие реального и финансового сектора в трансформационной экономике: материалы всероссийской научно-практической конференции. – Оренбург: РИК ГОУ ВПО ОГУ, 2006. – С. 505–510.

15 Реннер, А.Г. Моделирование рынка труда с учетом неоднородности данных / А.Г. Реннер, О.С. Бравичева // Вестник ОГУ. – 2005. – №10. Приложение. – С. 128–134.

16 Математические методы моделирования социально-экономических процессов (региональный аспект) / А.Г. Реннер [и др.]. – Самара: Изд-во СамНЦ РАН, 2008. – 182 с.



## **2 Методы многомерного статистического анализа в социально-экономических исследованиях**

Раздел состоит из восьми подразделов. Первые три подраздела – теоретические. В первом и втором представлен обзор методов многомерной классификации и методов снижения размерности признакового пространства, отмечены особенности методов и условия их применения. Третий подраздел посвящен решению актуальной задачи построения интегрального показателя качества (эффективности функционирования) объекта (системы). Здесь рассмотрены такие подходы как: построение интегрального показателя в случае работоспособности первой главной компоненты (методика С.А. Айвазяна), взвешенная процедура индивидуального рейтингования в пространстве модифицированных первых главных компонент, экспертно-статистический метод (при наличии экспертной информации в форме булевой матрицы). В подразделе 2.4 представлено территориальное зонирование города Оренбурга на основе факторов, определяющих цену жилья на вторичном рынке на основе методов многомерной классификации. Подраздел 2.5 посвящен кластеризации административно-территориальных образований по показателям экономической, демографической безопасности, в подразделе 2.5 представлена кластеризация муниципалитетов по показателям инвестиционной привлекательности. Реализация процедур многомерной классификации осуществлялась в ППП Statistica 6.0. В подразделе 2.7 решена задача многомерной классификации административно-территориальных образований Оренбургской области по показателям, характеризующим антропогенную нагрузку и эколого-экономический риск, методами многомерного шкалирования, которые реализованы в ППП SPSS. Подраздел 2.8 состоит из трех пунктов, посвященных построению интегральных показателей уровня демографической, экономической безопасности и инвестиционной привлекательности на примере Оренбургской области. Для обработки данных, необходимых для построения интегрального показателя, использовался ППП Statistica 6.0. В конце раздела приведены вопросы и задания по теорети-

ческому материалу, задания к исследовательской работе и список использованных источников.

## **2.1 Обзор методов многомерной классификации**

При анализе социально-экономических процессов, характеризующихся достаточно большим набором показателей, объекты могут существенно различаться по значениям этих показателей. Поэтому одной из важнейших задач многомерного статистического анализа является выявление однородных групп объектов, т.е. объектов схожих друг с другом по рассматриваемым свойствам. Решение такой задачи при небольшом наборе показателей традиционно осуществляется методами комбинационной группировки. Если же количество признаков велико, то требуется использование специальных методов многомерной классификации – кластерного, дискриминантного анализа и статистических пакетов, их реализующих.

Кластерный анализ – это совокупность методов, позволяющих классифицировать многомерные наблюдения на однородные, в определенном смысле, группы или классы. В отличие от комбинационных группировок кластерный анализ приводит к разбиению объектов на группы с учетом всех группировочных признаков одновременно. Методы кластерного анализа позволяют решать следующие задачи:

- 1) проведение классификации объектов с учетом признаков, отражающих сущность, природу объектов. Решение такой задачи, как правило, приводит к углублению знаний о совокупности классифицируемых объектов;
- 2) проверка выдвигаемых предположений о наличии некоторой структуры в изучаемой совокупности объектов, т.е. поиск существующей структуры;
- 3) построение новых классификаций для слабоизученных явлений, когда необходимо установить наличие связей внутри совокупности и попытаться привести в нее структуру.

В кластерном анализе однородными, как правило, считаются объекты, близкие между собой в смысле выбранной метрики. Выбор метрики – ключевой, но при этом наименее формализованный этап решения задачи многомерной классификации. Сходство или различие между классифицируемыми объектами устанавливается в зависимости от метрического расстояния между ними. Если каждый объект описывается  $k$  признаками, то он может быть представлен как точка в  $k$ -мерном пространстве, и сходство с другими объектами будет определяться как соответствующее расстояние. В кластерном анализе используются различные меры расстояния между объектами:

1) евклидово расстояние:

$$d_E(O_i, O_j) = \sqrt{\sum_{l=1}^k (x_{il} - x_{jl})^2}, \quad (2.1)$$

где  $x_{il}$ ,  $x_{jl}$  - величина  $l$ -й компоненты у  $i$ -го ( $j$ -го) объекта ( $l = 1, 2, \dots, k$ ;  $i, j = 1, 2, \dots, n$ ).

Данная мера различия объектов используется в трех случаях:

- наблюдения извлекаются из нормально распределенной генеральной совокупности с ковариационной матрицей вида  $\Sigma = \sigma^2 E_k$  (компоненты вектора  $X$  взаимно некоррелированы и имеют одинаковую дисперсию);

- компоненты вектора наблюдений  $X$  однородны по физическому смыслу и одинаково важны для классификации;

признаковое пространство совпадает с геометрическим пространством.

2) взвешенное евклидово расстояние:

$$d_{BE}(O_i, O_j) = \sqrt{\sum_{l=1}^k \omega_l (x_{il} - x_{jl})^2}. \quad (2.2)$$

Данная мера различия объектов применяется в случаях, когда каждой компоненте вектора наблюдений  $X$  удается приписать некоторый «вес»  $\omega_l$ , пропорциональный степени важности признака в задачи классификации.

3) расстояние Минковского:

$$d_M(O_i, O_j) = \left( \sum_{l=1}^k |x_{il} - x_{jl}|^p \right)^{\frac{1}{p}}. \quad (2.3)$$

4) Хеммингово расстояние (манхеттеновское расстояние, расстояние city-block) часто применяется как мера различия объектов, задаваемых дихотомическими признаками (частный случай расстояния Минковского при  $p = 1$ ):

$$d_H(O_i, O_j) = \sum_{l=1}^k |x_{il} - x_{jl}|. \quad (2.4)$$

5) в случае зависимых переменных и их различной значимости в решении вопроса классификации обычно используют обобщенное (взвешенное) расстояние Махаланобиса:

$$d(O_i, O_j) = \sqrt{(X_i - X_j)^T S^{-1} (X_i - X_j)}, \quad (2.5)$$

где  $X_i; X_j$  - векторы значений переменных у  $i$ -го и  $j$ -го объектов;

$S$  - общая ковариационная матрица.

В ряде процедур классификации используют понятия расстояния между группами объектов и меры близости двух групп объектов. Пусть  $S_i$  -  $i$ -я группа (класс, кластер), состоящая из  $n_i$  объектов;  $\bar{x}_i$  - «центр тяжести»  $i$ -ой группы;  $\rho(S_l, S_m)$  - расстояние между группами  $S_l, S_m$ . Наиболее употребительными расстояниями и мерами близости между классами объектов являются:

- расстояние, измеряемое по принципу «ближайшего соседа»:

$$\rho_{\min}(S_l, S_m) = \min_{\substack{x_i \in S_l \\ x_j \in S_m}} \rho(x_i, x_j); \quad (2.6)$$

- расстояние, измеряемое по принципу «дальнего соседа»:

$$\rho_{\max}(S_l, S_m) = \max_{\substack{x_i \in S_l \\ x_j \in S_m}} \rho(x_i, x_j); \quad (2.7)$$

- расстояние, измеряемое по «центрам тяжести» групп:

$$\rho(S_l, S_m) = \rho(\bar{x}_l, \bar{x}_m); \quad (2.8)$$

- расстояние, измеряемое по принципу «средней связи». Это расстояние определяется как среднее арифметическое всех попарных расстояний между представителями рассматриваемых групп:

$$\rho_{cp}(S_l, S_m) = \frac{1}{n_l n_m} \sum_{x_i \in S_l} \sum_{x_j \in S_m} \rho(x_i, x_j); \quad (2.9)$$

- обобщенное расстояние, которое в качестве частных случаев включает в себя все рассмотренные выше виды расстояний, основано на понятии степенного среднего и определяется формулой:

$$\rho_{об}(S_l, S_m) = \left[ \frac{1}{n_l n_m} \sum_{x_i \in S_l} \sum_{x_j \in S_m} \rho^r(x_i, x_j) \right]^{1/r}. \quad (2.10)$$

Существуют иерархические и итерационные методы классификации. Среди множества методов иерархического кластерного анализа наиболее распространены метод полных связей, метод средней связи, метод Уорда. Основным достоинством этих методов является возможность их реализации при не-

известном количестве классов. Недостатком методов является необходимость постоянного хранения и пересчета матрицы расстояний между объектами (классами).

Особенностью итерационных методов кластерного анализа является необходимость задания некоторых начальных условий классификации (количество образуемых кластеров, порог завершения процесса классификации и т.д.). Итерационные методы в большей степени, чем иерархические, требуют от пользователя интуиции при выборе типа классификационных процедур и задания начальных условий разбиения, так как большинство этих методов очень чувствительны к изменению задаваемых параметров. Например, выбранное случайным образом число кластеров может не только сильно увеличить трудоемкость процесса классификации, но и привести к образованию «размытых» или мало наполняемых кластеров. Поэтому целесообразно сначала провести классификацию по одному из иерархических методов или на основании экспертных оценок, а затем уже подбирать начальное разбиение и статистический критерий для работы итерационного алгоритма.

При использовании различных методов кластерного анализа для одной и той же совокупности данных могут быть получены различные варианты разбиения. Существенное влияние на результат классификации оказывают метрика расчета расстояния между объектами, алгоритм классификации, значения начальных условий и других параметров классификации. Возникает задача выбора «лучшего» разбиения. Для решения этой задачи используется некоторая мера качества классификации, которую принято называть функционалом или критерием качества. Наилучшим по выбранному функционалу считается такое разбиение, при котором достигается экстремальное (минимальное или максимальное) значение целевой функции – функционала качества. Выбор того или иного функционала качества, как правило, осуществляется весьма произвольно и опирается скорее на эмпирические и профессионально – интуитивные соображения, чем на строгие правила.

В отличие от кластерного анализа содержанием дискриминантного анализа является разработка методов решения задач различения (дискриминации) объектов наблюдения по определенным признакам в ситуации, когда исследователь обладает обучающей информацией, так называемыми обучающими выборками. Все процедуры дискриминантного анализа можно разбить на две группы. Первая группа процедур позволяет интерпретировать различия между существующими классами, вторая – проводить классификацию новых объектов в тех случаях, когда неизвестно заранее, к какому из существующих классов они принадлежат.

Следует отметить, что упомянутые методы многомерной классификации успешно работают при относительно устойчивом характере распределении экспериментальных данных в пространстве признаков. При высокой зашумленности экспериментальных данных и их противоречивости, а также когда распределение признаков отлично от нормального и характеризуется ярко выражено асимметричностью, эти методы являются неработоспособными. В этом случае рекомендуется проводить классификацию на базе нейронных сетей [1, 2].

К преимуществам классификации на основе нейронных сетей относятся:

- 1) нейронные сети справляются с решением задач классификации, когда распределение признаков отлично от нормального и асимметрично;
- 2) нейронные сети нечувствительны к корреляции признаков;
- 3) нейросетевые классификаторы оценивают апостериорную Байесовскую вероятность и поэтому аппроксимируют оптимальный статистический классификатор с минимальной ошибкой. Подобная статистическая интерпретация значений выходов нейронной сети позволяет, в частности, компенсировать обычно существующие диспропорции в объемах классов.

Основой классификации на базе нейронных сетей служит вектор параметров объекта. Компоненты этого вектора представляют собой различные характеристики объекта, которые влияют на принятие решения о том, к какому классу можно отнести данный образец. Объекты в пределах одного класса счи-

таются эквивалентными с точки зрения критерия разбиения. Сами классы часто бывают неизвестны заранее, а формируются динамически (как, например, в сетях Кохонена). Классы зависят от предъявляемых объектов, и поэтому добавление нового объекта требует корректирования системы классов.

Сети Кохонена рассчитаны на неуправляемое обучение. При управляемом обучении наблюдения, составляющие обучающиеся данные, вместе с входными переменными содержат также и соответствующие им выходные значения, переводящие первые во вторые. В случае неуправляемого обучения обучающие данные содержат только значения входных переменных. Сеть Кохонена учится понимать саму структуру данных.

Сеть Кохонена может распознавать кластеры в данных, а также устанавливать близость классов. Таким образом, пользователь может улучшить свое понимание структуры данных, чтобы затем уточнить нейросетевую модель. Сеть Кохонена можно использовать и в тех случаях, когда классы уже заданы – тогда преимущество будет в том, что сеть сможет выявить сходство между различными классами. Другая возможная область применения сети Кохонена – обнаружение новых явлений. Сеть Кохонена распознает кластеры в обучающих данных и относит все данные к тем или иным кластерам. Если после этого сеть встретится с набором данных, непохожим ни на один из известных образцов, то она не сможет классифицировать такой набор и тем самым выявит его новизну.

## **2.2 Методы снижения размерности признакового пространства**

Зачастую анализ социально-экономических процессов пытаются, из благих побуждений, проводить на основе как можно большего числа показателей  $x^{(1)}, x^{(2)}, \dots, x^{(k)}$ , создавая тем самым для исследователя такие ситуации, как:

- возможная мультиколлинеарность;
- невариативность данных;



- ненаглядность информации;
- скрытая агрегированность, не позволяющая увидеть интегральные свойства.

Одним из подходов к разрешению перечисленных ситуаций является переход к новому набору признаков, который может строиться в виде линейных комбинаций исходных центрированных (центрировано-нормированных признаков), например, из следующих соображений:

- новые признаки ортогональны;
- мера информативности новых признаков  $z^{(1)}, z^{(2)}, \dots, z^{(k')}$  определяется:

$$I_{k'}(z) = \frac{Dz^{(1)} + Dz^{(2)} + \dots + Dz^{(k')}}{Dx^{(1)} + Dx^{(2)} + \dots + Dx^{(k)}}, \quad (2.11)$$

где  $Dz^{(1)} \geq Dz^{(2)} \dots \geq Dz^{(k')}$ ,  $0 \leq I_{k'}(z) \leq 1$ .

Легко видеть, что мера информативности является возрастающей, ограниченной функцией дисперсий новых признаков.

Набор показателей, обладающих указанными свойствами, может быть построен путем последовательного применения критерия максимума дисперсии линейной комбинации исходных (центрированных или центрировано-нормированных) признаков, ортогональной с предшествующими. Построенные, таким образом, новые признаки называются главными компонентами. Их нахождение, из критерия (2.12)

$$\begin{cases} D(l^T x) \rightarrow \max \\ l^T l = 1 \end{cases} \quad (2.12)$$

сводится к нахождению собственных чисел и собственных векторов либо ковариационной матрицы исходных признаков (если все они измерены в одной шкале), либо корреляционной матрицы исходных признаков (если они измерены в разных шкалах).

Геометрически речь идет о переходе от исходной системы координат  $x_1, x_2, \dots, x_k$  к новой системе координат  $z_1, z_2, \dots, z_k$  с базисом из собственных векторов ковариационной (корреляционной) матрицы. Новые признаки нумеруются в порядке убывания дисперсий (собственных чисел  $\Sigma$  или  $R$ ), а мера информативности  $k'$  из них представляется в виде (2.13):

$$I_{k'}(z) = \frac{\lambda_1 + \dots + \lambda_{k'}}{\lambda_1 + \dots + \lambda_k}. \quad (2.13)$$

Следует отметить, что свойство ортогональности главных компонент позволяет решать проблему мультиколлинеарности.

Уменьшение дисперсии каждой последующей главной компоненты позволяет отбросить малоинформативные новые признаки, тем самым снизить размерность признакового пространства и, возможно, решить проблему наглядности. Если уровень информативности первой главной компоненты больше 55%, то ее рекомендуется брать в качестве латентного интегрального показателя, аккумулирующего влияние группы факторов.

Содержательную интерпретацию главным компонентам иногда удается дать на основе матрицы факторных нагрузок (2.14):

$$A = L\Lambda^{1/2}, \quad (2.14)$$

где  $L$  – матрица, вектор-столбцами которой являются собственные вектора;

$\Lambda^{1/2}$  - диагональная матрица, построенная на основе собственных чисел  $\lambda_i$ :

$$\Lambda^{1/2} = \begin{pmatrix} \lambda_1^{1/2} & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \lambda_k^{1/2} \end{pmatrix}.$$

Отметим, что если исходные признаки  $x_1, x_2, \dots, x_k$  центрированы и нормированы, то элементы матрицы нагрузок  $a_{ij} = \rho_{x_i z_j}$ . Отобрав признаки

$x_{i_1}, x_{i_2}, \dots, x_{i_s}$  наиболее тесно связанные с  $z_j$  ( $|\rho_{x_{i_l} z_j}| \geq 0,7, l = \overline{1, s}$ ) мы, возможно, дадим содержательную интерпретацию главной компоненте  $z_j, j = \overline{1, k}$ .

Еще одно важное свойство главных компонент – свойство наилучшей воспроизводимости, состоящее в том, что с помощью первых  $k'$  главных компонент  $z_1, z_2, \dots, z_{k'}$  ( $k' < k$ ) достигается наилучший прогноз исходных признаков среди всех прогнозов, которые можно построить с помощью  $k'$  линейных комбинаций набора из  $k$  произвольных признаков.

Приведенные выше рассуждения носят теоретический характер, а практическая реализация требует использования математико-статистического инструментария для того, чтобы по некоторой части генеральной совокупности (по выборке) производить оценку упомянутых характеристик и исследовать свойства оценок.

Свойства выборочных оценок, алгоритм построения оценок и их статистического анализа отражен в курсе «Многомерные статистические методы» или «Анализ данных» и в [3, 4]. Здесь можно только отметить, что результаты исследований опираются на предположение о нормальности исследуемой генеральной совокупности и взаимной независимости извлеченных из нее наблюдений. При этом уже на этапе перехода к «центрированным» наблюдениям исходных признаков невозможно соблюсти независимость. Однако, при достаточно большом  $n$  можем пренебречь эффектом нарушения независимости.

Следует сказать, что имеются и другие, в определенных ситуациях более эффективные методы снижения размерности признакового пространства, к примеру, метод главных компонент, как метод снижения размерности признакового пространства, можно считать частным случаем факторного анализа, в котором каждый из исходных признаков представляется в виде линейной комбинации новых признаков (общих факторов) и некоторого специфического (характерного) для каждого исходного признака фактора.

Другой пример – методы многомерного шкалирования, которые также позволяют выделять латентные показатели, лежащие в основе дифференци-

ции исследуемых объектов, однако имеют определенное преимущество перед методом главных компонент в меньших требованиях к шкале представления исходной информации. Основная идея многомерного шкалирования заключается в отображении информации о латентной конфигурации точек, содержащейся в матрице расстояний (или близости) между объектами, в виде геометрической конфигурации представляющих их точек в новом координатном пространстве небольшой размерности с расстояниями (или близостями), которые мало отличаются от исходной матрицы в смысле некоторого критерия. Методами многомерного шкалирования часто удается получить более компактное решение и учесть конфигурации данных более сложной структуры, чем в классическом методе главных компонент. Однако, при малой информативности двухфакторного решения, результаты шкалирования являются трудноинтерпретируемыми, в результате чего можно сформулировать некорректные выводы относительно структуры данных.

### **2.3 Методы построения интегральных показателей**

И в профессиональной деятельности, и в повседневной жизни человек постоянно сталкивается с ситуациями, когда ему приходится сравнивать между собой и упорядочивать, по некоторому, не поддающемуся непосредственному измерению свойству, ряд сложных систем. При этом общее представление о степени проявления анализируемых латентных (т.е. не поддающегося непосредственному количественному измерению) свойств складывается как результат определенного «суммирования» поддающихся измерению факторов. Необходимость использования интегрального показателя диктуется, с одной стороны, наличием большого числа данных, принимаемых во внимание при формировании латентной категории, с другой стороны, ограниченными возможностями человека за конечное время обобщать наборы разнородной информации [5].

Построение интегрального показателя возможно осуществить на основе:

1) компонентного анализа (в случае работоспособности первой главной компоненты);

2) экспертно-статистическим методом (при наличии экспертной информации в форме булевой матрицы).

Построение интегрального показателя в условиях отсутствия экспертных оценок может быть сведено к построению первой главной компоненты частных унифицированных показателей. Данная методика предложена С.А. Айвазяном [6], согласно которой, в качестве интегрального показателя, характеризующего некое латентное свойство, берется первая главная компонента, если выполняется достаточно жесткое требование о том, что уровень ее информативности превышает 55%. Преимущество данного метода заключается в том, что он позволяет отказаться от дорогостоящих услуг экспертов.

Перед тем, как начать процедуру свертки, необходимо привести все анализируемые показатели  $x^{(j)}$  ( $j = 1, 2, \dots, k$ ) к «общему знаменателю», т. е. применить к каждому из них такое преобразование, в результате которого область его возможных значений определится отрезком  $[0, 1]$ . При этом, нулевое значение преобразованного показателя будет означать самое низкое качество по данному свойству, а единичное – самое высокое. Таким образом, мы перейдем к унифицированной шкале в измерении интегрального критерия [7].

Все анализируемые показатели, характеризующие некое латентное свойство, могут быть разбиты на два непересекающихся класса:

- а) класс монотонно изменяющихся свойств (монотонные показатели);
- б) класс немонотонных показателей.

Класс (а) состоит из показателей  $x^{(j)}$  ( $j = 1, 2, \dots, k_1$ ), значения которых монотонно связаны с количественной оценкой анализируемого свойства. В нашем случае к этому классу относятся все рассматриваемые показатели, следовательно, следует перейти к преобразованным показателям  $\tilde{x}^{(j)}$ , связанным с исходными переменными  $x^{(j)}$  преобразованием:

$$\tilde{x}^{(j)} = \frac{x^{(j)} - x_{\min}^{(j)}}{x_{\max}^{(j)} - x_{\min}^{(j)}} \quad (2.15)$$

или

$$\tilde{x}^{(j)} = 1 - \frac{x^{(j)} - x_{\min}^{(j)}}{x_{\max}^{(j)} - x_{\min}^{(j)}}. \quad (2.16)$$

В соотношениях (2.15) и (2.16)  $x_{\min}^{(j)}$  и  $x_{\max}^{(j)}$ , соответственно, - наименьшее и наибольшее возможные значения  $j$ -ой исходной характеристики. При этом если большие значения  $x^{(j)}$  соответствуют более высокому качеству, то при переходе к нормированным показателям  $\tilde{x}^{(j)}$  следует воспользоваться преобразованием (2.15). Если же, напротив, относительно большие значения  $x^{(j)}$  свидетельствуют о плохом качестве, то при переходе к нормированным показателям  $\tilde{x}^{(j)}$  следует воспользоваться преобразованием (2.16).

Класс (б) состоит из исходных характеристик  $x^{(j)}$  ( $j = 1, 2, \dots, k_2$ ), для которых существует некоторое оптимальное значение  $x_{opt}^{(j)}$ , определенное внутри диапазона изменения этого показателя  $[x_{\min}^{(j)}, x_{\max}^{(j)}]$ . В этом случае переход к преобразованным переменным  $\tilde{x}^{(j)}$  рекомендуется осуществлять с помощью преобразования вида:

$$\tilde{x}^{(j)} = 1 - \frac{|x^{(j)} - x_{opt}^{(j)}|}{\max\{(x_{opt}^{(j)} - x_{\min}^{(j)}), (x_{\max}^{(j)} - x_{opt}^{(j)})\}} \quad (2.17)$$

Главные компоненты, согласно подразделу 2.2, строятся в виде:

$$Z = l^T X, \quad (2.18)$$

где  $Z$  - вектор центрированных главных компонент.

В ситуации неработоспособности первой главной компоненты и полного отсутствия какого-либо экспертного «обучения» для построения интегральных индикаторов и ранжировок объектов используют взвешенную процедуру индивидуального рейтингования в пространстве модифицированных первых главных компонент. В этом случае:

1) определяется размерность многокритериального пространства, т.е. определяется  $m_0$  первых главных компонент из условия:

$$m_0 = \min \left\{ m : \frac{\lambda_1 + \dots + \lambda_m}{\lambda_1 + \dots + \lambda_k} \geq 0.55 \right\}; \quad (2.19)$$

2) осуществляется разбиение анализируемого набора унифицированных частных критериев (регистрируемых статистических показателей)  $\tilde{x}^{(1)}, \dots, \tilde{x}^{(k)}$  на  $m_0$  однородных групп  $M_1, \dots, M_{m_0}$ , где принадлежность частных критериев к одной такой группе  $M_j$  определяется относительно высоким уровнем взаимной коррелированности;

3) проводится построение модифицированных первых главных компонент отдельно по частным критериям, входящим в каждую из групп  $M_1, \dots, M_{m_0}$ . Так если  $\tilde{x}^{(1)}(M_j), \dots, \tilde{x}^{(k_j)}(M_j)$  – унифицированные признаки, вошедшие в группу  $M_j$  (всего  $m_0$  групп), то по наблюдениям  $\tilde{x}_i^{(1)}(M_j), \dots, \tilde{x}_i^{(k_j)}(M_j)$  ( $i = 1, 2, \dots, n$ ) строится оценка ковариационной матрицы и определяется наибольшее собственное число  $\lambda_1(M_j)$ . Затем рассчитывается собственный вектор  $l(M_j) = (l_1(M_j), \dots, l_{k_j}(M_j))^T$  и значения модифицированной первой главной компоненты частных критериев, вошедших в группу  $M_j$ :

$$\tilde{y}_i^{(j)} = \sum_{q=1}^{k_j} l_q^2(M_j) \tilde{x}_i^{(q)}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m_0. \quad (2.20)$$

Возможны ситуации, когда  $M_j$  состоит из единственного частного критерия  $\tilde{x}^{(1)}(M_j)$ . Тогда значения интегральной характеристики  $\tilde{y}_i^{(j)}$  определяются значениями этого частного критерия, т.е.

$$\tilde{y}_i^{(j)} = \tilde{x}_i^{(1)}, i = 1, 2, \dots, n; \quad (2.20(a))$$

4) на следующем этапе определяется удельная значимость («вес»  $\nu_j$ ) каждой из построенных интегральных характеристик  $\tilde{y}^{(1)}, \dots, \tilde{y}^{(m_0)}$ . Удельная значимость (вес  $\tilde{\nu}_j$ ) интегральной характеристики  $j - \tilde{y}^{(j)}$  определяется долей объясненной ею дисперсии в суммарной дисперсии всех частных критериев  $\tilde{x}^{(1)}, \dots, \tilde{x}^{(k)}$ .

В дальнейшем применяют нормированные значения  $\nu_j$  весов  $\tilde{\nu}_j$ , подсчитанные по формулам:

$$\nu_j = \frac{\tilde{\nu}_j}{\sum_{j=1}^{m_0} \tilde{\nu}_j} = \frac{s_{\tilde{y}^{(j)}}^2}{\sum_{j=1}^{m_0} s_{\tilde{y}^{(j)}}^2}, \quad (2.21)$$

что обеспечивает выполнение условия:  $\sum_{j=1}^{m_0} \nu_j = 1, \nu_j \geq 0, j = 1, 2, \dots, m_0$ .

После определения весов необходимо расставить ранги  $R$  каждому муниципальному образованию следующим образом:

– на первое место ставят ту территорию, взвешенное евклидово расстояние от которой до эталона (объекта, характеризующегося минимальными значениями показателей) оказалось максимальным (или минимальным, если за эталон принят объект, характеризующийся максимальными значениями показателей):



$$R_{i_0} = 1, \text{ если } \rho_{i_0} = \sqrt{\sum_{j=1}^{m_0} v_j (\tilde{y}_{i_0}^{(j)} - \tilde{y}^{(\ominus)})^2} = \max_{1 \leq i \leq n} \sqrt{\sum_{j=1}^{m_0} v_j (\tilde{y}_i^{(j)} - \tilde{y}^{(\ominus)})^2}, \quad (2.22)$$

где  $v_j$  – удельная значимость («вес»  $v_j$ ) каждой из модифицированной главной компоненты  $\tilde{y}^{(1)}, \dots, \tilde{y}^{(m_0)}$ ;

$m_0$  – число модифицированных первых главных компонент;

$\tilde{y}^{(\ominus)}$  – «наихудшие» значения, принадлежащие объекту-эталону;

– на второе место следует поставить ту территорию, взвешенное евклидово расстояние от которого до эталона оказалось наибольшим среди всех остальных территорий за исключением  $i_0$ , т.е.

$$R_{i_1} = 2 \text{ если } \rho_{i_1} = \sqrt{\sum_{j=1}^{m_0} v_j (\tilde{y}_{i_1}^{(j)} - \tilde{y}^{(\ominus)})^2} = \max_{\substack{1 \leq i \leq n \\ i \neq i_0}} \sqrt{\sum_{j=1}^{m_0} v_j (\tilde{y}_i^{(j)} - \tilde{y}^{(\ominus)})^2}, \quad (2.22(a))$$

и т.д.

Другой подход, который используется в ситуациях, когда исследователь имеет принципиальную возможность получить с помощью экспертов так называемые «обучающие» или, хотя бы: «частично обучающие» выборки: т.е. наряду со значениями  $x_i^{(1)}, \dots, x_i^{(k)}$  показателей (частных критериев) апостериорного набора он имеет некоторую экспертную информацию и об оценках искомого интегрального свойства  $y_i$ . В этом случае построение интегрального индикатора осуществляется экспертно-статистическим методом [8]. Согласно этому методу рассматривают интегральное свойство как сводную характеристику  $f(\bar{X})$  – определяемую поддающимися учету и измерению признаками  $\bar{X}_i = (x_i^{(1)}, \dots, x_i^{(k)})$ . Сводная характеристика  $f(\bar{X})$  является латентной, то есть не поддающейся непосредственному количественному измерению. Интуитивное экспертное восприятие сводной характеристики можно рассмотреть в виде следующей модели латентного показателя (2.23):

$$y_i = f(\bar{X}_i) + \delta(\bar{X}_i), \quad i = 1, 2, \dots, n, \quad (2.23)$$

где  $\delta(\bar{X}_i)$  - искажения, носящие случайный характер, обусловленные как индивидуальностью эксперта, так и влиянием неучтенных входных показателей, оказывающих несущественное влияние на  $y$ .

Относительно остаточной случайной компоненты  $\delta(\bar{X})$  предполагается:  $M\delta(\bar{X}_i) = 0$ ;  $D\delta(\bar{X}_i) = \text{deg}^2(\bar{X}_i) < \infty$  [6].

Модель (2.22) интерпретируется как регрессионная модель  $y$  на  $\bar{X}$  с той спецификой, что вместо наблюдаемых значений  $y$ , необходимых для оценки  $f(\bar{X})$ , используются экспертные оценки  $y$ . Поэтому вместо обычной задачи регрессионного анализа – восстановления функции, ставится задача оценки  $f(\bar{X})$  с точностью до произвольного монотонного преобразования. Обычно рассматривают параметрическую модель:

$$y_i = \sum_{j=1}^k \theta_j x_i^{(j)} + \delta(\bar{X}_i), \quad (2.24)$$

где  $i = 1, 2, \dots, n$ ;  $j = 1, 2, \dots, k$ .

Исходные данные для решения задачи состоят из 2 частей:

1. экспертная часть, представленная в виде формы булевой матрицы попарных сравнений:

$$\gamma_m = \{\gamma_{ij}^{(m)}\}, \quad (2.25)$$

где  $i = \overline{1, n}$ ;  $j = \overline{1, n}$ ;

$$\gamma_{ij}^{(m)} = \begin{cases} 1, & \text{если } O_i \text{ и } O_j \text{ принадлежат одной группе объектов} \\ 0, & \text{в противном случае;} \end{cases}$$

$n$  – число объектов;

$m$  – показывает, каким экспертом предложена матрица  $\gamma_m$ ,  $m = \overline{1, l}$ .

$l$  – число экспертов;

2. статистическая часть исходных данных о показателях, представленная в виде матрицы  $X$  типа «объект-свойство».

Определив меру близости для любой пары объектов и построив разбиение объектов на однородные классы с помощью функции  $f(\bar{X}, \bar{\Theta})$  ( $\gamma_f$ - разбиение) параметры  $\bar{\Theta}$  оценивают, решая оптимизационную задачу о минимизации расстояний между разбиениями экспертов и разбиением  $\gamma_f$  [7].

На практике нередки ситуации, когда оперативно получить матрицы  $\gamma_m$  невозможно. В связи с этим матрицы парных сравнений предлагается получать на основании предварительного разбиения объектов исследования на классы с использованием методов кластерного анализа, классификации методами нейронных сетей, Парето-классификации [9].

#### **2.4 Территориальное зонирование города Оренбурга на основе многомерного статистического анализа факторов, определяющих цену жилья на вторичном рынке**

При решении различных задач анализа рынка жилья, например оценки стоимости жилья, требуется определять показатели различной степени дифференциации и обобщения. Для этого вся совокупность объектов жилой недвижимости на рынке (например, совокупность одно-, двух- или трехкомнатных квартир) должна быть разделена на группы (типы, классы) по качеству, размеру, месторасположению, а иногда и по другим типологическим признакам (например, для новостроек - по степени готовности в момент продажи). Группировка объектов жилой недвижимости по различным основаниям может происходить с различной степенью укрупнения и детализации. Так разделение по

месторасположению может осуществляться по укрупненным или детализированным зонам. Разделение может быть доведено до одного здания, и рассчитанная средняя по зданию удельная цена квартир представляет большой интерес для анализа динамики цен на объекты данного типа на данной территории. Аналогично необходимо поступать и при типизации объектов на рынке жилья по другим признакам. Совокупность объектов жилой недвижимости необходимо разделять по признакам качества на типы, классы (вплоть до единичного здания) и определять как среднюю удельную цену, так и иные показатели для каждого типа объектов. Так, например, средняя удельная цена однокомнатных, двухкомнатных и т.д. квартир одного типа в одном районе может различаться, и, соответственно, требуется дифференцированное определение средних показателей для этих групп объектов.

Таким образом, должна быть проведена типизация, то есть выделение однородных групп объектов, имеющих как можно меньшие различия по типологическим характеристикам.

Наиболее распространенная в риэлторской практике методика территориального зонирования – методика, использующая районную сетку города. Применение данной методики для г. Оренбурга показало, что районы города являются крупными территориальными ячейками (территория г.Оренбурга разделена на четыре района), чтобы по ним можно детально судить о реальных внутригородских различиях (таблица 2.1). Кроме того, районы весьма неравноценны по своему потенциалу. Численность населения в них варьирует от 101,2 тыс. чел. в Центральном районе и до 169,0 тыс. чел. в Ленинском. Поэтому если рассматривать район в его административных границах, то диапазон разброса цен на квартиры на вторичном рынке жилья г. Оренбурга становится чрезвычайно большим. В связи с этим территориальное зонирование города по комплексу показателей, характеризующих объекты жилой недвижимости (например, цене, площади, месторасположению) необходимо проводить с помощью методов многомерной классификации [10].

Таблица 2.1 – Вариация цен квартир, предлагаемых на продажу на вторичном рынке жилья г. Оренбурга в январе-июне 2005г., по районам, тыс. руб.

Показатели	Район			
	Дзержинский	Ленинский	Центральный	Промышлен- ный
Средняя цена на квартиры в целом по рынку, в том числе по типу квартир	915,6	948,5	1025,7	891,6
<b>однокомнатные</b>				
средняя	722,1	691,3	751,1	676,5
минимальная	500,0	500,0	590,0	400,0
максимальная	900,0	850,0	955,0	800,0
размах вариации	400,0	550,0	365,0	400,0
коэффициент вариации, %	34,12	36,71	28,32	33,96
<b>двухкомнатные</b>				
средняя	929,4	983,8	921,0	876,3
минимальная	760,0	690,0	600,0	640,0
максимальная	1300,0	1250,0	925,0	980,0
размах вариации	540,0	560,0	325,0	340,0
коэффициент вариации, %	42,82	37,45	24,56	26,78
<b>трехкомнатные</b>				
средняя	1095,5	1170,4	1405,0	1122,0
минимальная	910,0	650,0	720,0	730,0
максимальная	1430,0	1550,0	1350,0	1150,0
размах вариации	520,0	900,0	630,0	420,0
коэффициент вариации, %	39,12	45,62	32,48	25,36

Одной из проблем при проведении кластерного анализа с целью выделения однородных групп (кластеров), является выбор расстояния между исследуемыми объектами, так как от этого зависит состав и количество формируемых кластеров, а также степень сходства объектов внутри кластеров. В статистической литературе [3, 11-13] подробно рассмотрены и изложены основные виды расстояний для совокупности объектов, описываемых или количественными, или только качественными признаками, то есть измеренных в одной шкале. Поскольку объекты жилой недвижимости (квартиры, дома, коттеджи) характеризуются большим набором не только количественных, но и качественных показателей, то возникает необходимость приведения показателей к единой шкале измерения. В литературе предлагается два подхода к решению

этой проблемы [11, 13]. Первый подход заключается в том, что  $i$  – многомерное наблюдение представляется вектор-столбцом числа градаций признака  $x_k$ , причем компонентами этого вектора могут быть только нули и единицы. При таком подходе к достижению единообразия записи наблюдений многомерного признака смешанной природы вынуждены мириться, во-первых, с элементами субъективизма в выборе способов разбиения диапазонов изменения количественных признаков на интервалы группирования и, во-вторых, с определенной потерей информативности исходных данных, связанной с переходом от индивидуальных к группированным признакам по количественным переменным. В качестве альтернативного подхода может быть использована идея о том, что, руководствуясь некоторыми дополнительными соображениями (и допущениями), исследователь пытается преобразовать качественные и классификационные признаки в количественные, используя процесс так называемой «оцифровки», или шкалирования, неколичественных переменных [11]. «Оцифровка» качественного показателя (месторасположение дома, в котором находится квартира) осуществлялась различными методами неметрического шкалирования с применением ППП Statistica .

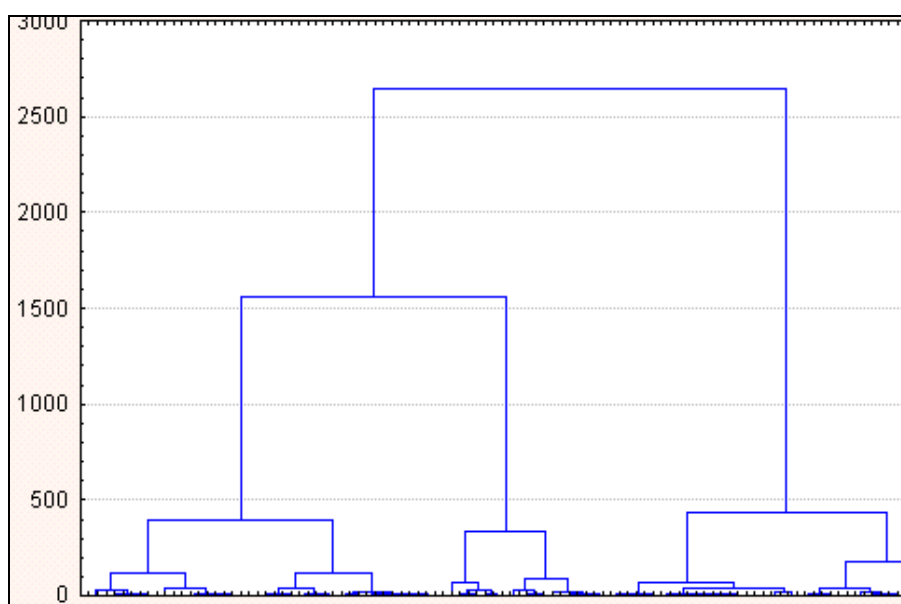


Рисунок 2.1 – Дендрограмма классификации однокомнатных квартир, предлагаемых на продажу на вторичном рынке жилья г. Оренбурга

В результате систематизации теоретических положений по проведению многомерной классификации пришли к выводу, что для решения поставленной задачи и достижения цели на первом этапе целесообразно использовать иерархические кластер-процедуры, в качестве правила объединения – метод Уорда, а в качестве меры расстояния – евклидово расстояние. Результаты классификации иерархическими методами представлены в виде дендограммы на рисунке 2.1.

Для уточнения разбиения объектов использовался метод k-средних. Для оценки качества классификации сопоставлялись расстояния от каждого объекта кластера до его центра, и использовался функционал качества – минимум внутриклассовой дисперсии. В таблице 2.2 представлена группировка квартир по выделенным классам, средние значения показателей, коэффициент вариации цен.

Таблица 2.2 – Результаты кластерного анализа по совокупности квартир, предлагаемых на продажу на вторичном рынке жилья г. Оренбурга

Тип кв-ры	Показатель	Номер кластера							
		1	2	3	4	5	6	7	
	коэффициент зонирования	0,1	0,2	0,3	0,4	0,5	0,6	0,7	
однокомнатные	средняя цена, тыс. руб.	480,83	406,30	384,61	365,63	347,27	326,15	302,86	
	средняя пл-дь, м <sup>2</sup>	общая	35,39	34,39	32,74	30,81	32,52	29,40	29,67
		жилая	18,29	18,06	17,86	16,44	17,40	16,15	16,21
		кухни	8,32	7,83	7,54	7,51	7,19	7,43	6,56
	K <sub>v</sub> , %	6,22	2,11	1,45	1,36	1,66	1,67	2,99	
	число объектов	18	23	13	8	22	13	7	
двухкомнатные	средняя цена, тыс. руб.	648,33	583,13	532,94	498,46	478,64	446,19	382,64	
	средняя пл-дь, м <sup>2</sup>	общая	51,02	50,23	50,89	46,62	50,11	44,59	44,17
		жилая	30,77	30,08	30,48	30,81	30,86	29,94	29,37
		кухни	8,33	7,83	8,38	7,86	8,19	6,19	6,74
	K <sub>v</sub> , %	4,49	2,50	2,37	1,55	0,94	2,77	2,89	
	число объектов	24	16	17	11	13	21	20	
трехкомнатные	средняя цена, тыс. руб.	936,67	831,82	755,38	726,67	690,77	628,57	499,46	
	средняя пл-дь, м <sup>2</sup>	общая	67,79	64,33	66,62	65,05	65,82	62,33	55,03
		жилая	42,84	43,80	43,28	42,62	42,92	42,28	38,56
		кухни	10,17	8,15	8,13	8,58	8,63	7,57	6,54
	K <sub>v</sub> , %	7,00	2,89	1,39	1,82	1,42	3,74	7,63	
	число объектов	6	11	13	6	13	21	26	

Наилучшие в содержательном плане результаты показала классификация квартир, предлагаемых на продажу, на семь кластеров (рисунок 2.2)

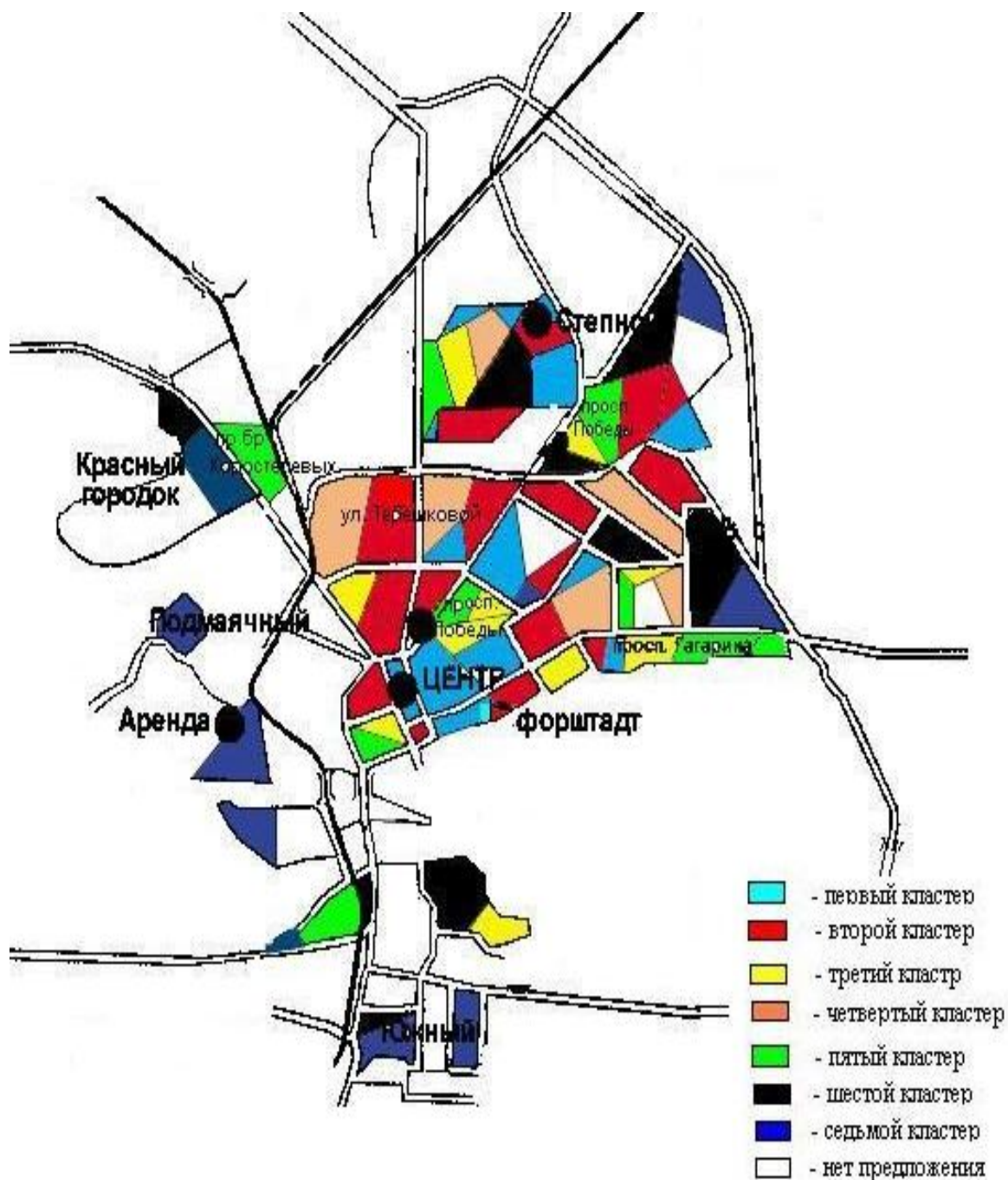


Рисунок 2.2 – Территориальное зонирование г. Оренбурга по всей совокупности квартир, предлагаемых на продажу на вторичном рынке жилья

Сравнение динамики цен квартир по выделенным зонам показало, что стоимость жилья в первой зоне значительно превышала стоимость по другим



группам: 665,0 тыс. руб. для однокомнатных квартир, 914,7 тыс. руб. – двухкомнатных и 1177,6 тыс. руб. – трехкомнатных.

Первая зона (кластер) – объекты расположены в непосредственной близости от центра и престижных районов северо-восточной части города. Высокий уровень цен на эту категорию жилья сложился за счет предложений к продаже квартир во вновь построенных домах последнего периода, успевших прийти на вторичный рынок жилья.

Объекты второй зоны (кластера) имеют четко выраженную характерную вытянутость с юго-востока на северо-восток, высокий уровень цен на квартиры. Кроме того, показатели площади объектов находятся на высоком уровне: жилая площадь в среднем составила 19,0 м<sup>2</sup> для однокомнатных квартир, 31,3 м<sup>2</sup> – двухкомнатных и 44,5 – трехкомнатных.

Третью зону (кластер) представляют объекты, располагающиеся в различных частях города. Картина зоны носит менее определенный, более размытый и фрагментарный характер, объекты обладают более низкой стоимостью по сравнению с предыдущими зонами, но хорошими потребительскими свойствами.

Четвертая зона (кластер) – это объекты, расположенные в районах наиболее удаленных от центра, обладают средними показателями общей, жилой площади и площади кухни, но и имеют небольшое отличие в цене по сравнению с объектами первых трех зон.

Объекты пятой зоны (кластера) сильно разбросаны по территории города, характеризуются средними показателями размера квартиры. Кроме того, уровень цен объектов ниже среднегородского и составил в среднем 488,9 тыс. руб. для однокомнатных квартир, 654,2 тыс. руб. – двухкомнатных и 783,9 тыс. руб. – трехкомнатных.

Шестую зону (кластер) представляют объекты, расположенные в северо-восточной части города старой застройки; средняя цена ниже среднегородского уровня, низкие потребительские свойства объектов (жилая площадь в среднем

составила 17,3 м<sup>2</sup> для однокомнатных квартир, 30,0 м<sup>2</sup> – двухкомнатных и 40,0 – трехкомнатных).

Самая низкая цена за весь анализируемый период наблюдалась в седьмой зоне: 394,1 тыс. руб. для однокомнатных квартир, 541,3 тыс. руб. – двухкомнатных и 677,6 тыс. руб. – трехкомнатных. Объекты зоны занимают периферийную территорию города Оренбурга, как бы “обнимают” более дорогие районы города полукольцом. При этом на севере эта зона наиболее близко подходит к центру города. Все эти районы относятся к числу достаточно непрестижных районов, поэтому их попадание в замыкающую группу не вызывает удивления.

Результаты исследования показали, что внутри каждого кластера наблюдалась незначительная вариация цен на квартиры. Коэффициент вариации цен изменился за анализируемый период с 0,58% до 9,55%. Следовательно, можно сделать вывод о том, что территориальное зонирование города методами многомерной классификации дает лучшие результаты по сравнению с применяемыми методиками.

Результаты полученной классификации использовались для построения обучающих выборок с целью проведения дискриминантного анализа, для последующего включения квартир в кластеры. Реализация процедуры дискриминантного анализа осуществлялась с применением ППП Statistica. В итоге получены следующие результаты:

- оценены коэффициенты дискриминантной функции, обеспечивающей отнесение объектов к данному классу, отдельно для каждого класса (таблица 2.3);
- получена таблица, где для каждого объекта указывается уровень значимости нулевой гипотезы о том, что объект может быть отнесен к данному классу (если уровень значимости больше 0,05, то нулевую гипотезу принимаем) и вероятность отнесения объекта к этому классу.

По мере поступления дополнительной информации, рекомендуется производить пересчет коэффициентов дискриминантных функций. Дискрими-

нантный анализ можно также использовать и как метод прогнозирования (предсказания) поведения наблюдаемых единиц статистической совокупности на основе имеющихся стереотипов поведения аналогичных объектов, входящих в состав объективно существующих или сформированных по определенному принципу множеств (обучающих выборок).

Таблица 2.3 – Линейные дискриминантные функции для совокупности однокомнатных, двухкомнатных, трехкомнатных квартир, предлагаемых на продажу на вторичном рынке жилья г. Оренбурга

Тип квартиры	Номер класса	Линейные дискриминантные функции
однокомнатные	1	$f_1(\bar{x}) = -90,94 + 0,20x_1 - 0,52x_2 + 4,0x_3 + 4,14x_4 - 11,3x_5$
	2	$f_2(\bar{x}) = -71,78 + 0,16x_1 - 0,04x_2 + 3,16x_3 + 3,60x_4 - 9,26x_5$
	3	$f_3(\bar{x}) = -77,07 + 0,15x_1 - 0,19x_2 + 4,04x_3 + 3,59x_4 - 8,59x_5$
	4	$f_4(\bar{x}) = -80,36 + 0,15x_1 - 0,07x_2 + 3,94x_3 + 3,68x_4 - 8,62x_5$
	5	$f_5(\bar{x}) = -84,05 + 0,17x_1 - 0,04x_2 + 3,68x_3 + 3,61x_4 - 8,44x_5$
	6	$f_6(\bar{x}) = -79,82 + 0,15x_1 + 0,10x_2 + 3,79x_3 + 3,71x_4 - 10,46x_5$
	7	$f_7(\bar{x}) = -71,00 + 0,14x_1 - 0,21x_2 + 4,06x_3 + 3,59x_4 - 9,50x_5$
двухкомнатные	1	$f_1(\bar{x}) = -953,3 + 2,62x_1 - 0,76x_2 + 4,63x_3 + 2,20x_4 - 3,56x_5$
	2	$f_2(\bar{x}) = -794,2 + 2,37x_1 + 1,04x_2 + 4,27x_3 + 1,56x_4 - 5,34x_5$
	3	$f_3(\bar{x}) = -691,0 + 2,18x_1 + 1,01x_2 + 4,25x_3 + 1,92x_4 - 6,12x_5$
	4	$f_4(\bar{x}) = -618,0 + 2,04x_1 + 1,11x_2 + 4,19x_3 + 1,99x_4 - 8,65x_5$
	5	$f_5(\bar{x}) = -574,7 + 1,96x_1 + 1,02x_2 + 4,28x_3 + 1,33x_4 - 7,13x_5$
	6	$f_6(\bar{x}) = -504,5 + 1,83x_1 + 1,11x_2 + 4,09x_3 + 0,80x_4 - 6,99x_5$
	7	$f_7(\bar{x}) = -436,2 + 1,66x_1 + 1,09x_2 + 4,14x_3 + 0,19x_4 - 6,94x_5$
трехкомнатные	1	$f_1(\bar{x}) = -66,13 + 0,06x_1 - 0,40x_2 + 2,43x_3 - 0,60x_4 + 4,04x_5$
	2	$f_2(\bar{x}) = -61,01 + 0,05x_1 - 0,28x_2 + 2,32x_3 - 0,75x_4 + 4,85x_5$
	3	$f_3(\bar{x}) = -58,3 + 0,15x_1 - 0,19x_2 + 4,04x_3 + 3,59x_4 - 8,59x_5$
	4	$f_4(\bar{x}) = -53,7 + 0,05x_1 - 0,43x_2 + 2,45x_3 - 0,44x_4 + 4,10x_5$
	5	$f_5(\bar{x}) = -52,73 + 0,039x_1 - 0,22x_2 + 2,32x_3 - 0,93x_4 + 2,15x_5$
	6	$f_6(\bar{x}) = -43,79 + 0,04x_1 - 0,17x_2 + 1,89x_3 - 0,321x_4 + 0,75x_5$
	7	$f_7(\bar{x}) = -31,00 + 0,14x_1 - 0,21x_2 + 1,06x_3 - 0,59x_4 + 0,50x_5$

где  $x_1$  – цена квартиры, тыс. руб.;  $x_2$  – общая площадь, кв. м;  $x_3$  – жилая площадь, кв. м;  $x_4$  – площадь кухни, кв. м.;  $x_5$  – месторасположение дома, в котором находится квартира.

## **2.5 Многомерная классификация муниципальных образований Оренбургской области по состоянию показателей, характеризующих экономическую и демографическую безопасность как основных составляющих национальной безопасности**

Особенности экономического, демографического и социального развития городов и районов области обуславливают межтерриториальные различия и указывают на региональную специфику экономической ситуации. С этих позиций представляет интерес изучение территориальной дифференциации муниципальных образований области по состоянию показателей, характеризующих экономическую и демографическую безопасность.

Будем характеризовать «экономическую безопасность» набором показателей, измеренных по городам и районам Оренбургской области:

$X_1$  – число зарегистрированных иностранных работников на 1000 человек населения;

$X_2$  – уровень безработицы (%);

$X_3$  – цепной темп роста числа предприятий и организаций по основным отраслям экономики (%);

$X_4$  – цепной темп роста среднесписочной численности работников (%);

$X_5$  – кредиторская задолженность на душу населения (тыс. руб.);

$X_6$  – оборот розничной торговли на душу населения, (руб.);

$X_7$  – средненоминальная заработная плата работников (руб.);

$X_8$  – инвестиции в основной капитал на душу населения (руб.);

$X_9$  – цепной темп роста численности населения в трудоспособном возрасте (%);

$X_{10}$  – удельный вес убыточных предприятий и организаций (в % от общего числа организаций и предприятий).

С целью выделения однородных групп, проведена многомерная классификация муниципальных образований Оренбургской области методами кластерного анализа. Расстояние между объектами рассчитывалось по формуле [11]:

$$d_E(O_i, O_j) = \sqrt{\sum_{l=1}^k (x_{il} - x_{jl})^2}, \quad (2.26)$$

где  $x_{il}$ ,  $x_{jl}$  - величина  $l$ -й компоненты у  $i$ -го ( $j$ -го) объекта ( $l = 1, 2, \dots, k$ ;  $i, j = 1, 2, \dots, n$ ).

Использование обычного евклидова расстояния оправдано в случаях, если: наблюдения берутся из генеральных совокупностей, имеющих многомерное нормальное распределение; компоненты вектора  $X$  однородны по физическому смыслу и одинаково важны для классификации [13].

Дендрограмма классификации муниципальных образований области методом Уорда, представлена на рисунке 2.3.

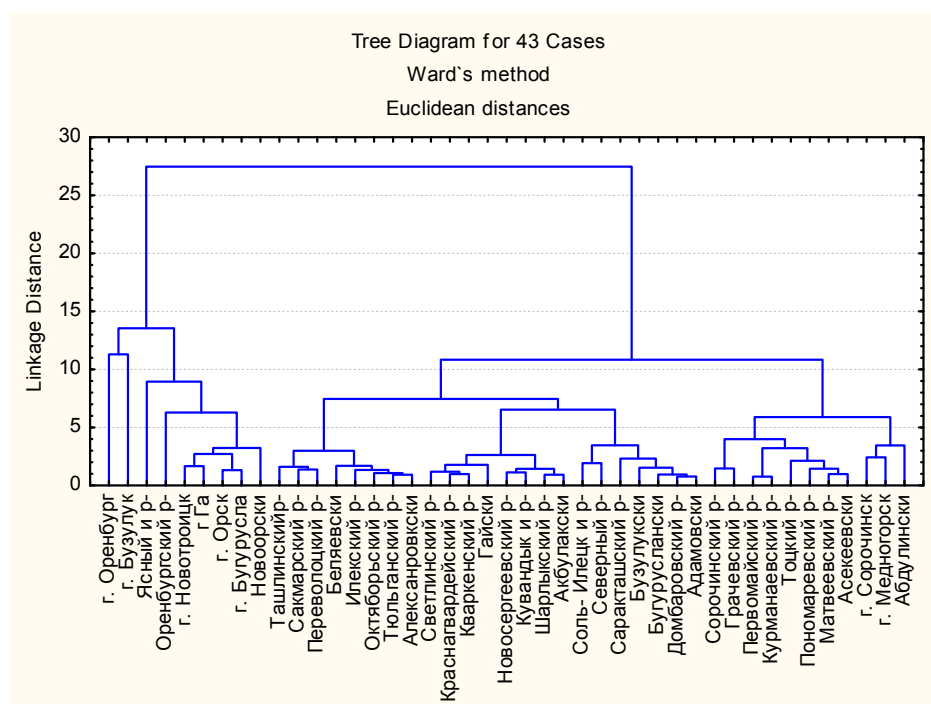


Рисунок 2.3 – Дендрограмма классификации городов и районов Оренбургской области по показателям, характеризующим экономическую безопасность региона, методом Уорда в 2007 году

На основе полученной дендрограммы можно выделить два кластера. Уточнение состава классов проведено с помощью метода k-средних, который дает лучшее разбиение по функционалу качества [14]. Разбиение муниципальных образований на большее число классов характеризуется несущественными различиями средних значений ряда признаков для некоторых классов. Результаты классификации методом k-средних приведены в таблице 2.4.

Таблица 2.4 – Классификация городов и районов Оренбургской области по показателям, характеризующим экономическую безопасность региона, методом k-средних в 2007 году

Номер кластера	Количество объектов в кластере	Состав класса
1	6	<b>Города:</b> Бузулук, Гай, Оренбург, Новотроицк, Орск. <b>Районы:</b> Оренбургский
2	37	<b>Города:</b> Бугуруслан, Медногорск, Сорочинск. <b>Районы:</b> Абдулинский, Адамовский, Акбулакский, Александровский, Асекеевский, Беляевский, Бугурусланский, Бузулукский, Гайский, Грачевский, Домбаровский, Илекский, Кваркенский, Красногвардейский, Кувандыкский, Курманаевский, Матвеевский, Новоорский, Новосергеевский, Октябрьский, Первомайский, Переволоцкий, Пономаревский, Сакмарский, Саракташский, Светленский, Северный, Соль-Илецкий, Сорочинский, Ташлинский, Тоцкий, Тюльганский, Шарлыкский, Ясенский.

График средних значений центрировано-нормированных признаков в классах приведен на рисунке 2.4.

В первый кластер вошли крупные города области: Бузулук, Гай, Новотроицк, Оренбург, Орск и Оренбургский район. Вхождение Оренбургского района в первый класс, объясняется его близостью к областному центру.

Для муниципальных образований, вошедших в первый класс характерны более высокие средние значения показателей числа зарегистрированных иностранных работников ( $\bar{x}^{(1)}$ ), темпа роста среднесписочной численности работников ( $\bar{x}^{(4)}$ ), кредиторской задолженности на душу населения ( $\bar{x}^{(5)}$ ), оборота розничной торговли на душу населения ( $\bar{x}^{(6)}$ ), средненоминальной заработной платы работников ( $\bar{x}^{(7)}$ ) и инвестиций в основной капитал на душу населения

( $\bar{x}^{(8)}$ ). Это объясняется тем, что объекты, вошедшие в первый класс, являются местами сосредоточения экономических и финансовых ресурсов. В частности, в города области прибывает население в трудоспособном возрасте, в том числе иностранные рабочие, привлеченные сравнительно более высоким заработком и большей возможностью для трудоустройства, что положительно отражается на уровне экономической безопасности муниципальных образований, вошедших в первый класс. Следовательно, муниципальные образования, вошедшие в первый кластер, можно охарактеризовать как объекты с более высоким уровнем экономической безопасности, по сравнению с уровнем экономической безопасности муниципальных образований, вошедших во второй класс.

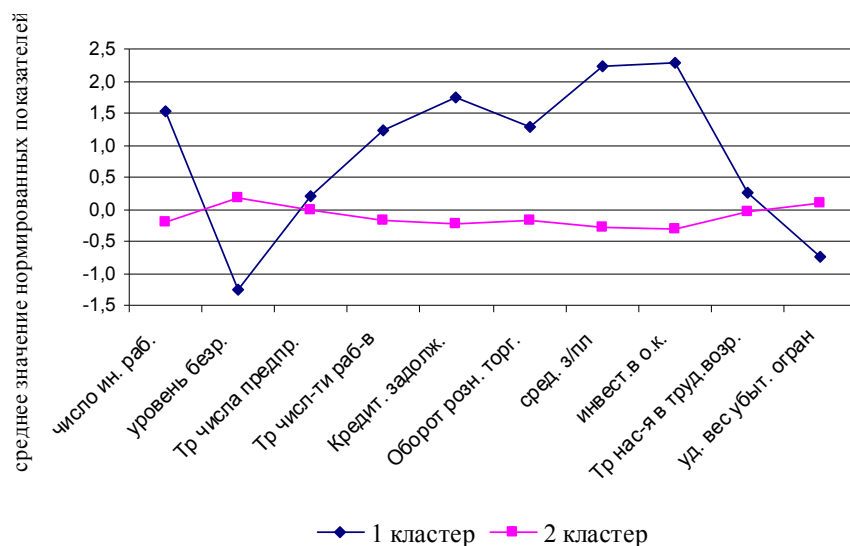


Рисунок 2.4 – График средних значений показателей, характеризующих экономическую безопасность городов и районов Оренбургской области в 2007 г.

Результаты классификации муниципальных образований Оренбургской области по показателям, характеризующим экономическую безопасность региона, за 2004-2006 гг. приведены в приложении А и свидетельствуют об отсутствии динамики в составе классов.

Таким образом, методами многомерной классификации было выделено два кластера, характеризующихся разными уровнями экономической безопасности.

Демографическая безопасность в современных условиях является важнейшей составной частью общенациональной безопасности страны. Существующие внешние и внутренние угрозы политического, экономического, социального, военного, экологического, техногенного и иного характера на демографическую безопасность без применения соответствующих упреждающих мер способны оказать дестабилизирующее воздействие на состояние защищенности демографических процессов [15].

Различия в экономическом и социальном развитии городов и районов области, разнообразие природно-климатических условий, экологической обстановки обуславливают интерес к изучению территориальной дифференциации административно-территориальных образований области по состоянию показателей, характеризующих демографическую безопасность [16]. Ставится задача разбиения административно-территориальных образований Оренбургской области на однородные группы по сформированному в работах [17, 18, 19] апостериорному набору из 23 показателей интегрального свойства демографической безопасности (таблица 2.5).

Таблица 2.5 – Апостериорные (редуцированные) наборы показателей интегрального свойства демографической безопасности

№ п/п	Код показателя	Наименование показателя
1	2	3
<b>1 Показатели, характеризующие качество населения</b>		
<b>1.1 Показатели воспроизводства населения</b>		
1	1.1.1	общий коэффициент рождаемости (на 1000 человек)
2	1.1.2	общий коэффициент смертности (на 1000 человек)
3	1.1.3	коэффициент младенческой смертности (на 1000 родившихся живыми)
4	1.1.4	уровень брачности населения (на 1000 человек)
5	1.1.5	уровень разводимости населения (на 1000 человек)
6	1.1.6	уровень внешней миграции (на 1000 человек)
<b>1.2 Показатели демографической структуры</b>		
7	1.2.1	удельный вес населения в трудоспособном возрасте (%)
8	1.2.3	соотношение мужчин и женщин (на 1000 мужчин приходится женщин)
<b>1.3 Показатели заболеваемости и состояния системы здравоохранения</b>		
9	1.3.1	показатель общей заболеваемости (на 1000 человек)
10	1.3.6	болезни системы кровообращения (на 1000 человек)



Продолжение таблицы 2.5

1	2	3
11	1.3.12	обеспеченность населения больничными койками (на 10000 человек)
<b>2 Показатели, характеризующие уровень благосостояния населения</b>		
12	2.2	соотношение среднего размера назначенных месячных пенсий с величиной прожиточного минимума (%)
13	2.3	площадь жилищ, приходящаяся в среднем на одного жителя (квадратных метров)
14	2.8	удельный вес учащихся государственных дневных общеобразовательных учреждений, занимающихся во вторую и третью смены (%)
<b>3 Показатели, характеризующие экономическое развитие</b>		
15	3.3	оборот розничной торговли на душу населения (рублей)
16	3.5	уровень официально зарегистрированной безработицы (%)
17	3.8	инвестиции в основной капитал на душу населения (рублей)
<b>4 Показатели, характеризующие риск в социальной сфере</b>		
18	4.1	численность лиц, состоящих на диспансерном учете по поводу алкоголизма и алкогольных психозов (на 10000 человек)
19	4.2	численность лиц, состоящих на диспансерном учете по поводу наркомании (на 10000 человек)
20	4.3	число пострадавших с утратой трудоспособности на один рабочий день и более и со смертельным исходом (на 1000 работающих)
21	4.4	травмы и отравления (на 1000 человек)
<b>5 Показатели, характеризующие качество окружающей среды</b>		
22	5.1	использование свежей воды ( $m^3 / z$ )
23	5.2	выбросы загрязняющих веществ в атмосферный воздух от стационарных источников (тысяч тонн/ $км^2$ )

Классификации подверглись 12 городов и 35 районов Оренбургской области. Результаты классификации методов Уорда представлены в виде дендрограммы (рисунок 2.5).

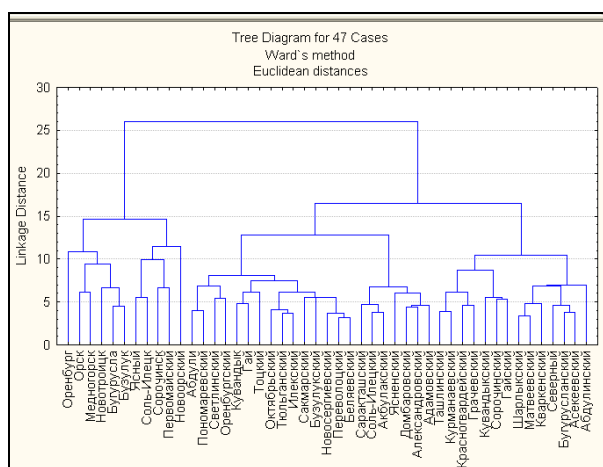


Рисунок 2.5 – Дендрограмма классификации городов и районов Оренбургской области по показателям демографической безопасности методом Уорда в 2007 году

На основе полученной дендрограммы можно выделить три кластера. Уточнение состава классов проведено с помощью метода k-средних. Результаты классификации методом k-средних для 2007 года приведены в таблице 2.6.

Таблица 2.6 – Классификация городов и районов Оренбургской области по показателям демографической безопасности методом k-средних в 2007 году

Номер кластера	Количество объектов в кластере	Состав класса
кластер 1	13	<b>Города:</b> Абдулино, Бугуруслан, Бузулук, Гай, Кувандык, Медногорск, Новотроицк, Оренбург, Орск, Соль-Илецк, Сорочинск, Ясный. <b>Район:</b> Оренбургский.
кластер 2	15	<b>Районы:</b> Абдулинский, Асекеевский, Бугурусланский, Бузулукский, Грачевский, Кваркенский, Курманаевский, Матвеевский, Новоорский, Новосергиевский, Первомайский, Пономаревский, Северный, Сорочинский, Шарлыкский.
кластер 3	19	<b>Районы:</b> Адамовский, Акбулакский, Александровский, Беляевский, Гайский, Домбаровский, Илекский, Красногвардейский, Кувандыкский, Октябрьский, Переволоцкий, Сакмарский, Саракташский, Светлинский, Соль-Илецкий, Ташлинский, Тоцкий, Тюльганский, Ясенский.

Суммарная внутриклассовая дисперсия для приведенной в таблице 2.5.3 классификации составила  $F = 62,19$ .

График средних значений центрировано-нормированных признаков приведен на рисунке 2.6.

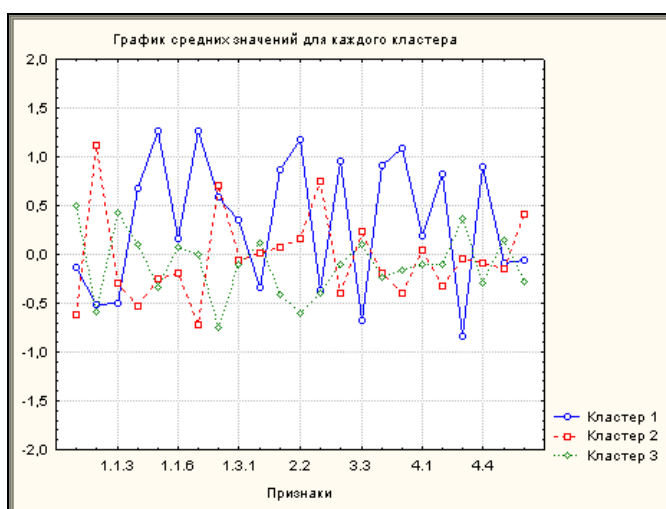


Рисунок 2.6 – График средних значений признаков в классах для 2007 года

Как видно из рисунка 2.6, первый класс характеризуется более высокими средними значениями большинства показателей, составляющих среду и систему обеспечения демографической безопасности. Исключением являются такие, как: уровень разводимости, коэффициент миграционного прироста, общая заболеваемость, охват детей дошкольными учреждениями, удельный вес учащихся государственных дневных общеобразовательных учреждений, занимающихся во вторую и третью смены, уровень безработицы, численность лиц, состоящих на диспансерном учете по поводу наркомании. Наиболее типичным объектом (то есть самым близким к центру кластера) является г. Гай. Наиболее нетипичным объектом данного кластера является г. Оренбург.

Объекты второго класса по сравнению с третьим характеризуются более высокими средними значениями таких показателей как, болезни системы кровообращения, площадь жилищ, приходящаяся в среднем на одного жителя, благоустройство жилищного фонда газом, обеспеченность населения собственными легковыми автомобилями, число пострадавших с утратой трудоспособности на один рабочий день и более и со смертельным исходом. Наиболее близок к центру второго кластера Матвеевский район. Наиболее нетипичным объектом данного кластера является Бузулукский район.

Третий класс лидирует лишь по одному значению показателя - общий коэффициент рождаемости. Однако остальные показатели, формирующие демографическую безопасность, оказались значительно ниже, чем в остальных классах. Наиболее близок к центру третьего кластера Беляевский район. Наиболее нетипичным объектом данного кластера является Гайский район.

Далее проведено разбиение объектов на однородные группы методами нейросетевой классификации. В качестве функционала качества разбиения рассмотрим суммарную внутриклассовую дисперсию. Чтобы правильно распределить плотность ядер классов (векторов весов) в соответствии с плотностью входных векторов в пространстве  $X$  для обучения сети будем применять метод выпуклой комбинации. Чтобы определить класс, к которому относится объект, нужно выбрать среди всех нейронов данного слоя один с максималь-

ным выходом. Рассмотренная сеть нейронов, использующая евклидову меру близости для классификации объектов, называется сетью Кохонена. Результаты классификации методом нейросетевой классификации для 2007 года представлены в таблице 2.7.

Таблица 2.7 – Классификация городов и районов Оренбургской области по показателям демографической безопасности с помощью сети Кохонена в 2007 году

Номер кластера	Количество объектов в кластере	Состав класса
кластер 1	14	<b>Города:</b> Абдулино, Бузулук, Бугуруслан, Гай, Кувандык, Медногорск, Новотроицк, Оренбург, Орск, Сорочинск, Соль-Илецк, Ясный. <b>Районы:</b> Новоорский, Оренбургский.
кластер 2	16	<b>Районы:</b> Абдулинский, Асекеевский, Беляевский, Бугурусланский, Гайский, Грачевский, Илекский, Кваркенский, Кувандыкский, Курманаевский, Матвеевский, Пономаревский, Саракташский, Северный, Соль-Илецкий, Сорочинский.
кластер 3	17	<b>Районы:</b> Адамовский, Акбулакский, Александровский, Бузулукский, Домбаровский, Красногвардейский, Новосергиевский, Октябрьский, Первомайский, Переволоцкий, Сакмарский, Светлинский, Ташлинский, Тоцкий, Тюльганский, Шарлыкский, Ясенский.

Суммарная внутриклассовая дисперсия для приведенной в таблице 2.5.4 классификации составила  $F = 58,88$ .

По сравнению с разбиением методом k-средних произошли следующие изменения в структуре классов: в первый класс перешел Новоорский район, из третьего класса во второй перешли Беляевский, Гайский, Кувандыкский и Саракташский районы. Из второго класса в третий перешли Бузулукский, Первомайский районы.

Результаты классификации муниципальных образований Оренбургской области по показателям демографической безопасности для 2004-2006 гг. приведены в приложении Б.

Анализируя значения функционала качества, можно утверждать, что наилучшим оказывается разбиение с помощью сетей Кохонена.

Классификация городов и районов по состоянию показателей, характеризующих демографическую безопасность, выявила существование в пределах Оренбургской области трех классов, каждый из которых представляет собой группу объектов, связанных единством демографического, экономического, географического, социального положения. По результатам классификации можно сделать следующие выводы. Показатели, обеспечивающие более высокий уровень демографической безопасности, характерны для всех городов области, а также для Новоорского и Оренбургского районов. Города в среднем существенно опережают сельскую местность по уровню демографической безопасности. Два сельских района: Оренбургский и Новоорский также вошли в первый класс, что объясняется их близостью к крупнейшим городам области. Второй класс характеризуется показателями, обеспечивающими средний уровень демографической безопасности. Это явление можно объяснить тем, что в условиях естественной убыли населения прирост числа жителей происходит только за счет миграционного притока. Третий класс состоит из сельских районов, которые характеризуются с показателями, обеспечивающими более низкий уровень демографической безопасности. Высокие показатели заболеваемости и смертности объясняются низким уровнем медицинского обслуживания, отсутствием экстренной медицинской помощи в сельских районах, а также достаточно высокой долей населения старше трудоспособного возраста.

Классификация городов и районов Оренбургской области по уровню демографической безопасности показала, что за исследуемый период происходили изменения в однородности. Изменения в структуре классов происходили в основном за счет перехода районов из третьего класса во второй, и наоборот. Наиболее неустойчивым оказалось положение следующих районов: Акбулакского, Адамовского, Кувандыкского, Красногвардейского, Первомайского, Сакмарского, Ташлинского и Тоцкого. В этих районах достаточно высока миграционная подвижность, однако из-за низкого медицинского обслуживания и отсутствия экстренной медицинской помощи высоки показатели заболеваемо-

сти и смертности. Это и обуславливает постоянный переход этих районов из класса в класс.

## **2.6 Классификация муниципальных образований Оренбургской области по показателям, характеризующим инвестиционную привлекательность региона**

Инвестиционная привлекательность представляет собой латентную категорию, которая характеризуется совокупностью признаков, отражающих сложившиеся на территории (в стране, в регионе) условия (экономические, политические, инфраструктурные, трудовые, финансовые и др.) и влияющие на ход инвестиционного процесса. Часть признаков, отражающая реальное развитие инвестиционной деятельности в регионе в виде капиталовложений в основной капитал отраслей экономики, характеризует инвестиционную активность. Другая часть признаков, «суммарно» отражающая объективные предпосылки для притока инвестиций, характеризует такую сторону инвестиционной привлекательности, как инвестиционный потенциал. Часть совокупности показателей, отражающая условия возможного возникновения непредвиденных финансовых потерь, характеризует инвестиционную привлекательность с позиции инвестиционного риска.

Для обеспечения комплексного анализа инвестиционной привлекательности в Оренбургской области сформирован перечень статистических показателей, представленный в таблице 2.8.

Таблица 2.8 – Система показателей, характеризующих инвестиционную привлекательность в Оренбургской области

Обозначения показателей	Наименование показателя
1	2
<b>Показатели, характеризующие инвестиционную активность</b>	
$y_1$	объем инвестиций в основной капитал на душу населения, рублей
$y_2$	ввод в действие жилых домов на 1000 человек населения, кв.метров общей площади

Продолжение таблицы 2.8

1	2
$y_3$	объем промышленной продукции на душу населения, рублей
$y_4$	производство (реализация) скота и птицы, центнеров на душу населения
<b>Показатели, характеризующие инвестиционный потенциал</b>	
<i>Показатели трудового потенциала</i>	
$x_{1.1}$	уровень безработицы, в процентах от населения в трудоспособном возрасте
$x_{1.2}$	доля населения в трудоспособном возрасте в общей численности населения, в процентах
$x_{1.3}$	доля лиц моложе трудоспособного возраста, в общей численности населения, в процентах
$x_{1.4}$	среднегодовая численность работников, занятых в сельскохозяйственном производстве, человек
$x_{1.5}$	среднегодовая численность работников, занятых в промышленности, человек
$x_{1.6}$	число зарегистрированных иностранных рабочих, в промилле от численности населения в трудоспособном возрасте
<i>Показатели финансового потенциала</i>	
$x_{2.1}$	сальдированный финансовый результат (прибыль минус убыток) на одно предприятие, рублей
$x_{2.2}$	уровень рентабельности реализованной продукции сельского хозяйства в сельскохозяйственных организациях, в процентах
$x_{2.3}$	уровень рентабельности реализованной продукции организаций промышленности, в процентах
<i>Показатели потребительского потенциала</i>	
$x_{3.1}$	оборот розничной торговли на душу населения, рублей
$x_{3.2}$	объем платных услуг на душу населения, рублей
<b>Показатели, характеризующие инвестиционный риск</b>	
$y_5$	задолженность организаций по заработной плате, в процентах от общего фонда заработной платы
$y_6$	удельный вес убыточных предприятий и организаций, в процентах от общего числа предприятий
$y_7$	просроченная кредиторская задолженность предприятий, в процентах от общей задолженности

Для характеристики региональной неоднородности инвестиционного пространства Оренбургской области, где под региональной неоднородностью инвестиционного пространства понимается уровень внутрирегиональных различий в относительных объемах привлекаемых капиталовложений, был рассчитан индекс концентрации Герфиндаля–Хиршманна (НИ) [20]:

$$HNI = \sum_{j=1}^k \left( \frac{i_j}{I} \cdot 100 \right)^2, \quad (2.27)$$

где  $k$  - количество муниципальных образований в регионе;

$i_j$  - объем инвестиций в  $j$ -ом муниципальном образовании;

$I$  - общий объем инвестиций в регионе.

Теоретически максимальное значение индекса ННІ, равное 10000, достижимо, когда весь объем инвестиций приходится на один муниципалитет. Минимальное его значение соответствует равномерному распределению инвестиций по 47 муниципалитетам области ( $212,77=10000/47$ ) (рисунок 2.7).

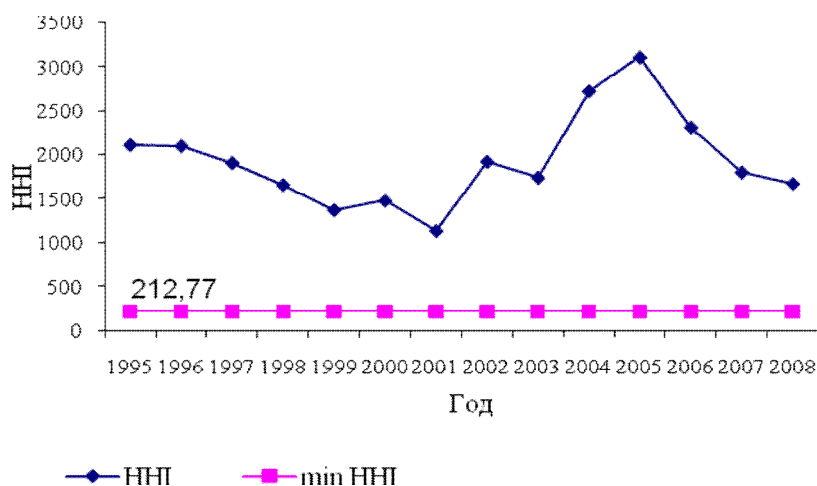


Рисунок 2.7 – Динамика индекса ННІ в Оренбургской области, 1995-2008гг.

Динамика индекса Герфиндаля-Хиршманна за период с 1995 по 2008гг. наглядно указывает на довольно высокий уровень концентрации инвестиций в основной капитал в отдельных образованиях. С 1995 по 1999гг. индекс ННІ уменьшался, что вызвано слабой инвестиционной активностью в этот период. Неравномерное распределение инвестиций в последние годы объясняется тем, что, во-первых, существенная их часть стала направляться в те образования, где сосредоточены запасы нефти и газа; во-вторых, значительно ослабла регулирующая роль государства, сократились государственные инвестиции на развитие региона. В результате существенная доля инвестиций стала аккумулиро-



ваться в отдельных, наиболее привлекательных муниципалитетах, что обусловило нарастание неоднородности инвестиционного пространства внутри региона [21].

Ставится задача разбиения административно-территориальных образований Оренбургской области на однородные группы по сформированному набору показателей, характеризующих инвестиционную привлекательность.

Целесообразно сначала провести классификацию по одному из агломеративных иерархических методов. На рисунке 2.8 представлена дендрограмма классификации 12 городов и 35 районов Оренбургской области, построенная методом Уорда с использованием обычного евклидова расстояния. При пороговом значении расстояния в 15 единиц объекты разбиваются на три класса.

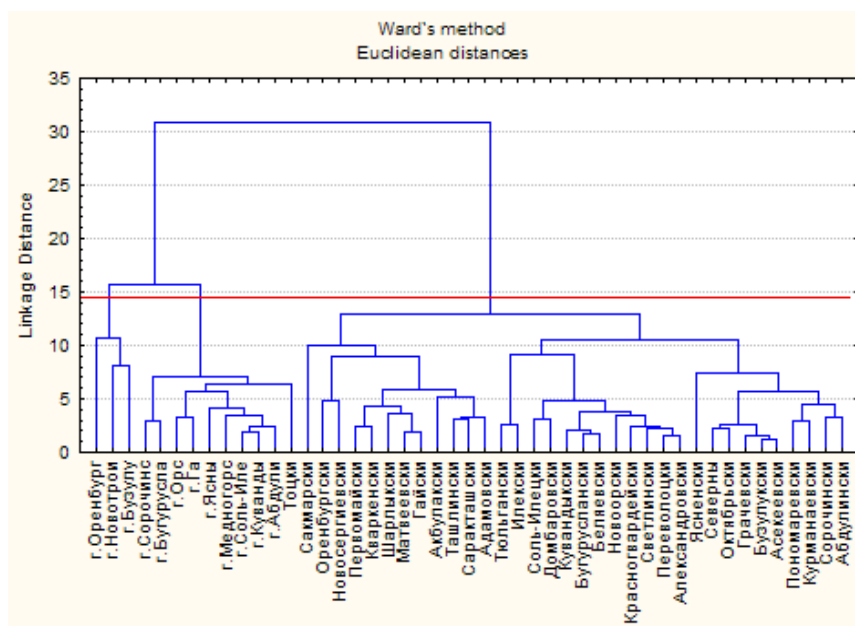


Рисунок 2.8 – Дендрограмма классификации городов и районов Оренбургской области методом Уорда в 2008 году

Уточнение состава классов проведено с помощью метода  $k$ -средних, который дает лучшее разбиение по функционалу качества. Результаты классификации методом  $k$ -средних для 2008 года приведены в таблице 2.9.

Таблица 2.9 – Классификация городов и районов Оренбургской области методом k-средних в 2008 году

Номер кластера	Количество объектов в кластере	Состав класса
кластер 1	13	<b>Города:</b> Абдулино, Бугуруслан, Бузулук Гай, Кувандык, Медногорск, Новотроицк, Оренбург, Орск, Соль-Илецк, Сорочинск, Ясный. <b>Район:</b> Оренбургский.
кластер 2	20	<b>Районы:</b> Адамовский, Акбулакский, Александровский, Асекеевский, Гайский, Кваркенский, Красногвардейский, Кувандыкский, Матвеевский, Новосергиевский, Новоорский, Октябрьский, Первомайский, Переволоцкий, Сакмарский, Саракташский, Северный, Сорочинский, Ташлинский, Шарлыкский.
кластер 3	14	<b>Районы:</b> Абдулинский, Беляевский, Бугурусланский, Бузулукский, Грачевский, Домбаровский, Илекский, Курманаевский, Пономаревский, Светлинский, Соль-Илецкий, Тоцкий, Тюльганский, Ясененский.

Суммарная внутриклассовая дисперсия  $F = 59,19$ .

Для интерпретации полученной классификации построены графики средних значений центрировано-нормированных признаков в классах, представленные на рисунке 2.9.

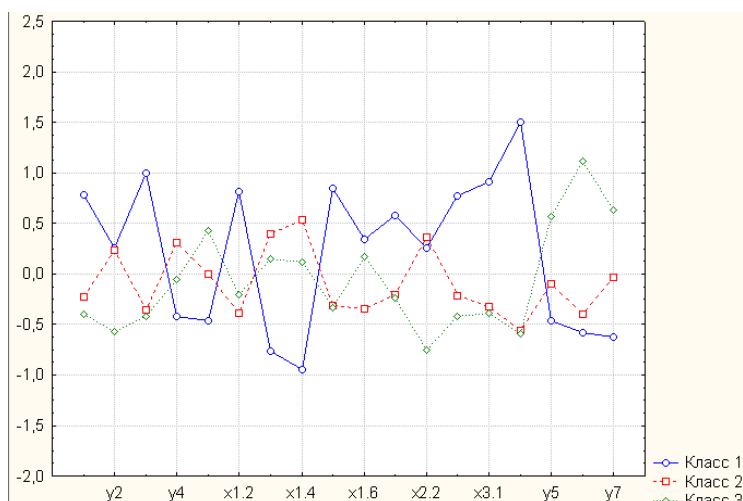


Рисунок 2.9 – График средних значений признаков в классах для 2008 года

В первый класс вошли все города Оренбургской области и Оренбургский район. Образования, вошедшие в первый класс, характеризуются более высокими, по сравнению с остальными классами, средними значениями показателей, характеризующих инвестиционную активность, исключение составил по-

казатель – объем сельскохозяйственной продукции на душу населения, более высокими значениями показателей финансового и потребительского потенциала и наименьшими значениями показателей, характеризующих инвестиционный риск.

Во второй класс вошли районы, отраслевой специализацией которых является сельское хозяйство. Они характеризуются сравнительно высокими средними значениями таких показателей, как производства птицы на душу населения, среднегодовая численность работников, занятых в сельскохозяйственном производстве; уровень рентабельности реализованной продукции сельского хозяйства.

В третьем классе, по сравнению с остальными, более высокие средние значения показателей, характеризующих инвестиционный риск и уровень зарегистрированной безработицы.

Проведем разбиение объектов на однородные группы методами нейросетевой классификации. Чтобы правильно распределить плотность ядер классов (векторов весов) в соответствии с плотностью входных векторов в пространстве  $X$  для обучения сети будем применять метод выпуклой комбинации. Чтобы определить класс, к которому относится объект, нужно выбрать среди всех нейронов данного слоя один с максимальным выходом. Рассмотренная сеть нейронов, использующая евклидову меру близости для классификации объектов, называется сетью Кохонена. Результаты классификации методом нейросетевой классификации для 2008 года представлены в таблице 2.10.

Таблица 2.10 – Классификация городов и районов Оренбургской области в 2008 году с помощью сети Кохонена

Номер кластера	Количество объектов в кластере	Состав класса
1	2	3
кластер 1	13	<b>Города:</b> Абдулино, Бугуруслан, Бузулук Гай, Кувандык, Медногорск, Новотроицк, Оренбург, Орск, Соль-Илецк, Сорочинск, Ясный. <b>Район:</b> Оренбургский.

Продолжение таблицы 2.10

1	2	3
кластер 2	14	<b>Районы:</b> Адамовский, Акбулакский, Александровский, Гайский, Красногвардейский Матвеевский, Новосергиевский, Новоорский, Октябрьский, Переволоцкий, Сакмарский, Саракташский, Северный, Шарлыкский.
кластер 3	20	<b>Районы:</b> Абдулинский, Асекеевский, Беляевский, Бугурусланский, Бузулукский, Грачевский, Домбаровский, Илекский, Кваркенский, Кувандыкский, Курманаевский, Первомайский, Пономаревский, Светлинский, Соль-Илецкий, Сорочинский, Ташлинский, Тоцкий, Тюльганский, Ясенский.

Суммарная внутриклассовая дисперсия  $F = 57,19$ .

По сравнению с разбиением методом  $k$ -средних произошли следующие изменения в структуре классов: из второго класса в третий перешли Асекеевский, Кваркенский, Кувандыкский, Первомайский, Сорочинский, Ташлинский районы.

Анализируя значения функционала качества, можно утверждать, что наилучшим оказывается разбиение с помощью сетей Кохонена.

Результаты классификации методом Уорда,  $k$ -средних и с помощью сети Кохонена в 2004-2007 гг. приведены в приложении В.

Таким образом, классификация городов и районов по показателям, характеризующим инвестиционную привлекательность, выявила существование в пределах Оренбургской области трех классов: со сравнительно высоким, средним и низким уровнем инвестиционной привлекательности.

Показатели, обеспечивающие сравнительно высокий уровень инвестиционной привлекательности, характерны для всех городов области, а также для Оренбургского района. Города в среднем существенно опережают сельскую местность по уровню инвестиционной привлекательности. Оренбургский район вошел в первый класс, что объясняется его близостью к крупнейшему городу области. Второй класс характеризуется показателями, обеспечивающими средний уровень инвестиционной привлекательности. Его состав в основном образуют сельские районы. Третий класс состоит из районов, которые характеризуются показателями, обеспечивающими более низкий уровень инвести-

онной привлекательности. Для этих районов характерны сравнительно высокие значения таких показателей, как удельный вес убыточных предприятий, задолженность по заработной плате, уровень безработицы.

Классификация городов и районов Оренбургской области по уровню инвестиционной привлекательности показала, что за исследуемый период происходили определенные сдвиги в структуре классов за счет, в основном, перехода районов из третьего класса во второй и наоборот. Эта динамика характеризует неустойчивость инвестиционных процессов ряда административно-территориальных образований. В частности, наиболее неустойчивым оказалось положение Первомайского и Красногвардейского районов.

## **2.7 Многомерная классификация административно-территориальных образований по показателям, характеризующим антропогенную нагрузку и эколого-экономический риск**

Моделирование антропогенной нагрузки и эколого-экономического риска региона будем проводить на примере Оренбургской области, поскольку районы Оренбуржья существенно различаются по природно-климатическим условиям, пригодности для сельского хозяйства, размещению промышленных центров, география распространения полезных ископаемых, уровню промышленного и экономического развития, что позволяет предполагать их неоднородность и по уровню антропогенной нагрузки и эколого-экономического риска [22].

Сформируем состав показателей, которые характеризуют антропогенную нагрузку. Возникновению эколого-экономического риска сопутствует снижение качества среды обитания, что предопределяет необходимость иметь некоторый показатель, количественно характеризующий это состояние или качество. Единого показателя, адекватно отражающего качество произвольной экосистемы, не существует как по причине отсутствия единого понимания самой

экосистемы и подходов к определению качества ее состояния, так и вследствие специфичных особенностей каждой исследуемой территории. Вместе с тем, причиной снижения качества окружающей среды в большинстве случаев является воздействие на нее антропогенной деятельности, и с увеличением уровня воздействия увеличивается вероятность возникновения неблагоприятных для среды последствий. Поэтому потенциально опасное снижением качества состояние среды будем характеризовать нагрузкой антропогенной деятельности на окружающую среду, в качестве компонент которой будем рассматривать атмосферу (воздушную среду), гидросферу (водную среду) и литосферу (почву).

Нагрузку на *водную среду* можно охарактеризовать таким показателем, как сброс сточных вод в поверхностные водоемы на единицу территории. Сброс сточных вод в поверхностные водоемы включает в себя объемы нормативно-чистых, нормативно-очищенных и загрязненных стоков (производственных и коммунальных), сброшенных в поверхностные водоемы [23].

Нагрузку на *воздушную среду* от стационарных источников будем характеризовать показателем выбросов загрязняющих веществ от стационарных источников на единицу территории. При этом под стационарными источниками выделения вредных веществ в атмосферный воздух понимаются, неподвижные технологические агрегаты (установки, устройства, аппараты и т.п.), выделяющие в процессе эксплуатации вредные вещества. Что касается объема выбросов от нестационарных (передвижных) источников, то будем предполагать, что в определенной степени это можно учесть через количество автомобилей, находящихся в личной собственности граждан, на единицу территории.

Для характеристики *нагрузки на почву* будем использовать показатель объема токсичных отходов на единицу территории [24]. Под токсичными отходами понимаются отходы производства и потребления, представляющие опасность для здоровья населения и окружающей среды.

Таким образом, для характеристики нагрузки антропогенной деятельности на окружающую среду будут использоваться 4 показателя:

- сброс сточных вод в поверхностные водоемы на единицу территории (SB, млн. куб. м. на 1 тыс. кв. м);
- выбросы загрязняющих веществ от стационарных источников на единицу территории (ST, тыс. тонн на 1 тыс. кв.м.);
- количество автомобилей, находящихся в личной собственности граждан, на единицу территории (NEST, штук на 1 тыс. кв.м.);
- объем токсичных отходов на единицу территории (TOX, тыс. тонн на 1 тыс. кв.м.).

Перейдем к формированию набора показателей для характеристики эколого-экономического риска, то есть риска, которому вынужденно подвергаются экономические агенты (в частности, население), функционируя в нарушенной среде. Прежде всего, это риск ухудшения их здоровья, приводящего к сокращению продолжительности трудовой жизни, снижению производительности труда, а в масштабах всей экономики - потерям трудового потенциала. Воздействие загрязненной окружающей среды на человека происходит за счет его контакта с загрязняющими веществами, содержащимися в окружающей среде (воздухе, воде, почве, продуктах растениеводства и животноводства), последствия воздействия зависят от концентрации вещества и продолжительности его воздействия (экспозиции) на человека. Поэтому для оценки рисков для человека от воздействия содержащихся в атмосферном воздухе и водопроводной воде загрязняющих веществ целесообразно использовать методику [25], [26].

Поскольку многие из веществ оказывают воздействие на одни и те же системы организма человека, осуществлен переход от коэффициентов опасности воздействия каждого вещества в отдельности к индексам опасности: для желудочно-кишечного тракта (ZKT), центральной нервной системы (CNS), почек (Pochki), сердечнососудистой системы (SSS), крови (Krov), печени (Pechen), иммунной системы (IS), репродуктивной системы (RS), гормональной системы (GS), кожи (Kozha), слизистой (Sliz), развития (Razv) и органов дыхания (OD). Рассчитаны также величины канцерогенного риска по мышьяку (Mish), свинцу (Svinec), хрому (VI)(Cr) и формальдегиду (Form). Получили, что

для характеристики эколого-экономического риска экономическим агентам будут использоваться 17 показателей.

Таким образом, переходим к решению задачи разбиения городов и районов области на однородные группы по сформированным наборам из 4 и 17 показателей, характеризующих антропогенную нагрузку на окружающую среду и эколого-экономический риск указанные категории. Классификацию будем проводить также по временным сечениям, поскольку достаточно высокая динамика протекающих в области экономических и социальных процессов позволяет предполагать неоднородность административно-территориальных образований области не только в пространстве, но и во времени.

В связи с необходимостью наглядного представления классифицируемых объектов в пространстве небольшой размерности и коррелированности признаков, по которым проводится классификация, для выделения однородных групп объектов нами предлагается использовать метод многомерного шкалирования.

Традиционно в многомерном шкалировании применяются понятия, которые не привычны или вовсе не встречаются в разделах классической статистики. Так, стимулом обычно называется объект, обладающий определенным набором присущих ему характеристик. Профиль стимула – набор количественных показателей, характеризующих стимулы. Стимульное пространство – это теоретическое пространство расположения стимулов. Шкала – ось теоретического пространства, которая является носителем значений обобщенного фактора. Стресс-формула – формула, применяемая для оценки соответствия эмпирических и теоретических ранговых данных. В качестве таких критериев обычно рассматриваются нормализованный стресс, так называемые стресс-1 и стресс-2, S-стресс и др. Эти критерии также используются при определении неизвестной заранее размерности теоретического пространства  $p$ , причем обычно рекомендуется получать решение в различных размерностях и выбирать одно из них, ориентируясь, например, на критерии соответствия данным, интерпретируемости и воспроизводимости. Один из способов оценки соответствия основывается на анализе графика зависимости величины стресса от размерности



(график каменистой осыпи) – нужно найти такую абсциссу на графике, в которой уменьшение стресса максимально замедляется, то есть изгиб графика должен приходиться именно на истинную размерность  $p$  [27]. Краскал и Виш отмечают, что не следует признавать решение со значением стресса выше 0,1, если это решение не одномерно. Однако если уровень ошибок измерения или выборки высок, то решение со значением стресса выше 0,1 признать можно. В размерности 1 стресс 0,15 и ниже представляет хорошее соответствие данным. Более того, Краскал и Виш пришли к выводу, что редко возникает необходимость добавлять больше координатных осей, чем это необходимо для снижения стресса ниже 0,05. Решение в более высокой размерности предпочитается решению в более низкой размерности, если существуют важные характеристики стимулов, проявляющиеся только в решение более высокой размерности. И наоборот, решение в более низкой размерности предпочитается, если нет таких существенных характеристик стимулов, которые не проявляются в решении низкой размерности [28]. При интерпретации результатов, анализируя итоговое расположение объектов в стимульном пространстве, следует в дополнение к существующим осям координат искать различные кластер-конфигурации (окружности, круги и т.д.) [29]. Также отметим, что получаемые в результате многомерного шкалирования оси всегда определены с точностью до поворота, выбор которого часто остается за исследователем; возможны ситуации, когда хорошо интерпретируемые решения получаются сразу при нескольких поворотах [30].

В нашем случае стимулами выступают города и районы Оренбургской области (4 города и 35 районов, всего 39 стимулов), профиль стимула содержит 4 количественных показателя, характеризующих антропогенную нагрузку, и 17 количественных показателей, характеризующих эколого-экономический риск. Приведем исходные данные к стандартизированному виду, затем для получения из исходной матрицы  $X$  матрицы различий  $\Delta = \{\delta_{ij}\}_{\substack{i=1,..,n \\ j=1,..,n}}$  используем обычное евклидово расстояние. Будем стремиться не аппроксимировать непосредствен-

но различия  $\delta_{ij}$ , а подбирать такую монотонную с исходными различиями последовательность чисел, которая была бы более близка к точным расстояниям.

Обозначим расстояние между  $i$  и  $j$  стимулами в теоретическом пространстве через  $d_{ij}$ , при этом  $d_{ij} = \sqrt{\sum_{i=1}^p (\tilde{x}_i - \tilde{x}_j)^2}$ , где  $\tilde{x}_i$  и  $\tilde{x}_j$  – вектора координат стимулов  $i$  и  $j$  соответственно в теоретическом пространстве размерности  $p$ . Преобразованное (уточненное) с учетом порядка исходных различий  $\delta_{ij}$  расстояние между  $i$  и  $j$  стимулами обозначим через  $\hat{d}_{ij}$ . Будем искать координаты стимулов  $\tilde{X}$  в новом теоретическом пространстве, минимизируя функцию потерь следующего вида:

$$\sigma^2 = \frac{\sum_{j,j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i,j} \hat{d}_{ij}^2}. \quad (2.28)$$

Данная функция называется нормализованным стрессом (normalized raw stress). Кроме (2.28) для оценки соответствия эмпирических и теоретических данных будем использовать стресс-1, стресс-2, S-стресс, долю учтенной вариации и коэффициент конгруэнции Тьюки. Чем меньше значения стрессов 1 и 2, S-стресса и чем ближе к 1 доля учтенной дисперсии и коэффициент конгруэнции, тем выше соответствие эмпирических и теоретических данных.

Ориентируясь на традиционные критерии определения размерности теоретического пространства, необходимость наглядного отображения данных, возможность достаточно ясной содержательной интерпретации и проанализировав графики зависимости нормализованного стресса от размерности (наиболее существенное его уменьшение, на 80%, наблюдается при переходе от одной шкалы к двум), для всех годов размерность теоретического пространства  $p$  была принята равной 2. Требование наличия не менее 5 стимулов на каждую координатную ось, возникновение которой ожидалось при анализе, для иссле-

дуемой совокупности из 39 стимулов было выполнено. Поиск конфигурации стимулов (собственно сам алгоритм шкалирования) был выполнен в ППП SPSS. Интерпретация получившихся шкал сделана на основе анализа коэффициентов корреляции между координатами стимулов в исходном пространстве  $X$  и координатами стимулов в новом теоретическом пространстве  $\tilde{X}$ : высокий значимый коэффициент корреляции свидетельствует о значительном вкладе данного конкретного показателя в содержательную интерпретацию оси [27]. Для расчета коэффициентов корреляции и для визуализации результатов использовалась Statistica.

Рассмотрим результаты классификации городов и районов области по показателям антропогенной нагрузки за 2007 г. Величина нормализованного стресса ( $\sigma^2 = 0,007$ ), стресс-формулы Краскала  $S1 = 0,084$  и  $S2 = 0,113$ , коэффициент  $S - stress = 0,021$  достаточно малы, а доля учтенной вариации и коэффициент конгруэнции Тьюки близки к 1 ( $D.A.F. = 0,993$  и  $TCC = 0,997$ ), что свидетельствует о хорошей степени соответствия теоретических результатов эмпирическим данным и о возможности корректного представления объектов в новой системе координат [27, 13]. С целью упрощения интерпретации был осуществлен поворот обеих осей на 45 градусов (по часовой стрелке). Первая шкала характеризует уровень антропогенной нагрузки на водную среду, а также на воздушную среду от нестационарных источников; вторая шкала тесно связана с показателем выбросов от стационарных источников и количеством токсичных отходов на единицу территории (таблица 2.11). Представим объекты в полученном стимульном пространстве (рисунки 2.10, 2.11).

Таблица 2.11 – Коэффициенты корреляции между координатами стимулов в исходном и новом пространствах за 2007 г. (после вращения)

Новые признаки	Исходные признаки			
	SB	NEST	ST	TOX
Шкала 1 (d1)	0,06	0,32*	0,84**	0,78**
Шкала 2 (d2)	0,92**	0,88**	0,44**	0,23

\* - значим на уровне 0,05; \*\* - значим на уровне 0,01

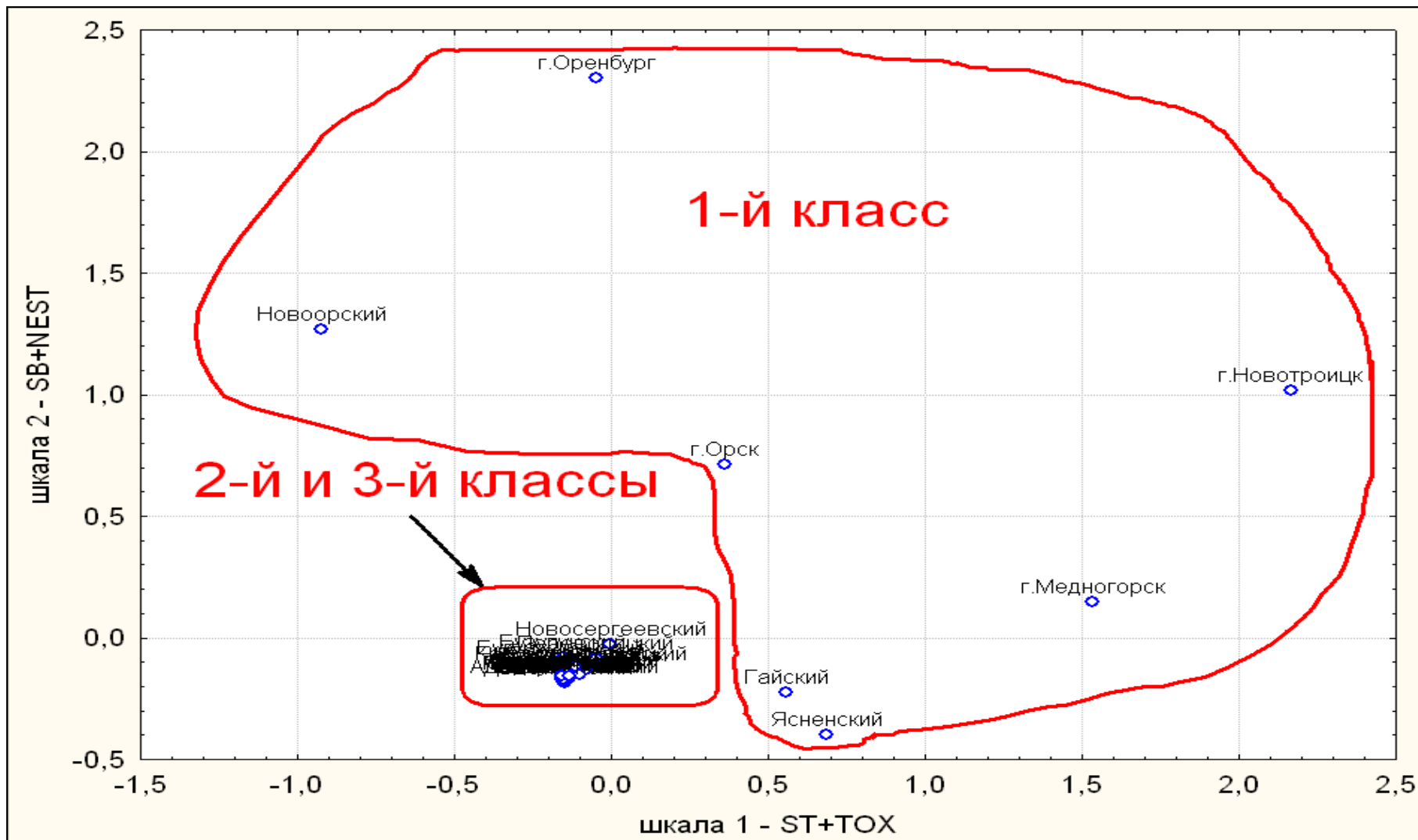


Рисунок 2.10 – Расположение городов и районов Оренбургской области в пространстве новых переменных за 2007 г.

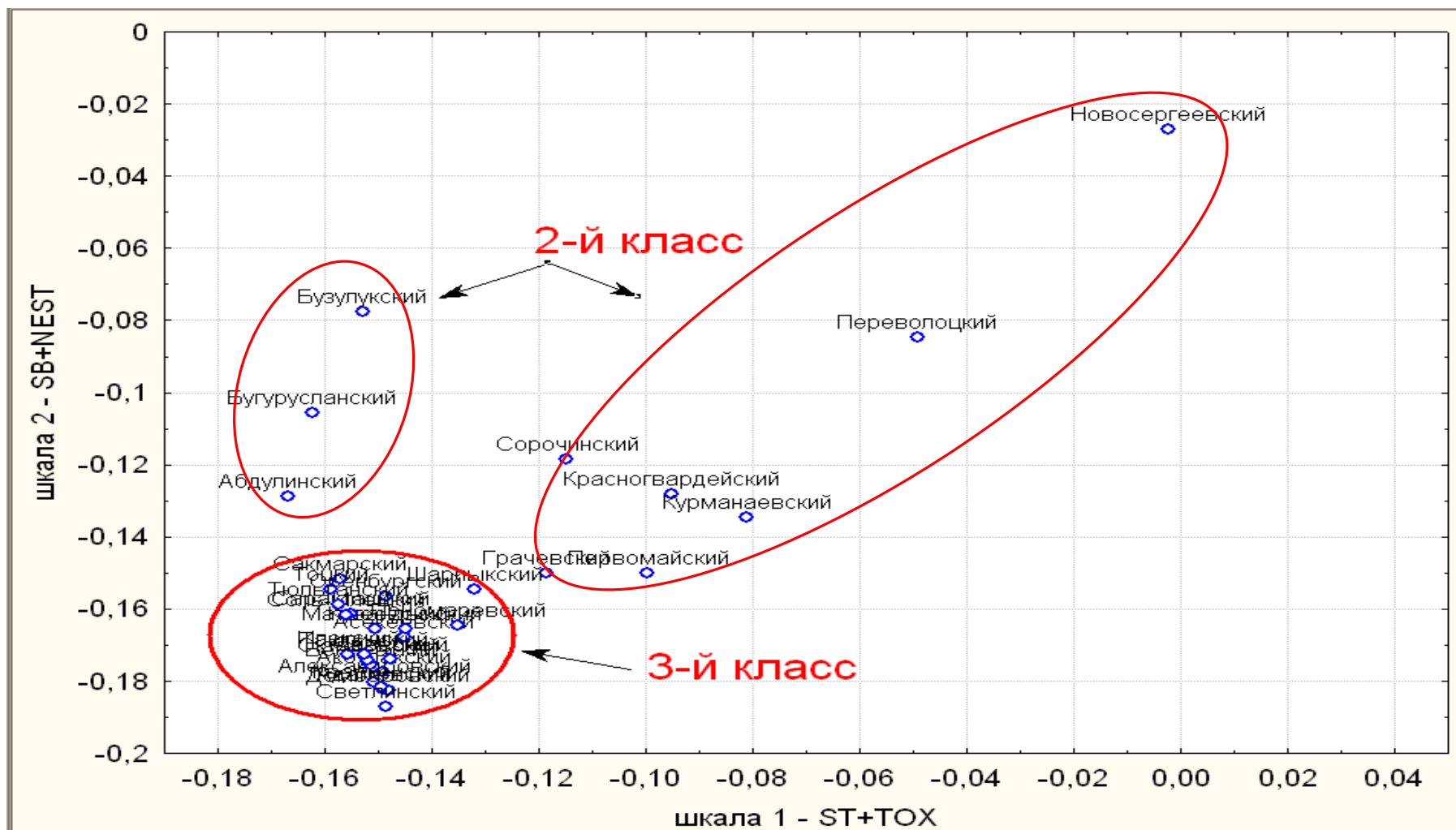


Рисунок 2.11 – Расположение городов и районов Оренбургской области, относящихся ко второму и третьему классам, в пространстве новых переменных за 2007 г.

В соответствии с интерпретацией шкал объекты, располагающиеся в углах графика, характеризуются наибольшим уровнем антропогенного воздействия на окружающую среду, а сосредоточенные вокруг нулевого значения - наименьшим. Как видно из рисунка 2.10, наиболее выделяющимися объектами (расположенными на значительном расстоянии от начала координат) являются центр области город Оренбург и относящиеся к восточной зоне Оренбуржья города Новотроицк, Медногорск и Орск, а также Новоорский, Гайский и Ясненский районы. Эти объекты мы отнесем в класс под номером 1, характеризующийся высокой антропогенной нагрузкой на окружающую среду. Предполагается, что чем больше нагрузка на окружающую среду, тем больше вероятность наступления событий, имеющих неблагоприятные последствия для нее, то есть тем больше, согласно определению, экологический риск. Для остальных объектов характерна меньшая антропогенная нагрузка, но изменив масштаб изображения стимулов (рисунок 2.11), видим, что среди них выделяется «ядро» районов с минимальной нагрузкой на окружающую среду: выделим их в класс под номером 3, оставшиеся районы – в класс под номером 2 (средняя нагрузка). Третий класс является более однородным в отличие от второго и особенно первого класса, объекты в которых значительно больше отличаются друг от друга. Состав каждого класса приведен в таблице 2.12.

Использование многомерного шкалирования позволило не только разнести города и районы области по классам с разным уровнем антропогенной нагрузки и говорить о том, что для объектов данного класса в среднем характерны высокие, средние или низкие значения показателей, характеризующих антропогенную нагрузку, но и наглядно показать, по каким именно группам показателей различаются объекты внутри каждого класса. Так, Оренбург загрязняет окружающую среду выбросами от нестационарных источников, промышленные города Новотроицк и Медногорск - от стационарных. Для Гайского и Ясненского районов значима проблема токсичных отходов, для Новоорского – загрязнения водного бассейна. Анализируя результаты шкалирования за 2000-2007 гг., можно сделать вывод об увеличении антропогенной нагрузки

(экологического риска) в Грачевском, Красногвардейском, Курманаевском и Новосергиевском районах (переход из третьего класса во второй), некотором уменьшении – в Домбаровском и Саракташском районах (переход из класса 2 в класс 3).

Таблица 2.12 – Состав классов, выделенных по показателям, характеризующим антропогенную нагрузку на окружающую среду, 2007 г

Класс	Количество объектов в классе	Состав класса
Высокий риск (нагрузка)	7	<b>Города:</b> Медногорск, Новотроицк, Оренбург, Орск <b>Районы:</b> Гайский, Новоорский, Ясненский
Средний риск (нагрузка)	10	<b>Районы:</b> Абдулинский, Бугурусланский, Бузулукский, Грачевский, Красногвардейский, Курманаевский, Новосергиевский, Первомайский, Переволоцкий, Сорочинский
Низкий риск (нагрузка)	22	<b>Районы:</b> Адамовский, Акбулакский, Александровский, Асекевский, Беляевский, Домбаровский, Илекский, Кваркенский, Кувандыкский, Матвеевский, Октябрьский, Оренбургский, Пономаревский, Сакмарский, Саракташский, Северный, Светлинский, Соль-Илецкий, Ташлинский, Тоцкий, Тюльганский, Шарлыкский

Аналогичным образом проведем классификацию городов и районов Оренбургской области по показателям, характеризующим эколого-экономический риск населению за 2007 год (таблицы 2.13 и 2.14).

Таблица 2.13 – Коэффициенты корреляции между координатами стимулов в исходном и новом пространствах за 2007 год (индексы опасности для критических систем организма)

Исх. признаки (Н1)	Новые признаки		Исх. признаки (Н1)	Новые признаки	
	Шкала 1 (d1)	Шкала 2 (d2)		Шкала 1 (d1)	Шкала 2 (d2)
1	2	3	4	5	6
почки	0,333*	0,809**	иммунная система	0,229	-0,356*
печень	0,321*	0,812**	кожа	0,229	-0,356*
слизистая	0,346*	0,807**	репродукт. система	0,135	-0,217
ЖКТ	0,403*	0,570**	развитие	0,135	-0,217

Продолжение таблицы 2.13

1	2	3	4	5	6
ССС	0,220	-0,413*	кровь	0,118	0,172
гормон. система	0,220	-0,361*	органы дыхания	0,998**	-0,011
ЦНС	0,230	-0,358*	-	-	-

\* - значим на уровне 0,05; \*\* - значим на уровне 0,01

Таблица 2.14 Коэффициенты корреляции между координатами стимулов в исходном и новом пространствах за 2007 год (канцерогенные риски)

Исходные признаки (канцерогенный риск)	Новые признаки	
	Шкала 1 (d1)	Шкала 2 (d2)
от воздействия мышьяка	0,221	-0,363*
от воздействия свинца	0,135	-0,211
от воздействия хрома	0,333*	0,807**
от воздействия формальдегида	0,606**	0,183

\* - значим на уровне 0,05; \*\* - значим на уровне 0,01

Размерность теоретического пространства также была принята равной двум. Первая шкала тесно связана с показателями неканцерогенного риска для желудочно-кишечного тракта, почек, печени, слизистых и органов дыхания; вторая шкала – также с показателем неканцерогенного риска для желудочно-кишечного тракта, а также почек, печени, центральной нервной, сердечно-сосудистой, иммунной и гормональной систем, кожи. Кроме того, первая шкала тесно связана с канцерогенными рисками по хромю и формальдегиду, а вторая – с канцерогенными рисками по мышьяку и хромю (увеличение рисков по хромю и формальдегиду при движении от начала координат вправо и по мышьяку и хромю при движении от начала координат вверх).



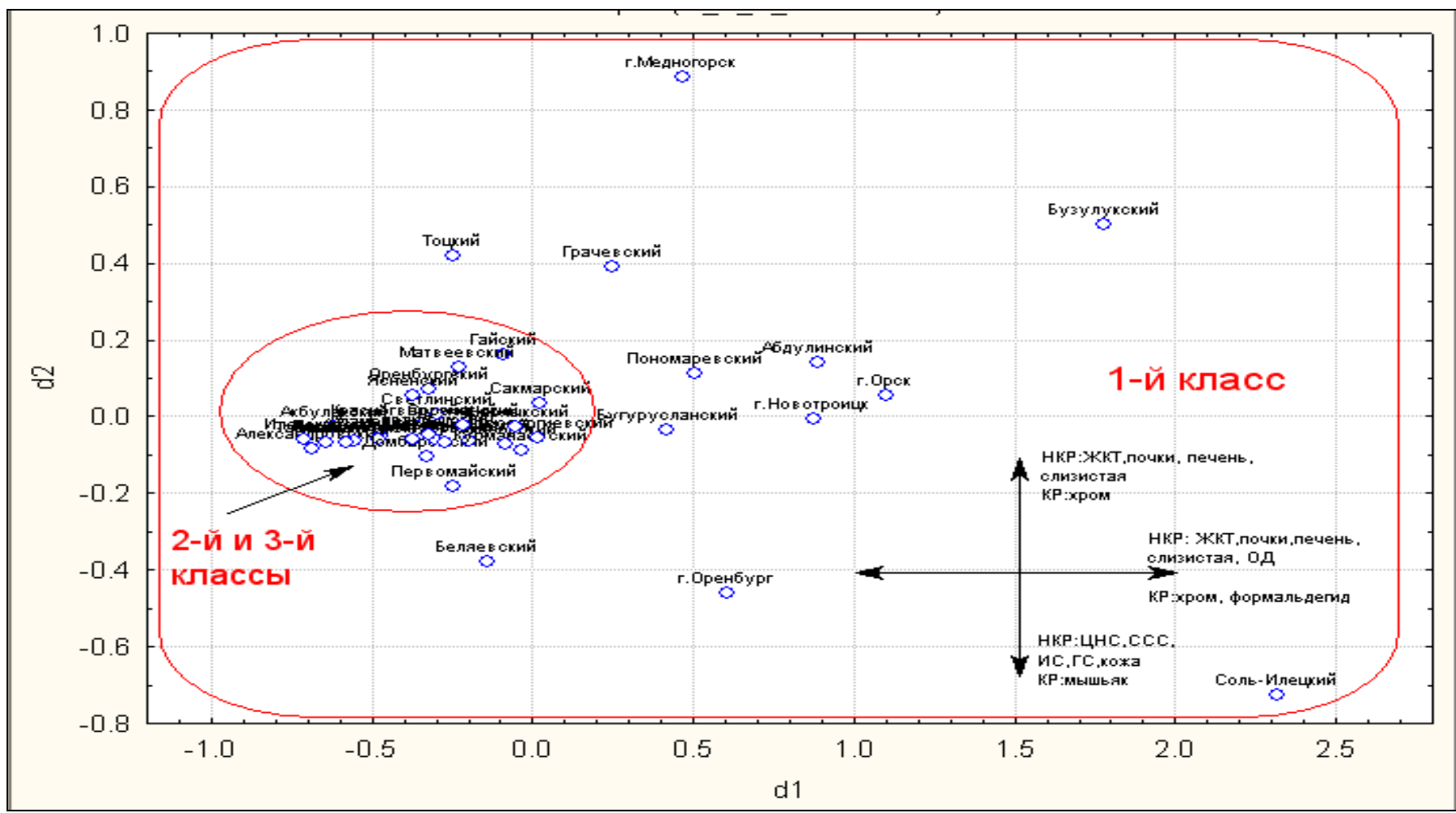


Рисунок 2.12 – Расположение городов и районов Оренбургской области в пространстве новых переменных за 2007 г.

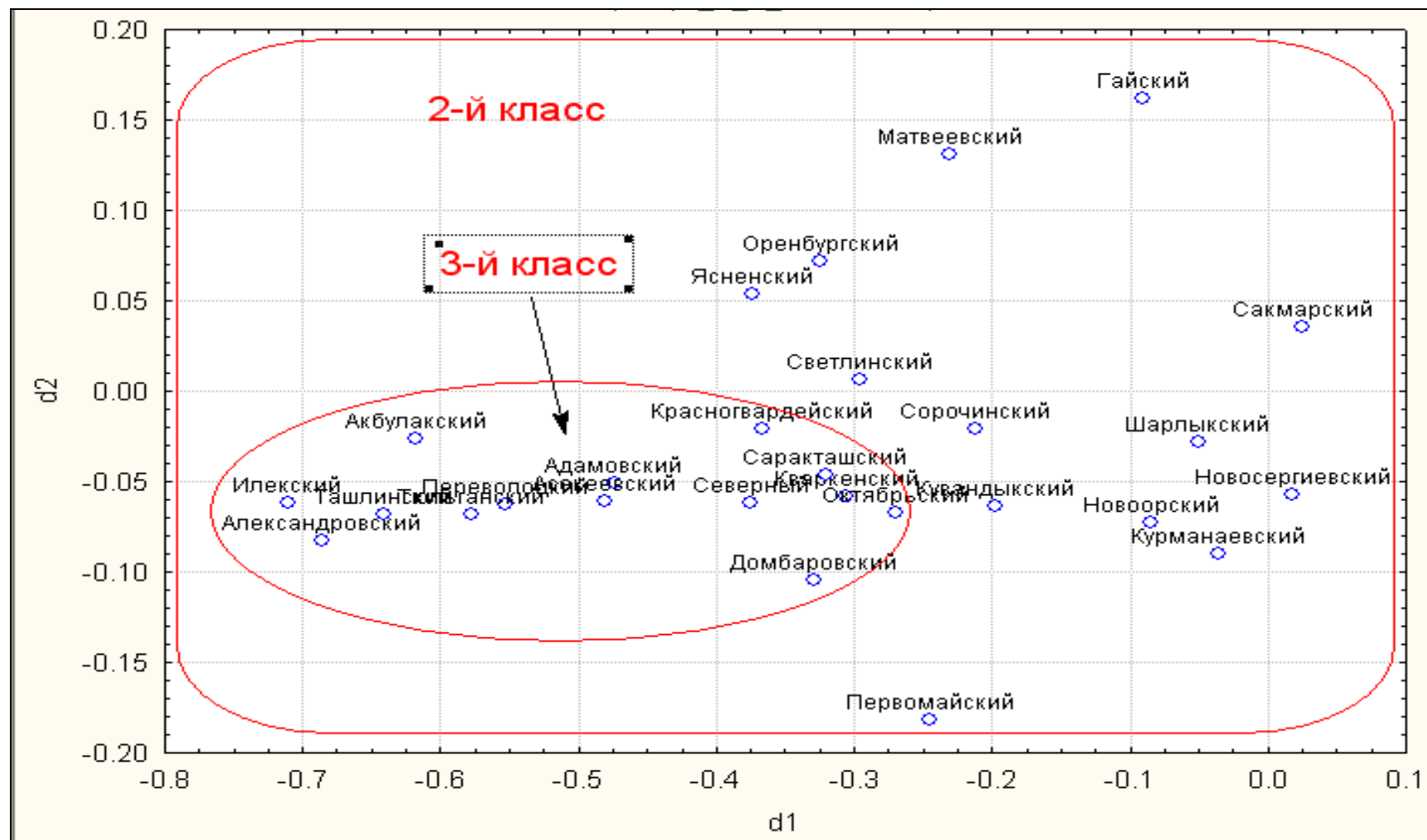


Рисунок 2.13 – Расположение городов и районов Оренбургской области, относящихся ко второму и третьему классам, в пространстве новых переменных за 2007 г.

Представим объекты в полученном стимульном пространстве (рисунки 2.12, 2.13). Из рисунка 2.12 видно, что около двух третьих всех объектов группируются в окрестности начала координат (на рисунке они выделены в овал) – это районы Оренбургской области со сравнительно низким уровнем канцерогенных или неканцерогенных рисков. Данную группу объектов целесообразно разбить на 2 класса: выделить в отдельный класс среднего риска районы, располагающиеся ближе к правой внутренней границе овала, которые характеризуются повышенными уровнями рисками для той или иной системы организма. Это, например, Гайский (суммарный неканцерогенный риск для ЖКТ, печени и почек) или Первомайский (центральная нервная система) районы и др. (рисунки 2.13).

Наиболее значительно выделяющимися объектами являются: города Оренбург, Новотроицк, Медногорск и Орск, а также Абдулинский, Беляевский, Бузулукский, Бугурусланский, Грачевский, Пономаревский, Соль-Илецкий и Тоцкий районы. Эти объекты мы отнесем в класс под номером 1, характеризующийся наибольшими рисками. Отметим, что данный класс является значительно менее однородным и плотным (разброс по первой оси 2,8 и по второй оси 1,6), чем даже совокупность объектов 2 и 3 классов в целом (разброс по первой оси 0,7 и по второй оси 0,4). Это объясняется тем, что каждый из этих объектов очень специфичен в отношении приоритетных для управления рисков, и объединяет их в один класс сам факт наличия высокого риска для той или иной системы организма. Другими словами, все эти объекты проблемны, но проблемы в каждом из них свои. Так, для Соль-Илецкого района (самая правая точка на рисунке) это суммарный неканцерогенный риск для органов дыхания, что связано с существенно ухудшившейся в этом районе ситуацией с загрязнением атмосферы; для г. Медногорска (самая верхняя точка) это риски для ЖКТ, печени, почек, слизистых и т.д. Полный состав каждого класса приведен в таблице 2.15.

Аналогичным образом проведено шкалирование за 2000-2006 гг.

В результате классификации городов и районов Оренбургской области по показателям, характеризующим эколого-экономический риск населению, за период 2000-2007 гг. можно сделать вывод, что особое внимание следует уделить сле-

дующим муниципальным образованиям области: городам Медногорск и Орск, Пономаревскому и Соль-Илецкому районам (ежегодно относились к классу высокого риска); городам Оренбург и Новотроицк, Абдулинскому, Беляевскому, Бузулукскому, Грачевскому, Курманаевскому, Матвеевскому, Первомайскому, Светлинскому, Сорочинскому, Тоцкому и Шарлыкскому районам (переход в класс с большей степенью риска).

Таблица 2.15 – Состав классов по показателям, характеризующим эколого-экономический риск населению (за 2007 г.)

Класс	Количество объектов в классе	Состав класса
Высокий риск	12	<b>Города:</b> Медногорск, Новотроицк, Оренбург, Орск <b>Районы:</b> Абдулинский, Беляевский, Бугурусланский, Бузулукский, Грачевский, Пономаревский, Соль-Илецкий, Тоцкий
Средний риск	13	<b>Районы:</b> Гайский, Кувандыкский, Курманаевский, Матвеевский, Новоорский, Новосергеевский, Оренбургский, Первомайский, Сакмарский, Светлинский, Сорочинский, Шарлыкский, Ясенский
Низкий риск	14	<b>Районы:</b> Адамовский, Акбулакский, Александровский, Асекеевский, Домбаровский, Илекский, Кваркенский, Красногвардейский, Октябрьский, Переволоцкий, Саракташский, Северный, Ташлинский, Тюльганский

## 2.8 Моделирование интегральных показателей, характеризующих социально-экономические процессы региона и сравнительный анализ муниципальных образований Оренбургской области по синтетическим категориям

### 2.8.1 Сравнительный анализ муниципальных образований Оренбургской области по уровню демографической безопасности региона

Для сравнительного анализа муниципальных образований региона по уровню демографической безопасности реализован экспертно-статистический метод. Матрицы парных сравнений  $\gamma_m$  построены по результатам разбиения объектов ис-

следования на классы с использованием методов кластерного анализа и нейросетевой классификации. Таким образом, три вида разбиений (метод Уорда, k-средних, нейросетевая классификация) позволили получить матрицы парных сравнений, эквивалентные оценкам экспертов (приложение Г). Располагая экспертной и статистической информацией, были оценены коэффициенты целевой функции.

Оценка влияния факторов на уровень демографической безопасности для 2004 года имеет вид:

$$f_{2004}(X, \Theta) = 0,74x_{1.1.1} - 0,67x_{1.1.2} - 0,12x_{1.1.3} + 0,39x_{1.1.4} - 0,60x_{1.1.5} + 0,62x_{1.1.6} + 0,26x_{1.2.1} + 1,51x_{1.2.3} - 0,66x_{1.3.1} - 0,37x_{1.3.6} + 0,19x_{1.3.12} + 0,50x_{2.2} + 0,72x_{2.3} - 0,49x_{2.8} + 1,18x_{3.3} - 0,59x_{3.5} + 0,09x_{3.8} - 0,11x_{4.1} - 0,34x_{4.2} - 0,50x_{4.3} - 0,26x_{4.4} + 0,14x_{5.1} - 0,25x_{5.2}$$

Оценка влияния факторов на уровень демографической безопасности и значения интегрального показателя для 2005-2007 гг. представлены в приложении Д. По значениям интегрального показателя произведено ранжирование муниципальных единиц Оренбургской области по заданному интегральному свойству.

Далее рассчитаны характеристики, которые позволят отслеживать улучшение или ухудшение демографической безопасности в динамике. Простое приращение (уменьшение) во времени  $t$  значения  $\tilde{y}_{i,t}$ , в действительности может не означать соответственно улучшения (ухудшения) уровня демографической безопасности как по отношению к самому себе, так и по своему положению среди других объектов: ведь во времени меняются и «объекты-эталоны» (то есть  $x_{\max}$  и  $x_{\min}$ ), и «объекты-конкуренты» (соседи по рейтингу), и если они меняются к лучшему более быстрыми темпами, чем значение  $\tilde{y}_{i,t}$ , для  $i$ -го объекта, то и межрегиональная динамика демографической безопасности для этого региона будет отрицательной (то есть будет сигнализировать об относительном ухудшении), несмотря на некоторое увеличение значения  $\tilde{y}_{i,t}$ . При измерении межрегиональной динамики каждого конкретного объекта естественно ориентироваться на динамику его положе-

ния (ранга) в ряду других объектов, то есть на величину  $\delta_{i,t} = R(\tilde{y}_{i,t-1}) - R(\tilde{y}_{i,t})$ , где  $R(\tilde{y}_{i,t}) = R_i$  - ранг  $i$ -го объекта в момент времени  $t$ .

Очевидно, положительные значения  $\delta_{i,t}$  будут свидетельствовать о положительной межрегиональной динамике объекта  $i$ . В таблице 2.16 приведены значения рейтингов для объектов Оренбургской области и значения  $\delta_{i,t}$ .

Таблица 2.16 – Ранжирование городов и районов Оренбургской области по уровню демографической безопасности за период 2004-2007 гг.

Города и районы	Рейтинг						
	2004г.	2005г.		2006г.		2007г.	
		факт	$\delta_{i,2004}$	факт	$\delta_{i,2005}$	факт	$\delta_{i,2006}$
1	2	3	4	5	6	7	8
Абдулинский	32	27	+5	28	-1	30	-2
Адамовский	16	17	-1	16	+1	17	-1
Акбулакский	29	31	-2	33	-2	33	0
Александровский	38	38	0	37	+1	34	+3
Асекеевский	26	21	+5	21	0	26	-5
Беляевский	23	19	+4	27	-8	24	+3
Бугурусланский	31	34	-3	32	+2	29	+3
Бузулукский	17	18	-1	18	0	19	-1
Гайский	36	36	0	31	+5	32	-1
Грачевский	37	35	+2	36	-1	35	+1
Домбаровский	25	28	-3	25	+3	28	-3
Илекский	35	33	+2	38	-5	38	0
Кваркенский	41	39	+2	40	-1	42	-2
Красногвардейский	45	46	-1	47	-1	46	+1
Кувандыкский	47	44	+3	43	+1	44	-1
Курманаевский	24	25	-1	26	-1	23	+3
Матвеевский	46	45	+1	42	+3	45	-3
Новоорский	14	6	+8	9	-3	8	+1
Новосергиевский	33	37	-4	35	2	36	-1
Октябрьский	27	26	+1	22	+4	25	-3
Оренбургский	4	3	+1	1	+2	1	0
Первомайский	15	12	+3	15	-3	14	+1
Переволоцкий	34	32	+2	34	-2	37	-3
Пономаревский	42	40	+2	44	-4	43	+1
Сакмарский	30	29	+1	29	0	21	+8
Саракташский	22	22	0	19	+3	18	+1
Светлинский	43	47	-4	45	+2	39	+6
Северный	28	30	-2	30	0	31	-1
Соль-Илецкий	21	24	-3	24	0	27	-3
Сорочинский	20	23	-3	23	0	20	+3
Ташлинский	19	16	+3	17	-1	16	+1
Тоцкий	18	20	-2	20	0	22	-2
Тюльганский	40	43	-3	41	+2	40	+1
Шарлыкский	39	41	-2	39	+2	41	-2
Ясенский	44	42	+2	46	-4	47	-1
Абдулино	10	13	-3	11	+2	11	0
Бузулук	5	5	0	4	+1	6	-2
Бугуруслан	7	7	0	8	-1	7	+1
Гай	6	10	-4	6	+4	5	+1

Продолжение таблицы 2.16

1	2	3	4	5	6	7	8
Кувандык	13	14	-1	14	0	15	-1
Медногорск	12	15	-3	13	+2	13	0
Новотроицк	3	4	-1	5	-1	4	+1
Оренбург	1	1	0	2	-1	2	0
Орск	2	2	0	3	-1	3	0
Сорочинск	8	9	-1	7	+2	9	-2
Соль-Илецк	9	8	+1	10	-2	12	-2
Ясный	11	11	0	12	-1	10	+2

Наибольший рост рейтинга за 2005 год наблюдался у Новоорского района (+8 пунктов), за 2006 год – у Гайского района (+ 5 пунктов), за 2007 год – у Сакмарского района (+ 8 пунктов). Наибольшее падение рейтинга за 2005 год наблюдалось у г. Гай (- 4 пункта), а также у Новоорского и Светлинского районов (-4 пункта). За 2006 год – у Кувандыкского района (упал на 5 пунктов), за 2007 год – у Асекеевского района (- 5 пунктов).

За исследуемый период количество муниципальных образований, в которых наблюдалось повышение уровня демографической безопасности, немного превышало количество муниципальных образований, в которых наблюдалось снижение уровня демографической безопасности. За 2005, 2006гг. рейтинг не изменился у 8 административно-территориальных единиц, за 2007 год – у 7 административно-территориальных единиц [30].

### **2.8.2 Сравнительный анализ муниципальных образований Оренбургской области по уровню экономической безопасности региона**

Для сравнительного анализа муниципальных образований региона по уровню экономической безопасности построен интегральный показатель и проведено ранжирование, на основе методики предложенной С.А.Айвазяном, согласно которой, в качестве интегрального показателя, характеризующего некое латентное свойство, берется первая главная компонента, если выполняется требование о том, что уровень ее информативности превышает 55% [32].

Перед проведением компонентного анализа произведен переход от исходных показателей к унифицированным признакам. Оценка матрицы парных коэффициентов корреляции признаков приведена в приложении Ж. Гипотеза о незначимости матрицы парных коэффициентов корреляции отвергнута [33, 13]. Оценки собственных чисел и вклад главных компонент в суммарную дисперсию исходных признаков представлены в таблице 2.17.

Таблица 2.17 – Вклад главных компонент в суммарную дисперсию исходных признаков в 2007 году

Номер главной компоненты (i)	Собственные числа $\lambda_i$	Относительный вклад каждой главной компоненты в суммарную дисперсию	Накопленные значения собственных чисел	Накопленный относительный вклад
1	5,565	79,496	5,565	79,496
2	0,752	10,746	6,317	90,242
3	0,367	5,239	6,684	95,480
4	0,193	2,761	6,877	98,242
5	0,117	1,677	6,994	99,919
6	0,004	0,057	6,998	99,976
7	0,002	0,029	7,000	100,00

Уровень информативности первой главной компоненты составил 79,5%, следовательно, можно взять ее в качестве интегрального индикатора экономической безопасности.

Вектор унифицированных показателей связан с центрированно-нормированными главными компонентами следующим образом:

$$\bar{X}^* = A\bar{f}, \quad (2.29)$$

где  $\bar{f}$  - вектор центрированно-нормированных главных компонент;

$\bar{X}^*$  - вектор унифицированных показателей;

$A$  – матрица факторных нагрузок.

Значения элементов матрицы факторных нагрузок приведены в таблице 2.18. Тесную положительную связь интегральный показатель экономической безопасности региона имеет со следующими показателями: число зарегистрированных



иностранных работников на 1000 человек населения ( $X_1$ ), темп роста среднесписочной численности работников ( $X_4$ ), кредиторская задолженность ( $X_5$ ), оборот розничной торговли на душу населения ( $X_6$ ), средненоминальная заработная плата работников ( $X_7$ ) и инвестиции в основной капитал ( $X_8$ ). Таким образом, эти показатели положительно влияют на экономическую безопасность региона. Отрицательная тесная связь наблюдается между интегральным показателем и уровень безработицы ( $X_2$ )

Таблица 2.18 - Матрица факторных нагрузок

	F1	F2	F3	F4	F5	F6	F7
$X_1$	0,719	-0,351	-0,052	0,066	0,023	-0,013	0,034
$X_2$	-0,751	0,215	-0,457	-0,283	0,098	-0,022	-0,021
$X_4$	0,655	-0,275	0,249	-0,161	0,113	-0,032	0,008
$X_5$	0,750	-0,454	-0,129	0,22	0,112	0,024	0,018
$X_6$	0,708	-0,325	-0,099	0,017	0,151	0,017	-0,011
$X_7$	0,838	-0,244	-0,127	-0,185	-0,241	0,033	0,005
$X_8$	0,667	-0,372	-0,225	0,017	0,014	-0,011	-0,009

Результаты компонентного анализа за 2004-2006гг. приведены в приложении Ж.

Индивидуальные значения первой главной компоненты, характеризующие уровень экономической безопасности региона, и соответствующие ранги для муниципальных образований Оренбургской области в 2007 г. представлены в таблице 2.19.

Таблица 2.19 – Значение интегрального показателя и ранговая позиция объектов Оренбургской области по уровню экономической безопасности в 2007 году

Города и районы	Значение интегрального показателя	Ранг
1	2	3
г. Оренбург	5,9091	1
г. Бузулук	1,5370	2
Оренбургский р-н	0,6817	3
г. Орск	0,5697	4
г. Новотроицк	0,5470	5
г. Гай	0,2597	6

Продолжение таблицы 2.19

1	2	3
г. Бугуруслан	0,1097	7
г. Медногорск	0,0923	8
г. Сорочинск	0,0869	9
Новоорский	0,0735	10
Курманаевский р-н	0,0344	11
Первомайский р-н	0,0086	12
Грачевский р-н	-0,0217	13
Абдулинский	-0,1159	14
Ясный и р-н	-0,1267	15
Сакмарский р-н	-0,1314	16
Северный р-н	-0,2099	17
Светлинский р-н	-0,2391	18
Октябрьский р-н	-0,2562	19
Бузулукский	-0,2581	20
Тоцкий р-н	-0,2867	21
Новосергеевский р-н	-0,2876	22
Ташлинский р-н	-0,3134	23
Переволоцкий р-н	-0,3160	24
Краснагвардейский р-н	-0,3212	25
Адамовский	-0,3278	26
Соль-Илецк и р-н	-0,3444	27
Шарлыкский р-н	-0,3450	28
Пономаревский р-н	-0,3562	29
Кувандык и р-н	-0,3565	30
Саракташский р-н	-0,3675	31
Кваркенский р-н	-0,3717	32
Тюльганский р-н	-0,3717	33
Бугурусланский	-0,3771	34
Акбулакский	-0,3832	35
Беляевский	-0,3861	36
Александровский	-0,3916	37
Матвеевский р-н	-0,3936	38
Сорочинский р-н	-0,3986	39
Илекский р-н	-0,4116	40
Гайский	-0,4328	41
Асекеевский	-0,5046	42
Домбаровский р-н	-0,5056	43

Сравнительно более высокий уровень экономической безопасности имеют города области и Оренбургский район. По уровню экономической безопасности города области существенно опережают сельскую местность: здесь сосредоточены предприятия, предоставляющие торговые, социальные и финансовые услуги. Легко объяснить, почему сельский Оренбургский район имеет более высокий уровень экономической безопасности, по сравнению с экономической безопасностью дру-

гих сельских районов: примыкая к областному центру, он использует преимущество своего положения. Сравнительно более низкий уровень экономической безопасности наблюдается практически у всех сельских районов области. Сокращение производства сельскохозяйственной продукции и убыточность 60-70 % сельскохозяйственных предприятий объясняется медленной адаптацией сельского населения к рыночным отношениям, сокращением инвестиций в аграрный сектор экономики и ухудшением материально-технической базы сельского хозяйства [34].

Для выявления динамики рейтинговых позиций муниципальных образований построен интегральный показатель за 2004-2007 гг. В таблице 2.20 приведены значения рейтингов для муниципальных образований Оренбургской области

Таблица 2.20 – Ранжирование муниципальных образований Оренбургской области по уровню экономической безопасности за период 2004-2007 годы

Города и районы	Рейтинг						
	2004 г. ранг	2005 г.		2006 г.		2007 г.	
		ранг	$\delta_{i,2004}$	ранг	$\delta_{i,2005}$	ранг	$\delta_{i,2006}$
1	2	3	4	5	6	7	8
Абдулинский р-н	19	12	7	12	0	12	1
Адамовский р-н	28	26	2	26	-1	27	-2
Акбулакский р-н	20	31	-11	31	-1	32	2
Александровский р-н	37	37	0	37	0	37	3
Асекеевский р-н	33	36	-3	36	3	33	-3
Беляевский р-н	36	35	1	35	-3	38	-1
Бугурусланский	39	38	1	38	2	36	-4
Бузулукский р-н	21	28	-7	28	0	28	-4
г. Бугуруслан	7	8	-1	8	0	8	1
г. Бузулук	4	4	0	4	1	3	-1
г. Гай	8	7	1	7	0	7	-2
г. Медногорск	9	9	0	9	-1	10	4
г. Новотроицк	3	3	0	3	-1	4	1
г. Оренбург	1	1	0	1	0	1	0
г. Орск	2	2	0	2	0	2	0
г. Сорочинск	6	6	0	6	0	6	-4
Гайский р-н	41	42	-1	42	-1	43	0
Грачевский р-н	34	34	0	34	-1	35	2
Домбаровский р-н	35	40	-5	40	1	39	-2
Илекский р-н	38	33	5	33	-1	34	-1
Кваркенский р-н	40	39	1	39	-1	40	3
Краснагвардейский р-н	12	21	-9	21	-1	22	-1
Кувандык и р-н	10	11	-1	11	2	9	-3
Курманаевский р-н	27	32	-5	32	12	20	1
Матвеевский р-н	43	41	2	41	0	41	3
Новоорский р-н	14	18	-4	18	0	18	3
Новосергеевский р-н	15	13	2	13	-3	16	2
Октябрьский р-н	30	30	0	30	4	26	2
Оренбургский р-н	5	5	0	5	0	5	0
Первомайский р-н	16	15	1	15	4	11	-6
Переволоцкий р-н	22	19	3	19	2	17	-1

Продолжение таблицы 2.20

1	2	3	4	5	6	7	8
Пономаревский р-н	29	24	5	24	-6	30	10
Сакмарский р-н	23	20	3	20	1	19	-2
Саракташский р-н	25	16	9	16	-9	25	0
Светлинский р-н	26	27	-1	27	-4	31	0
Северный р-н	31	29	2	29	5	24	2
Соль- Илецк и р-н	13	10	3	10	-5	15	2
Сорочинский р-н	42	43	-1	43	1	42	0
Ташлинский р-н	32	22	10	22	-7	29	2
Тоцкий р-н	18	17	1	17	4	13	-3
Тюльганский р-н	17	23	-6	23	2	21	-7
Шарлыкский р-н	24	25	-1	25	2	23	-3
Ясный и р-н	11	14	-3	14	0	14	6

Наибольший рост рейтинга за 2004-2007 гг. наблюдался в Пономаревском (+9 пунктов), Абдулинском (+8 пунктов) и Северном (+9 пунктов) районах. Это объясняется тем, что в этих районах за период с 2004 по 2007 гг. наблюдалось повышение значений показателей, характеризующих экономическое развитие. Так, например, оборот розничной торговли на душу населения повысился примерно в три раза; сократилась, практически до нуля, задолженность организаций по заработной плате, уменьшился удельный вес убыточных организаций. Показатель числа зарегистрированных иностранных работников на 1000 человек населения в этих районах в несколько раз выше, по сравнению с аналогичным показателем других сельских районов. К тому же соседство с Самарской областью и Башкортостаном благоприятно отражается на развитии районов: оно создает возможности для производственного кооперирования, торговли и культурного обмена.

Наибольшее падение рейтинга за 2004-2007 гг. наблюдалось у районов, расположенных вдоль границы с Казахстаном (Акбулакский, Светлинский, Беляевский, Домбаровский, Илекский районы), что объясняется высоким уровнем безработицы, относительно низким фондом оплаты труда на одного рабочего, ростом удельного веса убыточных предприятий, низким оборотом розничной торговли на душу населения.

### 2.8.3 Сравнительный анализ муниципальных образований Оренбургской области по уровню инвестиционной привлекательности региона

Для сравнительного анализа муниципальных образований Оренбургской области по таким латентным категориям, как инвестиционный потенциал, активность и инвестиционный риск, построены интегральные показатели и проведено ранжирование, на основе методики, предложенной С.А. Айвазяном, согласно которой, в качестве интегрального показателя, характеризующего некое латентное свойство, берется первая главная компонента, если выполняется достаточно жесткое требование о том, что уровень ее информативности превышает 55%. В случае неработоспособности первой главной компоненты (уровень информативности составил меньше 0,55) проводится взвешенная процедура индивидуального рейтингования территорий в пространстве модифицированных первых главных компонент.

Уровень информативности первой главной компоненты, построенной по унифицированным показателям, характеризующим инвестиционный потенциал, приведен в таблице 2.21. Аналогичные таблицы для показателей инвестиционной активности и инвестиционного риска приведены в приложении И.

Таблица 2.21 – Уровень информативности первой главной компоненты, построенной по показателям, характеризующим инвестиционный потенциал

Год	2004	2005	2006	2007	2008
$I_1$	42,51	37,33	42,48	42,87	37,08

В связи с неработоспособностью первой главной компоненты (уровень информативности первой главной компоненты не превысил 0,55 ни в одном из случаев) была проведена взвешенная процедура индивидуального рейтингования территорий в пространстве модифицированных первых главных компонент. Количество выделенных главных компонент для обеспечения уровня информативности  $I(m) \geq 0,55$  представлено в таблице (2.22) и в приложении И.

Таблица 2.22 – Уровень информативности выделенных главных компонент, построенных по показателям, характеризующим инвестиционный потенциал

Собственные числа	Год				
	2004	2005	2006	2007	2008
$\lambda_1$	4,041	3,437	3,803	3,833	3,708
$\lambda_2$	1,858	1,949	1,428	1,221	1,431
$\lambda_3$	–	1,046	1,068	1,141	1,249
<b>I<sub>3</sub></b>	<b>0,5899</b>	<b>0,6433</b>	<b>0,6299</b>	<b>0,6195</b>	<b>0,6388</b>

Выделение двух главных компонент для 2004 года и трех главных компонент для 2005-2008 гг. достаточно для обеспечения заданного уровня информативности в 55%.

На следующем этапе было осуществлено разбиение анализируемого набора унифицированных частных критериев (регистрируемых статистических показателей)  $\tilde{x}^{(1)}, \dots, \tilde{x}^{(k)}$  на  $m_0$  однородных групп  $M_1, \dots, M_{m_0}$ , где принадлежность частных критериев к одной такой группе  $M_j$  определяется относительно высоким уровнем взаимной коррелированности. Матрицы факторных нагрузок приведены в приложении К.

После построения модифицированных первых главных компонент отдельно по частным критериям, входящим в каждую из групп  $M_1, \dots, M_{m_0}$  была определена удельная значимость («вес»  $\nu_j$ ) каждой из построенных интегральных характеристик и расставлены ранги  $R$  каждому муниципальному образованию.

При измерении межрегиональной динамики каждого конкретного объекта будем ориентироваться на динамику его положения (ранга) в ряду других объектов, то есть на величину:

$$\gamma_i(t) = R_i(t-1) - R_i(t), \quad (2.30)$$

где  $R_i(t)$  – ранг  $i$  – го муниципального образования в рейтинге муниципальных образований в момент времени  $t$ ;

$R_i(t-1)$  – ранг  $i$  – го муниципального образования в рейтинге муниципальных образований в момент времени  $t-1$ .

Очевидно, положительные значения  $\gamma_i(t)$  будут свидетельствовать о положительной динамике  $i$ -го муниципального образования: чем выше ранг, тем выше инвестиционная активность, потенциал и ниже риск.

В таблицах (2.23), (2.24), (2.25) приведены значения рейтингов для муниципальных образований Оренбургской области и значения  $\gamma_i(t)$ . Значения интегральных показателей приведены в приложении Л.

Таблица 2.23 – Ранжирование муниципальных образований Оренбургской области по уровню инвестиционного потенциала за период с 2004 по 2008гг.

Города и районы	Рейтинг								
	2004	2005		2006		2007		2008	
	ранг	ранг	$\gamma_i(2005)$	ранг	$\gamma_i(2006)$	ранг	$\gamma_i(2007)$	ранг	$\gamma_i(2008)$
1	2	3	4	5	6	7	8	9	10
Абдулинский	45	42	+3	45	-3	42	+3	45	-3
Адамовский	12	18	-6	22	-4	20	+2	23	-3
Акбулакский	22	23	-1	20	+3	30	-10	32	-2
Александровский	30	40	-10	38	+2	40	-2	38	+2
Асекеевский	23	25	-2	26	-1	32	-6	39	-7
Беляевский	41	35	+6	43	-8	38	+5	36	+2
Бугурусланский	28	22	+6	32	-10	25	+7	22	+3
Бузулукский	31	38	-7	21	+17	27	-6	21	+6
Гайский	39	44	-5	35	+9	41	-6	41	0
Грачевский	19	28	-9	29	-1	33	-4	40	-7
Домбаровский	25	33	-8	43	-10	43	0	46	-3
Илекский	32	25	+7	28	-3	36	-8	42	-6
Кваркенский	35	32	+3	24	+8	35	-11	28	+7
Красногвардейский	34	37	-3	30	+7	29	+1	33	-4
Кувандыкский	38	46	-8	40	+6	45	-5	47	-2
Курманаевский	29	28	+1	33	-5	39	-6	43	-4
Матвеевский	33	36	-3	39	-3	28	+11	30	-2
Новоорский	24	24	0	16	+8	17	-1	15	+2
Новосергиевский	20	14	+6	13	+1	14	-1	13	+1
Октябрьский	17	19	-2	27	-8	24	+3	26	-2
Оренбургский	18	11	+7	14	-3	13	+1	10	+3
Первомайский	37	34	+3	23	+11	31	-8	35	-4
Переволоцкий	36	39	-3	37	+2	37	0	31	+6
Пономаревский	46	45	+1	46	-1	34	+12	29	+5
Сакмарский	15	27	-12	34	-7	26	+8	17	+9
Саракташский	21	16	+5	15	+1	16	-1	18	-2
Светлинский	27	20	+7	18	+2	22	-4	16	+6
Северный	44	41	+3	47	-6	18	+29	25	-7
Соль-Илецкий	16	17	-1	25	-8	21	+4	20	+1
Сорочинский	26	30	-4	41	-11	46	-5	37	+9
Ташлинский	10	15	-5	17	-2	15	2	19	-4
Тоцкий	42	43	-1	31	+12	44	-13	34	+10
Тюльганский	43	30	+13	19	+11	19	0	27	-8
Шарлыкский	40	21	+19	36	-15	22	+14	24	-2
Ясенский	47	47	0	42	+5	47	-5	43	4

Продолжение таблицы 2.23

1	2	3	4	5	6	7	8	9	10
г.Абдулино	14	10	+4	10	0	10	0	11	-1
г.Бугуруслан	7	8	-1	6	2	7	-1	8	-1
г.Бузулук	6	5	+1	3	2	3	0	5	-2
г.Гай	3	2	+1	7	-5	4	+3	4	0
г.Кувандык	9	9	0	9	0	9	0	9	0
г.Медногорск	4	6	-2	2	4	6	-4	7	-1
г.Новотроицк	1	4	-3	4	0	2	+2	3	-1
г.Оренбург	2	1	+1	1	0	1	0	1	0
г.Орск	5	3	+2	8	-5	5	3	2	+3
г.Соль-Илецк	11	13	-2	12	+1	11	1	12	-1
г.Сорочинск	13	12	+1	11	+1	12	-1	14	-2
г.Ясный	8	7	+1	5	+2	8	-3	6	2

Ведущие позиции среди всех муниципальных образований области по уровню инвестиционного потенциала занимают города Оренбургской области. Они характеризуются значительным развитием промышленного производства, которое представлено практически всеми отраслями: топливной, электроэнергетической, металлургической, химической и нефтехимической, машиностроением, легкой и пищевой. Наибольшее развитие получили предприятия топливной промышленности, которые осуществляют добычу, переработку и продажу нефти и газа, как на территории области, так и за ее пределами. Крупнейшие предприятия топливной промышленности располагаются, в основном в таких городах, как Оренбург, Бузулук, Бугуруслан, которые имеют сравнительно высокий уровень инвестиционного потенциала. Сравнительно высоким уровнем инвестиционного потенциала характеризуются такие районы, как Оренбургский, Новосергиевский, Новоорский. Наличие в Новоорском районе месторождений медно-цинковых руд «Барсучий лог», глин (Кумакское, Новоорское, Приморское), строительного камня (Мусогоатское, Новоорское, Ульяновское), песчано-гравийных смесей (Озёрное, Оринское и др.), известняков (Ириклинское) обусловило развитие крупных промышленных предприятий – Обоганительная фабрика ЗАО «Ормет» с карьером «Барсучий лог», которое занимается добычей руды и выпуском медно-цинкового концентрата. В Новосергиевском районе получили развитие пищевая промышленность и машиностроение.



Такие районы, как Адамовский, Бугурусланский, Новоорский, Новосергиевский Октябрьский Саракташский Светлинский Соль–Илецкий Ташлинский обладают средним потенциалом, здесь сравнительно высокие значения таких показателей, как удельный вес населения моложе трудоспособного возраста, среднегодовая численность работников, занятых в сельскохозяйственном производстве; уровень рентабельности реализованной продукции сельского хозяйства в сельскохозяйственных организациях (в процентах). Сравнительно низким потенциалом обладают такие районы, как Абдулинский, Александровский, Беляевский, Гайский, Кувандыкский, Переволоцкий, Тоцкий, Ясененский.

Таблица 2.24 – Ранжирование муниципальных образований Оренбургской области по уровню инвестиционной активности за период с 2004 по 2008гг.

Города и районы	Рейтинг								
	2004	2005		2006		2007		2008	
	ранг	ранг	$\gamma_i(2003)$	ранг	$\gamma_i(2004)$	ранг	$\gamma_i(2005)$	ранг	$\gamma_i(2006)$
1	2	3	4	5	6	7	8	9	10
Абдулинский	47	37	+10	47	-10	44	+3	45	-1
Адамовский	22	12	+10	27	-15	33	-6	16	+17
Акбулакский	21	16	+5	24	-8	13	+11	4	+9
Александровский	41	44	-3	39	+5	41	-2	41	0
Асекеевский	40	13	+27	6	+7	17	-11	21	-4
Беляевский	35	19	+16	30	-11	26	+4	31	-5
Бугурусланский	26	32	-6	38	-6	34	+4	38	-4
Бузулукский	17	29	-12	37	-8	38	-1	32	+6
Гайский	46	36	+10	31	+5	24	+7	24	0
Грачевский	28	20	+8	26	-6	21	+5	29	-8
Домбаровский	18	18	0	25	-7	30	-5	34	-4
Илекский	36	9	+27	18	-9	25	-7	20	+5
Кваркенский	45	47	-2	44	+3	42	+2	40	+2
Красногвардейский	42	38	+4	32	+6	36	-4	27	+9
Кувандыкский	44	42	+2	45	-3	47	-2	37	+10
Курманаевский	37	40	-3	34	+6	31	+3	36	-5
Матвеевский	23	33	-10	22	+11	35	-13	28	+7
Новоорский	13	8	+5	9	-1	11	-2	6	+5
Новосергиевский	24	10	+14	10	0	14	-4	13	+1
Октябрьский	27	22	+5	7	+15	12	-5	10	+2
Оренбургский	9	3	+6	1	+2	4	-3	2	+2
Первомайский	1	2	-1	4	-2	39	-35	30	+9
Переволоцкий	39	14	+25	8	+6	23	-15	18	+5
Пономаревский	43	25	+18	16	+9	22	-6	19	+3
Сакмарский	20	11	+9	12	-1	6	+6	12	-6
Саракташский	34	27	+7	20	+7	15	+5	14	+1
Светлинский	14	43	-29	42	+1	32	+10	42	-10
Северный	32	15	+17	19	-4	28	-9	26	+2

Продолжение таблицы 2.24

1	2	3	4	5	6	7	8	9	10
Соль-Илецкий	25	41	-16	46	-5	43	+3	47	-4
Сорочинский	30	39	-9	36	+3	18	+18	33	-15
Ташлинский	16	21	-5	13	+8	9	+4	17	-8
Тоцкий	29	46	-17	40	+6	46	-6	43	+3
Тюльганский	38	24	+14	43	-19	37	+6	35	+2
Шарлыкский	33	30	+3	29	+1	27	+2	25	+2
Ясненский	31	45	-14	41	+4	45	-4	46	-1
г.Абдулино	19	35	-16	11	+24	40	-29	11	+29
г.Бугуруслан	6	6	0	14	-8	8	+6	8	0
г.Бузулук	7	7	0	2	+5	1	+1	3	-2
г.Гай	8	17	-9	17	0	5	+12	15	-10
г.Кувандык	15	28	-13	28	0	29	-1	22	+7
г.Медногорск	11	34	-23	35	-1	20	+15	39	-19
г.Новотроицк	4	4	0	5	-1	2	+3	7	-5
г.Оренбург	2	5	-3	3	+2	7	-4	1	+6
г.Орск	3	26	-23	23	+3	10	+13	23	-13
г.Соль-Илецк	10	23	-13	15	+8	16	-1	9	+7
г.Сорочинск	5	1	+4	21	-20	3	+18	5	-2
г.Ясный	12	31	-19	33	-2	19	+14	44	-25

Ведущие позиции среди всех муниципальных образований области по уровню инвестиционной активности занимают такие города, как Бузулук, Оренбург, Новотроицк, Сорочинск. Наличие развитых крупных промышленных предприятий в этих городах способствует постоянному притоку инвестиций, что обуславливает рост таких показателей как инвестиции в основной капитал, объем промышленной продукции, ввод в действие жилых домов. Сравнительно высокий уровень инвестиционной активности наблюдается в таких районах, как Оренбургский, Новосергиевский, Новоорский, Первомайский (2004-2008гг.). Первомайский район известный своими крупными месторождениями нефти, горючих сланцев, газа, фосфитов и строительного сырья долгое время имел самый высокий в области объем инвестиций на душу населения. В результате снижения объемов добычи нефти, в районе произошло сокращение инвестиций в основной капитал и падение рейтинга по уровню инвестиционной активности.

В целом, можно отметить, что часть районов (Акбулакский, Новоорский, Новосергиевский, Октябрьский, Сакмарский, Ташлинский) обладает средней инвестиционной активностью, здесь сравнительно высокие значения таких показателей, как производства скота и птицы на душу населения, ввод в действие жилых

домов. Такие районы, как Абдулинский, Александровский, Кваркенский, Кувандыкский, Тоцкий обладали низкой инвестиционной активностью на протяжении исследуемого периода.

Ранжирование муниципальных единиц по уровню инвестиционного риска представлено в таблице 2.25, сравнительно низкий уровень риска наблюдается в городах – Бузулуке, Оренбурге, Сорочинске и в Оренбургском районе, а сравнительно высокий уровень риска в Абдулинском, Асекеевском, Бузулукском, Курманаевском, Пономаревском, Соль-Илецком, Тюльганском районах. Рост рейтинга в 2008г. в Красногвардейском районе на 25 пунктов, возможно, связан со снижением просроченной кредиторской задолженности в 8 раз. В Домбаровском и Октябрьском районах снижение рейтинга инвестиционного риска на 32 и 26 пунктов соответственно, возможно связано с увеличением убыточных предприятий и просроченной кредиторской задолженностью (в 4 раза и в 9 раз соответственно).

Таблица 2.25 – Ранжирование муниципальных образований Оренбургской области по уровню инвестиционного риска за период с 2004 по 2008гг.

Города и районы	Рейтинг								
	2004	2005		2006		2007		2008	
	ранг	ранг	$\gamma_i(2005)$	ранг	$\gamma_i(2006)$	ранг	$\gamma_i(2007)$	ранг	$\gamma_i(2008)$
1	2	3	4	5	6	7	8	9	10
Абдулинский	39	44	-5	40	+4	40	0	39	+1
Адамовский	8	6	+2	12	-6	28	-16	21	+7
Акбулакский	16	26	-10	23	+3	38	-15	35	+3
Александровский	13	8	+5	11	-3	11	0	14	-3
Асекеевский	44	38	+6	22	+16	39	-17	36	+3
Беляевский	26	22	+4	3	+19	33	-30	33	0
Бугурусланский	24	30	-6	37	-7	31	+6	32	-1
Бузулукский	34	35	-1	31	+4	43	-12	40	+3
Гайский	45	36	+9	9	+27	10	-1	9	+1
Грачевский	31	34	-3	38	-4	47	-9	41	+6
Домбаровский	17	10	+7	4	+6	6	-2	38	-32
Илекский	43	45	-2	43	+2	32	+11	46	-14
Кваркенский	38	23	+15	29	-6	21	+8	37	-16
Красногвардейский	23	21	+2	34	-13	37	-3	12	+25
Кувандыкский	21	25	-4	27	-2	26	1	31	-5
Курманаевский	42	46	-4	33	+13	45	-12	45	0
Матвеевский	37	41	-4	30	+11	30	0	19	+11
Новоорский	11	16	-5	39	-23	19	+20	24	-5
Новосергиевский	10	12	-2	7	+5	22	-15	20	+2
Октябрьский	25	17	+8	8	+9	2	+6	28	-26
Оренбургский	3	3	0	19	-16	9	+10	11	-2
Первомайский	28	27	+1	25	+2	25	0	22	+3

Продолжение таблицы 2.25

1	2	3	4	5	6	7	8	9	10
Переволоцкий	27	14	+13	5	+9	24	-19	10	+14
Пономаревский	46	47	-1	44	+3	46	-2	44	+2
Сакмарский	4	15	-11	15	0	3	+12	1	+2
Саракташский	20	28	-8	26	+2	34	-8	17	+17
Светлинский	18	19	-1	6	+13	17	-11	27	-10
Северный	32	32	0	35	-3	44	-9	29	+15
Соль-Илецкий	35	37	-2	46	-9	36	+10	42	-6
Сорочинский	40	42	-2	42	0	35	+7	26	+9
Ташлинский	6	2	+4	1	+1	12	-11	30	-18
Тоцкий	36	39	-3	32	+7	29	+3	23	+6
Тюльганский	47	43	+4	45	-2	23	+22	43	-20
Шарлыкский	33	33	0	47	-14	41	+6	25	+16
Ясненский	41	29	+12	41	-12	42	-1	47	-5
г.Абдулино	29	40	-11	24	+16	20	+4	16	+4
г.Бугуруслан	14	13	+1	14	-1	13	+1	4	+9
г.Бузулук	12	11	+1	17	-6	1	+16	6	-5
г.Гай	19	31	-12	28	-3	14	+14	3	+11
г.Кувандык	9	5	+4	21	-16	7	+14	5	2
г.Медногорск	2	7	-5	18	-11	8	+10	8	0
г.Новотроицк	22	18	+4	20	-2	18	+2	18	0
г.Оренбург	5	4	+1	13	-9	4	+9	7	-3
г.Орск	15	24	-9	36	-12	27	+9	13	+14
г.Соль-Илецк	30	20	+10	16	+4	16	0	15	+1
г.Сорочинск	7	9	-2	10	-1	5	+5	2	+3
г.Ясный	1	1	0	2	-1	15	-13	34	-19

По результатам рейтинговых оценок муниципальных образований Оренбургской области по уровням инвестиционного потенциала, активности и риска можно сделать следующие выводы. Лидерами по уровню инвестиционной привлекательности являются следующие образования: г.Оренбург, г.Бузулук, Оренбургский район. Последние места по уровню инвестиционной привлекательности занимают Абдулинский, Кувандыкский, Курманаевский, Тюльганский районы, так как имеют низкие уровни инвестиционного потенциала, активности и высокий уровень риска. Относительно остальных образований затруднительно сделать однозначный вывод об уровне инвестиционной привлекательности. В связи с этим, важным вопросом является построение сводного интегрального показателя, характеризующего инвестиционную привлекательность.

Используя результаты ранжирования муниципальных образований по инвестиционному потенциалу, активности и риску, выявлена теснота и направление

связи между указанными показателями с помощью коэффициентов ранговой корреляции Спирмена и Кендалла (таблица 2.26).

Таблица 2.26 – Результаты расчета коэффициентов ранговой корреляции Спирмена и Кендалла

Коэффициент	Результаты расчетов				
	2004	2005	2006	2007	2008
$r(rA(t), r\Pi(t))$ - коэффициент ранговой корреляции Спирмена между активностью и потенциалом	$r=0,751$ $t$ -статистика=7,623 $p=0,000$	$r=0,429$ $t=3,185$ $p=0,003$	$r=0,385$ $t=2,800$ $p=0,007$	$r=0,614$ $t=5,219$ $p=0,000$	$r=0,505$ $t=3,920$ $p=0,000$
$\tau(rA(t), r\Pi(t))$ - коэффициент ранговой корреляции Кендалла между активностью и потенциалом	$\tau=0,576$ $z$ -статистика=5,713 $p=0,000$	$\tau=0,296$ $z=2,939$ $p=0,003$	$\tau=0,257$ $z=2,551$ $p=0,011$	$\tau=0,457$ $z=4,532$ $p=0,000$	$\tau=0,374$ $z=3,707$ $p=0,000$
$r(rA(t), rP(t))$ - коэффициент ранговой корреляции Спирмена между активностью и риском	$r=-0,640$ $t$ -статистика=-5,586 $p=0,000$	$r=-0,374$ $t=-2,705$ $p=0,009$	$r=-0,337$ $t=-2,404$ $p=0,020$	$r=-0,459$ $t=-3,466$ $p=0,001$	$r=-0,536$ $t=-4,261$ $p=0,001$
$\tau(rA(t), rP(t))$ - коэффициент ранговой корреляции Кендалла между активностью и риском	$\tau=-0,451$ $z$ -статистика=-4,466 $p=0,000$	$\tau=-0,260$ $z=-2,577$ $p=0,009$	$\tau=-0,216$ $z=-2,137$ $p=0,033$	$\tau=-0,304$ $z=-3,017$ $p=0,003$	$\tau=-0,382$ $z=-3,787$ $p=0,000$
$r(r\Pi(t), rP(t))$ - коэффициент ранговой корреляции Спирмена между потенциалом и риском	$r=-0,689$ $t$ -статистика=-6,376 $p=0,000$	$r=-0,493$ $t=-3,796$ $p=0,000$	$r=-0,293$ $t=-2,053$ $p=0,046$	$r=-0,474$ $t=-3,609$ $p=0,001$	$r=-0,597$ $t=-4,492$ $p=0,000$
$\tau(r\Pi(t), rP(t))$ - коэффициент ранговой корреляции Кендалла между потенциалом и риском	$\tau=-0,482$ $z$ -статистика=-4,778 $p=0,000$	$\tau=-0,330$ $z=-3,269$ $p=0,001$	$\tau=-0,204$ $z=-2,018$ $p=0,044$	$\tau=-0,341$ $z=-3,376$ $p=0,001$	$\tau=-0,426$ $z=-4,22$ $p=0,000$

Согласно таблице 2.26, между показателями инвестиционной активности и инвестиционного потенциала существует значимая положительная связь, между инвестиционной активностью и инвестиционным риском, а также между инвестиционным потенциалом и инвестиционным риском наблюдается отрицательная связь.

В связи с этим на основе построенных интегральных показателей, характеризующих отдельные составляющие инвестиционной привлекательности, построен сводный интегральный показатель, являющийся неубывающей функцией от ин-

вестиционного потенциала  $\rho(\Pi)$  и инвестиционной активности  $\rho(A)$  и невозрастающей от инвестиционного риска  $\rho(P)$ :

$$U = v_1\rho(\Pi) + v_2\rho(A) - v_3\rho(P), \quad (2.31)$$

где  $U$  – значение сводного интегрального показателя, характеризующего уровень инвестиционной привлекательности;

$v_1, v_2, v_3$  – вес каждой из построенной латентной категории (потенциала, активности и риска), найденный, как евклидово расстояние от объекта до эталона, по исходным признакам:

$$v_{il} = \sqrt{\sum_{j=1}^k (\tilde{x}_{ij} - \tilde{x}^{(э)})^2}, \quad i = 1 \dots n, \quad (2.32)$$

где  $\tilde{x}_{ij}$  – унифицированные значения  $j$ -го признака для  $i$ -го объекта;

$\tilde{x}^{(э)}$  – «наихудшие» значения унифицированных признаков, принадлежащие объекту-эталону;

Ранжирование муниципальных образований по уровню инвестиционной привлекательности представлено в таблице 2.27.

Таблица 2.27 – Ранжирование муниципальных образований по уровню инвестиционной привлекательности (2008г.)

Города и районы	Инвестиционный потенциал			Инвестиционная активность			Инвестиционный риск			Инвестиционная привлекательность $U = v_1\rho(\Pi) + v_2\rho(A) - v_3\rho(P)$	
	вес	$\rho(\Pi)$	ранг	вес	$\rho(A)$	ранг	вес	$\rho(P)$	ранг	$U$	ранг
1	2	3	4	5	6	7	8	9	10	11	12
г.Оренбург	1,916	0,967	1	1,369	0,766	1	0,158	0,08	7	2,889	1
г.Орск	1,365	0,955	2	0,461	0,186	23	0,267	0,139	13	1,352	4
г.Новотроицк	1,254	0,896	3	0,838	0,376	7	0,372	0,187	18	1,369	3
г.Гай	0,802	0,88	4	0,564	0,264	15	0,093	0,035	3	0,851	8
г.Бузулук	1,178	0,837	5	1,269	0,581	3	0,198	0,076	6	1,708	2
г.Ясный	1,062	0,779	6	0,112	0,055	44	0,706	0,368	34	0,573	12
г.Медногорск	1,094	0,778	7	0,203	0,097	39	0,236	0,094	8	0,848	9
г.Бугуруслан	1,038	0,739	8	0,693	0,365	8	0,099	0,041	4	1,016	6

Продолжение таблицы 2.27

1	2	3	4	5	6	7	8	9	10	11	12
г.Кувандык	1,027	0,608	9	0,252	0,188	22	0,192	0,074	5	0,658	10
Оренбургский	0,816	0,587	10	1,018	0,749	2	0,244	0,112	11	1,214	5
г.Абдулино	0,813	0,552	11	0,426	0,325	11	0,365	0,182	16	0,521	13
г.Соль–Илецк	1,043	0,502	12	0,456	0,346	9	0,373	0,164	15	0,62	11
Новосергиевск.	0,786	0,496	13	0,407	0,307	13	0,42	0,218	20	0,423	15
г.Сорочинск	0,786	0,487	14	1,09	0,472	5	0,048	0,025	2	0,896	7
Новоорский	0,75	0,482	15	0,565	0,387	6	0,585	0,251	24	0,433	14
Светлинский	0,745	0,364	16	0,111	0,076	42	0,633	0,291	27	0,095	25
Сакмарский	0,646	0,357	17	0,407	0,309	12	0,051	0,025	1	0,355	17
Саракташский	0,696	0,345	18	0,353	0,268	14	0,37	0,183	17	0,267	18
Ташлинский	0,654	0,342	19	0,338	0,25	17	0,611	0,319	30	0,114	24
Соль–Илецкий	0,633	0,318	20	0,038	0,029	47	0,972	0,51	42	-0,294	40
Бузулукский	0,624	0,311	21	0,17	0,129	32	0,989	0,5	40	-0,279	39
Бугурусланский	0,631	0,308	22	0,143	0,101	38	0,635	0,325	32	0,002	31
Адамовский	0,567	0,3	23	0,335	0,252	16	0,552	0,231	21	0,127	23
Шарлыкский	0,55	0,297	24	0,232	0,174	25	0,548	0,259	25	0,062	28
Северный	0,548	0,296	25	0,223	0,17	26	0,708	0,305	29	-0,016	32
Октябрьский	0,545	0,287	26	0,481	0,33	10	0,602	0,294	28	0,138	20
Тюльганский	0,545	0,282	27	0,15	0,111	35	0,85	0,524	43	-0,275	38
Кваркенский	0,544	0,28	28	0,127	0,096	40	0,691	0,402	37	-0,114	34
Пономаревский	0,544	0,261	29	0,313	0,239	19	1,009	0,525	44	-0,314	42
Матвеевский	0,528	0,257	30	0,212	0,162	28	0,404	0,21	19	0,085	26
Переволоцкий	0,54	0,249	31	0,315	0,239	18	0,249	0,109	10	0,183	19
Акбулакский	0,525	0,247	32	1,003	0,522	4	0,738	0,386	35	0,368	16
Красногвардейск	0,521	0,237	33	0,224	0,168	27	0,229	0,129	12	0,131	21
Тоцкий	0,518	0,233	34	0,086	0,065	43	0,515	0,239	23	0,003	30
Первомайский	0,515	0,231	35	0,193	0,148	30	0,388	0,236	22	0,056	29
Беляевский	0,514	0,229	36	0,187	0,14	31	0,75	0,362	33	-0,127	35
Сорочинский	0,503	0,223	37	0,171	0,128	33	0,549	0,276	26	-0,018	33
Александровск.	0,496	0,221	38	0,126	0,096	41	0,394	0,146	14	0,064	27
Асекеевский	0,493	0,215	39	0,281	0,213	21	0,778	0,394	36	-0,141	36
Грачевский	0,492	0,214	40	0,196	0,149	29	1,003	0,510	41	-0,377	43
Гайский	0,431	0,212	41	0,304	0,180	24	0,194	0,094	9	0,128	22
Илекский	0,425	0,19	42	0,280	0,213	20	1,066	0,688	46	-0,593	45
Курманаевский	0,265	0,189	43	0,145	0,11	36	1,139	0,625	45	-0,646	46
Ясненский	0,397	0,189	43	0,106	0,034	46	1,426	0,752	47	-0,993	47
Абдулинский	0,217	0,177	45	0,057	0,044	45	0,96	0,481	39	-0,42	44
Домбаровский	0,243	0,149	46	0,154	0,117	34	0,853	0,409	38	-0,295	41
Кувандыкский	0,217	0,138	47	0,140	0,105	37	0,619	0,323	31	-0,155	37
Вес	x	$v_1 = 0,295$ $v_2 = 0,341$ $v_3 = 0,364$	x	x	$v_1 = 0,586$ $v_2 = 0,414$	x	x	$v_1 = 0,551$ $v_2 = 0,449$	x	x	x

Далее проведен анализ результатов ранжирования в соответствии с построенными интегральными показателями и выделены муниципальные образования со сравнительно высоким, средним и низким уровнем инвестиционной привлекательности (таблица 2.28).

Таблица 2.28 – Результаты группировки муниципальных образований Оренбургской области по инвестиционной привлекательности (2008г.)

Города и районы	Инвестиционный потенциал			Инвестиционная активность			Инвестиционный риск			ИП $U = v_1\rho(II) + v_2\rho(A) - v_3\rho(P)$	
	вес	$\rho(II)$	ранг	вес	$\rho(A)$	ранг	вес	$\rho(P)$	ранг	$U$	ранг
	Высокий (ранг 1–15)			Высокая (ранг 1–15)			Низкий (ранг 1–15)				
г.Оренбург	1,916	0,967	1	1,369	0,766	1	0,158	0,08	7	2,889	1
г.Гай	0,802	0,88	4	0,564	0,264	15	0,093	0,035	3	0,851	8
г.Бузулук	1,178	0,837	5	1,269	0,581	3	0,198	0,076	6	1,708	2
г.Бугуруслан	1,038	0,739	8	0,693	0,365	8	0,099	0,041	4	1,016	6
Оренбургский	0,816	0,587	10	1,018	0,749	2	0,244	0,112	11	1,214	5
г.Соль–Илецк	1,043	0,502	12	0,456	0,346	9	0,373	0,164	15	0,62	11
г.Сорочинск	0,786	0,487	14	1,09	0,472	5	0,048	0,025	2	0,896	7
	Высокий (ранг 1–15)			Высокая (ранг 1–15)			Средний (ранг 16–31)				
г.Новотроицк	1,254	0,896	3	0,838	0,376	7	0,372	0,187	18	1,369	3
г.Абдулино	0,813	0,552	11	0,426	0,325	11	0,365	0,182	16	0,521	13
Новоорский	0,75	0,482	15	0,565	0,387	6	0,585	0,251	24	0,433	14
Новосергиевск.	0,786	0,496	13	0,407	0,307	13	0,42	0,218	20	0,423	15
	Высокий (ранг 1–15)			Средняя (ранг 16–31)			Низкий (ранг 1–15)				
г.Орск	1,365	0,955	2	0,461	0,186	23	0,267	0,139	13	1,352	4
г.Кувандык	1,027	0,608	9	0,252	0,188	22	0,192	0,074	5	0,658	10
	Средний (ранг 16–31)			Низкая (ранг 32–47)			Высокий (ранг 32–47)				
Соль–Илецкий	0,633	0,318	20	0,038	0,029	47	0,972	0,51	42	–0,294	40
Бузулукский	0,624	0,311	21	0,17	0,129	32	0,989	0,5	40	–0,279	39
Тюльганский	0,545	0,282	27	0,15	0,111	35	0,85	0,524	43	–0,275	38
Кваркенский	0,544	0,28	28	0,127	0,096	40	0,691	0,402	37	–0,114	34
	Низкий (ранг 32–47)			Средняя (ранг 16–31)			Высокий (ранг 32–47)				
Беляевский	0,514	0,229	36	0,187	0,14	31	0,75	0,362	33	–0,127	35
Асекеевский	0,493	0,215	39	0,281	0,213	21	0,778	0,394	36	–0,141	36
Грачевский	0,492	0,214	40	0,196	0,149	29	1,003	0,51	41	–0,377	43
Илекский	0,425	0,19	42	0,28	0,213	20	1,066	0,688	46	–0,593	45
	Низкий (ранг 32–47)			Низкая (ранг 32–47)			Высокий (ранг 32–47)				
Курманаевский	0,265	0,189	43	0,145	0,11	36	1,139	0,625	45	–0,646	46
Ясненский	0,397	0,189	43	0,106	0,034	46	1,426	0,752	47	–0,993	47
Абдулинский	0,217	0,177	45	0,057	0,044	45	0,96	0,481	39	–0,42	44
Домбаровский	0,243	0,149	46	0,154	0,117	34	0,853	0,409	38	–0,295	41
	Средний (ранг 16–31)			Высокая (ранг 1–15)			Средний (ранг 16–31)				
Саракташский	0,696	0,345	18	0,353	0,268	14	0,37	0,183	17	0,267	18
Октябрьский	0,545	0,287	26	0,481	0,33	10	0,602	0,294	28	0,138	20
	Средний (ранг 16–31)			Средняя (ранг 16–31)			Средний (ранг 16–31)				
Ташлинский	0,654	0,342	19	0,338	0,25	17	0,611	0,319	30	0,114	24
Адамовский	0,567	0,3	23	0,335	0,252	16	0,552	0,231	21	0,127	23
Шарлыкский	0,55	0,297	24	0,232	0,174	25	0,548	0,259	25	0,062	28
Матвеевский	0,528	0,257	30	0,212	0,162	28	0,404	0,21	19	0,085	26
...	...	...	...	...	...	...	...	...	...	...	...



Анализ результатов ранжирования показал, что сравнительно более высокие ранги по инвестиционному потенциалу определяют сравнительно более высокую инвестиционную привлекательность объектов. К таковым относятся города Оренбургской области, поскольку им соответствуют более высокие значения показателей, характеризующих финансовый и потребительский потенциал, здесь сосредоточены предприятия, предоставляющие торговые, финансовые услуги и осуществляющие строительную и промышленную деятельность, и районы (Оренбургский, Новоорский и Новосергиевский), которые примыкают к крупнейшим городам области и используют преимущества своего положения.

Объекты, которые характеризуются сравнительно высоким инвестиционным риском, в большей степени имеют низкую инвестиционную привлекательность. Это районы, расположенные на окраине региона и характеризующиеся сравнительно высокими значениями таких показателей, как удельный вес убыточных предприятий и организаций, задолженность организаций по заработной плате.

Инвестиционная активность оказывается существенной для инвестиционной привлекательности объектов со средними показателями инвестиционного потенциала и риска. К этим группам относятся в основном районы, отраслевой специализацией которых является сельское хозяйство.

Ранжирование муниципальных образований Оренбургской области по уровню инвестиционной привлекательности за 2004-2008 гг. представлено в таблице Л.5.

По результатам рейтинговой оценки городов и районов Оренбургской области можно сделать следующие выводы: лидерами по уровню инвестиционной привлекательности являются такие города, как Оренбург, Бузулук и Оренбургский район. Ряд районов (Адамовский, Александровский, Октябрьский, Саракташский), где получило развитие сельское хозяйство, характеризуются показателями, обеспечивающими средний уровень инвестиционной привлекательности. Сельское хозяйство – жизненно важная отрасль в Оренбургской области, оно обеспечивает работой большинство жителей села, так как каждый пятый трудящийся Оренбуржья работает в сельском хозяйстве. Основными проблемами здесь являются – низкий

уровень производительности труда, низкий уровень оплаты и, как считают многие специалисты-аграрии, заниженные цены на продукцию. Для развития сельского хозяйства необходимо обновление основных фондов, создание условий для повышения конкурентоспособности сельскохозяйственной продукции, улучшения ее качества.

Сравнительно низким уровнем инвестиционной привлекательности обладают следующие районы – Абдулинский, Бузулукский, Курманаевский, Пономаревский, Соль-Илецкий, Тюльганский, Ясенский. Это либо районы, как было уже отмечено, расположенные на окраине региона, либо образования со значительными запасами полезных ископаемых осадочного происхождения (нефть, уголь), но в результате того, что ближайшие города оттягивают на себя трудовые и финансовые ресурсы, к тому же некоторые виды ресурсов топливной промышленности не пользуются спросом, эти районы являются непривлекательными.

Для сравнения проведено построение интегрального показателя инвестиционной привлекательности по методике, предложенной С.А. Айвазяном [6]. Интуитивное экспертное восприятие сводной характеристики можно представить в виде модели латентного показателя.

Для получения матрицы парных сравнений  $\gamma_m$  были использованы результаты разбиения объектов на классы с использованием методов кластерного анализа и нейросетевой классификации, полученных ранее. Таким образом, три вида разбиений (метод Уорда, k–средних, нейросетевая классификация) позволили получить матрицы парных сравнений, эквивалентные оценкам экспертов (приложение М). Располагая «экспертной» и статистической информацией, были оценены коэффициенты целевой функции.

Оценка влияния факторов на уровень инвестиционной привлекательности для 2008 года имеет вид:

$$f_{2008}(X, \theta) = 20,1y_1 + 30,8y_2 + 35,0y_3 + 26,8y_4 + 27,5x_{1.1} + 25,9x_{1.2} + 29,6x_{1.4} + 21,8x_{1.5} + 19,1x_{1.6} + 34,6x_{2.1} + 35,1x_{2.2} + 25,0x_{2.3} + 38,5x_{3.1} - 23,6y_5 - 29,8y_6 - 22,1y_7.$$

Оценка влияния факторов на уровень инвестиционной привлекательности и значения интегрального показателя для 2004-2007гг. представлены в приложении Н.

По значениям интегрального показателя произведено ранжирование муниципальных образований Оренбургской области по заданному интегральному свойству (таблица 2.29).

Таблица 2.29 – Ранжирование муниципальных образований Оренбургской области по уровню инвестиционной привлекательности за период с 2004 по 2008 гг.

Города и район	Рейтинг								
	2004	2005		2006		2007		2008	
		факт	$\gamma_i(2005)$	факт	$\gamma_i(2006)$	факт	$\gamma_i(2007)$	факт	$\gamma_i(2008)$
1	2	3	4	5	6	7	8	9	10
Абдулинский	47	47	0	47	0	47	0	47	0
Адамовский	16	10	6	7	3	11	-4	19	-8
Акбулакский	25	19	6	18	1	29	-11	14	+15
Александровский	23	29	-6	22	7	30	-8	29	+1
Асекеевский	41	36	5	44	-8	33	11	37	-4
Беляевский	20	16	4	16	0	39	-23	38	+1
Бугурусланский	34	33	1	38	-5	37	1	36	+1
Бузулукский	27	34	-7	33	1	38	-5	42	-4
Гайский	42	39	3	9	30	13	-4	12	+1
Грачевский	29	23	6	30	-7	41	-11	41	0
Домбаровский	21	18	3	27	-9	25	2	43	-18
Илекский	46	45	1	43	2	46	-3	44	+2
Кваркенский	30	26	4	34	-8	22	12	28	-6
Красногвардейский	33	37	-4	45	-8	32	13	31	+1
Кувандыкский	32	35	-3	32	3	35	-3	34	+1
Курманаевский	39	40	-1	37	3	44	-7	46	-2
Матвеевский	38	43	-5	41	2	36	5	22	+14
Новоорский	19	28	-9	23	5	23	0	24	-1
Новосергиевский	24	21	3	12	9	9	3	13	-4
Октябрьский	28	24	4	13	11	18	-5	23	-5
Оренбургский	2	2	0	3	-1	3	0	3	0
Первомайский	5	4	1	11	-7	27	-16	17	+10
Перволюцкий	35	27	8	31	-4	28	3	26	+2
Пономаревский	45	46	-1	42	4	45	-3	45	0
Сакмарский	12	25	-13	15	10	12	3	11	+1
Саракташский	22	30	-8	28	2	31	-3	27	+4
Светлинский	17	17	0	20	-3	24	-4	30	-6
Северный	43	42	1	46	-4	42	4	35	7
Соль-Илецкий	26	22	4	36	-14	40	-4	39	+1
Сорочинский	14	11	3	21	-10	19	2	20	-1
Ташлинский	11	9	2	10	-1	17	-7	15	+2
Тоцкий	37	41	-4	40	1	43	-3	33	+10
Тюльганский	44	44	0	39	5	34	5	40	-6
Шарлыкский	36	38	-2	35	3	26	9	25	+1

Продолжение таблицы 2.29

1	2	3	4	5	6	7	8	9	10
Ясненский	40	32	8	26	6	21	5	32	-11
г.Абдулино	31	31	0	25	6	20	5	10	+10
г.Бугуруслан	3	5	-2	14	-9	6	8	6	0
г.Бузулук	4	7	-3	2	5	2	0	2	0
г.Гай	13	13	0	19	-6	10	9	9	+1
г.Кувандык	18	20	-2	29	-9	15	14	18	-3
г.Медногорск	15	14	1	5	9	16	-11	16	0
г.Новотроицк	8	6	2	6	0	4	2	4	0
г.Оренбург	1	1	0	1	0	1	0	1	0
г.Орск	10	12	-2	17	-5	8	9	7	+1
г.Соль-Илецк	9	15	-6	24	-9	14	10	8	+6
г.Сорочинск	6	3	3	4	-1	5	-1	5	0
г.Ясный	7	8	-1	8	0	7	1	21	-14

Ряд сельских районов, (Адамовский, Александровский, Акбулакский, Октябрьский, Первомайский, Сакмарский, Ташлинский) характеризуются показателями, обеспечивающими средний уровень инвестиционной привлекательности. Такие районы, как Абдулинский, Илекский, Курманаевский, Тюльганский, Северный, Пономаревский являются непривлекательными для инвестирования.

В целом, ранжирование муниципальных образований по потенциалу, активности и риску, а также по сводной латентной категории – инвестиционной привлекательности, позволило провести сравнительный анализ муниципальных образований и выделить группы районов, менее привлекательные, которые нуждаются в поддержке администрации, а также районы, с более высоким уровнем инвестиционной привлекательности.

Для оценки согласованности двух рейтингов муниципальных образований: рейтинга, построенного на основании функции от инвестиционного потенциала, активности и риска, и рейтинга, построенного на основании «экспертной» информации, рассчитаны ранговые коэффициенты корреляции Спирмена и Кендалла (таблица 2.30).

Таблица 2.30 – Результаты расчета коэффициентов ранговой корреляции Спирмена и Кендалла

Коэффициент	Результаты расчетов				
	2004	2005	2006	2007	2008
$r(rU(t), rI(t))$ - коэффициент ранговой корреляции Спирмена	$r=0,900$ $t$ -статистика=13,868 $p = 0,000$	$r=0,864$ $t = 11,506$ $p = 0,000$	$r=0,802$ $t = 9,016$ $p = 0,000$	$r=0,881$ $t = 12,514$ $p = 0,000$	$r=0,911$ $t = 14,804$ $p = 0,000$
$\tau(rU(t), rI(t))$ - коэффициент ранговой корреляции Кендалла	$\tau = 0,758$ $z$ -статистика =7,511 $p = 0,000$	$\tau = 0,715$ $z = 7,089$ $p = 0,000$	$\tau = 0,613$ $z = 6,080$ $p = 0,000$	$\tau = 0,722$ $z = 7,162$ $p = 0,000$	$\tau = 0,763$ $z = 7,566$ $p = 0,000$

Высокие значения коэффициентов Спирмена и Кендалла свидетельствуют о существенной связи между показателями инвестиционной привлекательности, полученными на основании функции от инвестиционного потенциала, активности и риска и на основании функции, определяемой поддающимися учету и измерению признаками.

Таким образом, существенных различий в полученных результатах нет, но второй подход требует реализации трудоемкого и недостаточно теоретически обоснованного алгоритма, поэтому предпочтение отдано первому подходу.

## Вопросы и задания по теоретическому материалу раздела 2

1. В чем заключаются особенности известных Вам методов классификации/снижения размерности признакового пространства? Приведите примеры задач, в которых преимущества того или иного метода проявляются в наибольшей степени
2. Опишите известные Вам подходы к формированию названий латентных показателей, построенных методом главных компонент/факторным анализом/методом многомерного шкалирования
3. В каком случае возможно построение интегрального показателя экспертно-статистическим методом? Опишите алгоритм построения интегрального показателя экспертно-статистическим методом

4. В каком случае возможно построение интегрального показателя на основе компонентного анализа? Опишите алгоритм построения интегрального показателя на основе первой главной компоненты

5. Каким образом можно осуществить ранжирование объектов по латентной категории в ситуации неработоспособности первой главной компоненты и полного отсутствия какого-либо экспертного «обучения»?

6. В чем суть взвешенной процедуры индивидуального рейтингования в пространстве модифицированных первых главных компонент?

7. Каким образом можно осуществить анализ динамики рейтинговых позиций объектов и сравнение рейтинговых позиций административно-территориальных образований по уровням нескольких латентных категорий?

8. По показателям, характеризующим инвестиционный риск, определен вклад главных компонент в суммарную дисперсию исходных признаков и построена матрица факторных нагрузок:

**Вклад главных компонент в суммарную дисперсию исходных признаков**

Номер главной компоненты (i)	Собственные числа, $\lambda_i$	Относительный вклад каждой главной компоненты в суммарную дисперсию, %	Накопленные значения собственных чисел	Накопленный относительный вклад, %
1	1,50	50,15	1,50	50,147
2	1,00	33,34	2,50	83,490
3	0,495	16,51	3,00	100

**Матрица факторных нагрузок, построенная по показателям, характеризующим инвестиционный риск**

	F1	F2	F3
$y_1$	0,358	0,911	0,206
$y_2$	0,867	0,001	-0,498
$y_3$	0,790	-0,413	0,453

где  $y_1$  – задолженность организаций по заработной плате, в % от общего фонда заработной платы;

$y_2$  – удельный вес убыточных предприятий и организаций, в процентах от общего числа предприятий;

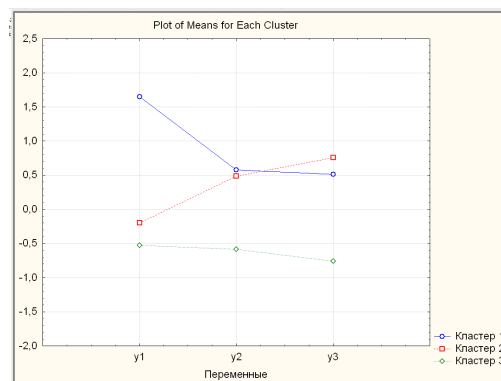
$y_3$  – просроченная кредиторская задолженность предприятий, в процентах от общей задолженности.

Прокомментируйте полученные результаты. Дайте интерпретацию главным компонентам. Какой метод следует применить для ранжирования объектов по инвестиционному риску?

9. В таблице представлены результаты классификации муниципальных образований Оренбургской области по показателям, характеризующим инвестиционный риск:

Номер кластера	Количество объектов в кластере	Состав класса
кластер 1	9	<b>Районы:</b> Асекеевский, Илекский, Курманаевский, Матвеевский, Первомайский, Пономаревский, Соль-Илецкий, Сорочинский, Тюльганский
кластер 2	16	<b>Города:</b> Абдулино, Гай, Орск, Соль-Илецк <b>Районы:</b> Абдулинский, Акбулакский, Бугурусланский, Бузулукский, Гайский, Грачевский, Кваркенский, Кувандыкский, Саракташский, Северный, Тоцкий, Шарлыкский
кластер 3	22	<b>Города:</b> Бугуруслан, Бузулук Кувандык, Медногорск, Новотроицк, Оренбург, Сорочинск, Ясный <b>Районы:</b> Адамовский, Александровский, Беляевский, Домбаровский, Красногвардейский, Новоорский, Новосергиевский, Октябрьский, Оренбургский, Переволоцкий, Сакмарский, Светлинский, Ташлинский, Ясенский

График средних значений признаков, имеет вид:



где  $y_1$  – задолженность организаций по заработной плате, в % от общего фонда заработной платы;

$y_2$  – удельный вес убыточных предприятий и организаций, в процентах от общего числа предприятий;

$y_3$  – просроченная кредиторская задолженность предприятий, в процентах от общей задолженности.

Дайте содержательную интерпретацию результатов классификации.

## **Задание к исследовательской работе**

Осуществите сравнительный анализ объектов (субъектов РФ, муниципальных образований области, края, республики) по одному из следующих интегральных показателей:

- уровню демографической безопасности;
- уровню экономической безопасности;
- уровню инвестиционной привлекательности;
- уровню антропогенной нагрузки;
- уровню эколого-экономического риска.

Выявите динамику рейтинговых позиций объектов по анализируемому латентному свойству.

Замечания:

- 1) объект и предмет исследования выберите самостоятельно;
- 2) для составления информационной базы рекомендуется воспользоваться статистическими данными Федеральной службы государственной статистики РФ и ее территориальных подразделений ([www.gks.ru](http://www.gks.ru)).

## **Список использованных источников к разделу 2**

- 1 Оссовский, С. Нейронные сети для обработки информации / С. Оссовский; пер. с польского И.Д. Руденко. – М.: Финансы и статистика, 2004. – 344 с.
- 2 Левин, В.С. Прогнозирование и классификация экономических систем в условиях неопределенности методами искусственных нейронных сетей / В.С. Левин, В.И. Смирнов. – Оренбург: Издательский центр ОГАУ, 2004. – 188 с.
- 3 Прикладная статистика. Основы эконометрики: учебник для вузов: В 2 т. 2-е изд., испр. – Т.1: Айвазян С.А., Мхитарян В.С. Теория вероятностей и прикладная статистика. – М.: ЮНИТИ-ДАНА, 2001. – 656 с.



- 4 Андерсон, Т. Введение в многомерный статистический анализ / Т. Андерсон; под ред. Б.В. Гнеденко. – М.: Физматгиз, 1963. – 500 с.
- 5 Айвазян, С. А. Прикладная статистика. Основа эконометрики : в 2 т.: учеб. для вузов / С. А. Айвазян, В. С. Мхитарян . - М. : ЮНИТИ-ДАНА, 2001. – 432 с.
- 6 Айвазян, С. А. Эмпирический анализ синтетических категорий качества жизни населения / С.А. Айвазян // Экономика и математические методы. – 2003. – Т. 39, № 3. – С. 19 –53.
- 7 Айвазян, С. А. К методологии измерения синтетических категорий качества жизни населения / С.А. Айвазян // Экономика и математические методы. – 2003. – Т.39, №2. – С.33 – 53.
- 8 Айвазян, С. А. Эконометрическое моделирование: учебное пособие – Выпуск 4: Межстрановой анализ интегральных характеристик категории качества жизни населения / С. А. Айвазян.– М.: МЭСИ, 2002. – 63 с.
- 9 Реннер, А.Г. Построение интегрального показателя эффективности функционирования (качества) экономического объекта /А.Г. Реннер, О.С. Бравичева // Проблемы теории и практики статистики: Сборник научных трудов ОГАУ. – Оренбург: Издательский центр ОГАУ, 2002. – 420 с.
- 10 Реннер, А. Г. Моделирование стоимости жилья на вторичном рынке жилья / А. Г. Реннер, О. И. Стебунова // Вестник Оренбургского государственного университета. - 2005. - Т. 1, N 10. - С. 179-182.
- 11 Айвазян, С.А. Классификация многомерных наблюдений / С.А. Айвазян, З.И. Бежаева, О.В. Староверов. – М.: Статистика, 1974. – 240 с.
- 12 Дубров, А.М. Многомерные статистические методы / А.М. Дубров, В.С. Мхитарян, Л.И. Трошин. – Финансы и статистика, 1998. – 352 с.
- 13 Сошникова, Л.А. Многомерный статистический анализ в экономике: учеб. пособие для вузов / Под ред. проф. В.Н. Тамашевича. – М.: ЮНИТИ-ДАНА, 1999. – 598 с.
- 14 Мхитарян, В.С. Использование языка Statistica Basic в ППП «Statistica» в задачах классификации: методические указания для самостоятельных компью-

терных исследований студентов / В.С. Мхитарян, Т.А. Иванова; Московский государственный университет экономики, статистики и информатики. М., 2002. – 18 с.

15 Бантикова, О.И. Демографическое состояние как реальная и потенциальная угроза устойчивому развитию общества / О.И. Бантикова // Взаимодействие реального и финансового секторов в трансформационной экономике: материалы Вроссисокй научно-практической конференции. – Оренбург: РИК ГОУ ВПО ОГУ, 2006. – С. 92-97.

16 Бантикова, О.И. Методы многомерной классификации в исследовании территориальной дифференциации Оренбургской области по состоянию показателей, характеризующих демографическую безопасность / О.И. Бантикова // Развитие Университетского комплекса как фактор повышения инновационного и образовательного потенциала региона: материалы Всероссийской научно-практической конференции. – Оренбург, 2007. – С. 14-20.

17 Бантикова, О.И. Анализ факторов, влияющих на демографическую безопасность Оренбургской области / О.И. Бантикова // Проблемы демографии, медицины и здоровья населения России: история и современность: сборник материалов II Международной научно-практической конференции. – Пенза, 2006. – С. 41-45.

18 Математические методы моделирования социально-экономических процессов (региональный аспект) / А.Г. Реннер [и др.]. – Самара: Изд-во СамНЦ РАН, 2008. – 182 с.

19 Анализ и моделирование демографических и миграционных процессов в контексте национальной безопасности (региональный аспект) / В.П. Ковалевский, О.В. Буреш, А.Г. Реннер, О.И. Бантикова, В.И. Васянина. Под редакцией А.Г. Реннера. – Самара: Изд-во СамНЦ РАН, 2009. – 226 с.

20 Валиуллин, Х.Х. Неоднородность инвестиционного пространства России: региональный аспект. Проблемы прогнозирования / Х.Х. Валиуллин, Э.Р. Шакирова, 2004 - №1 С.157–165.

21 Жемчужникова, Ю.А. Классификация муниципалитетов Оренбургской области по уровню инвестиционной привлекательности / Ю.А. Жемчужникова //

Воспроизводственный потенциал региона: сборник материалов III Международной научно-практической конференции. – Уфа: РИЦ БашГУ, 2007. – С. 357-360.

22 Седова, Е.Н. Многомерная классификация городов и районов Оренбургской области по показателям, характеризующим риски для населения от воздействия окружающей среды / Е.Н. Седова // Сборник статей VII Международной научно-практической конференции «Состояние биосферы и здоровье людей». – Пенза: РИО ПГСХА, 2007.- С. 180-184.

23 Города и районы Оренбургской области: Стат.сб./ Территориальный орган Федеральной службы государственной статистики по Оренбургской области. – Оренбург, 1998–2007 гг.

24 Бакуменко, Л. П. Интегральная оценка качества и степени экологической устойчивости окружающей среды региона (на примере республики Марий Эл) / Л.П. Бакуменко, П. А. Коротков // Прикладная эконометрика, 2008. – №1. – С. 73-92.

25 Руководство по оценке риска для здоровья населения при воздействии химических веществ, загрязняющих окружающую среду: руководство. Р 2.1.10.1920-04 (утв. главным государственным санитарным врачом РФ 05.03.2004) [Электронный ресурс].- Режим доступа адрес [www.infopravo.by.ru](http://www.infopravo.by.ru)

26 Руководство по оценке риска для здоровья населения при воздействии химических веществ, загрязняющих окружающую среду (Human Health Risk Assessment from Environmental Chemicals) – М.: Федеральный центр Госсанэпиднадзора Минздрава России, 2004. – 143 с.

27 Дэйвисон, М. Многомерное шкалирование: методы наглядного представления данных – М.: Финансы и статистика, 1988. – 254 с

28 Терехина, А. Ю. Многомерное шкалирование в психологии / А.Ю. Терехина // Психологический журнал.-Том 4.- №1. - 1983. - С.76-88.

29 Толстова, Ю.Н. Основы многомерного шкалирования: учебное пособие. – М.: КДУ, 2006. – 160 с.

30 Бантикова, О.И. Построение интегрального показателя, характеризующего уровень демографической безопасности / О.И. Бантикова // Вестник ОГУ. – 2007. – № 2. – С. 98 – 104.

31 Балаш, В.А. Модели линейной регрессии для панельных данных: учебное пособие для ВУЗов / В.А. Балаш, О.С. Балаш. – М., 2002. – 65 с.

32 Васянина, В.И. Моделирование интегрального показателя экономической безопасности региона / В.И. Васянина // Развитие университетского комплекса как фактор повышения инновационного и образовательного потенциала региона: материалы Всероссийской научно-практической конференции. – Оренбург, 2007. – С. 21-25.

33 Многомерные статистические методы и основы эконометрики: учеб.-практ. пособие / Сост. А.М. Дубров [и др.]. - М.: МЭСИ, 1998. – 108 с.

34 Васянина, В.И. Сравнительный анализ административно-территориальных образований региона по уровню экономической безопасности / В.И. Васянина // Вестник ОГУ. – 2008. – № 10. С. 76-80.

### **3 Моделирование социально-экономических явлений и процессов со специфическими зависимыми переменными**

При моделировании многих экономических и социальных процессов и явлений приходится сталкиваться с разного рода ограничениями на значения зависимой переменной. Так, зависимая переменная по смыслу задачи может быть только целочисленной и неотрицательной: количество детей в семье, количество комнат в доме, количество случаев обращения к врачу и т.д. Более того, зависимая переменная может принимать вообще нечисловые значения: наличие или отсутствие высшего образования у главы семьи, наличие или отсутствие работы у индивидуума, уровень образования у главы семьи и т.д. Для удобства обработки эти нечисловые значения подлежат оцифровке. Таким образом, далее речь пойдет о так называемых моделях с дискретными зависимыми переменными.

Модели с дискретными зависимыми переменными могут быть классифицированы, например, по типу зависимой переменной, на модели выбора среди конечного числа альтернативных вариантов и модели счетных данных [1]. Модели первого типа в зависимости от числа вариантов, среди которых осуществляется выбор, подразделяются на модели бинарного выбора и модели множественного выбора. К моделям множественного выбора относятся модели с неупорядоченными и упорядоченными альтернативами. В моделях упорядоченного выбора варианты можно расположить в определенном порядке (ранжировать). Например, это могут быть оценки, полученные на экзамене, выбор ценных бумаг, исходя из их рейтинга и т.д. Модели с неупорядоченными альтернативными вариантами применяются, когда значения результирующей переменной не могут быть упорядочены. Например, выбор вида страхования, выбор типа отдыха и др.

Раздел организован следующим образом. В подразделе 3.1 рассматриваются модели бинарного выбора: дается постановка задачи, в том числе, с использованием понятия «полезности», записывается модель, кратко описывается получение оценок коэффициентов модели методом максимального правдоподобия, затрагиваются вопросы оценки качества модели и интерпретации коэффициентов. Не-

большой подраздел 3.2 посвящен обобщению моделей бинарного выбора – моделям множественного упорядоченного выбора. В подразделе 3.3 дается описание двумерных пробит-моделей, представляющих собой модели бинарного выбора, объединенные в одну систему. В подразделе 3.4 на основе модели бинарного выбора осуществляется ранжирование категорий безработных по степени их востребованности на рынке труда. В подразделах 3.5 и 3.7 модели бинарного и множественного выбора используются для построения интегральных показателей экономической безопасности, антропогенной нагрузки и эколого-экономического риска. В подразделе 3.6 на основе моделей бинарного и множественного выбора осуществляется выявление факторов, связанных с уровнем эколого-экономического риска. И, наконец, в подразделе 3.8 на основе двумерных пробит-моделей исследуются взаимосвязи эколого-экономического риска с уровнем социально-экономического развития и уровнем заболеваемости населения региона.

### 3.1 Модели бинарного выбора

Зависимую переменную, принимающую всего два значения, называют бинарной. Для удобства значения бинарной переменной, которые могут, вообще говоря, иметь и нечисловую природу, чаще всего кодируют нулем и единицей. Например, зависимая переменная «занятость» может принимать два значения: 0, если индивидуум безработный, и 1, если имеет работу. Зависимая переменная «риск банкротства» может принимать значение 1, если банк имеет высокий риск банкротства, и 0 в противном случае.

Нас будет интересовать моделирование зависимости между бинарной результирующей переменной  $y$  и объясняющими переменными  $x = (x_1, x_2, \dots, x_p)^T$ , информация о которых, как обычно, представлена матрицей  $X$  типа «объект-свойство» размерности  $n \times (p + 1)$ . Значения факторных признаков для  $i$ -го объекта обозначены через  $X_i$ . Естественно описывать интересующую нас связь, используя функцию регрессии.

Для дискретной случайной величины, принимающей только два возможных значения 0 и 1, условное по  $x = (x_1, x_2, \dots, x_p)$  функция регрессии имеет вид:

$$M(y|x) = 1 \cdot P(y=1|x) + 0 \cdot P(y=0|x) = P(y=1|x). \quad (3.1)$$

Значит, задача моделирования связи бинарной результативной переменной и факторов  $x_1, x_2, \dots, x_p$  сводится к задаче моделирования условной *вероятности* того, что результативная переменная принимает значение 1.

Если по аналогии с обычным регрессионным анализом использовать линейную аппроксимацию  $P(y=1|x) = \beta_0 + \beta_1 \cdot x_1 + \dots + \beta_p \cdot x_p = x^T \cdot \beta$ , где  $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$ , то линейная функция в правой части будет принимать значения, выходящие за диапазон от 0 до 1, в то время как стоящая слева вероятность ограничена указанным диапазоном. Поэтому подберем некоторую функцию, область значений которой есть  $[0;1]$ :

$$P(y=1|x) = F(\beta_0 + \beta_1 \cdot x_1 + \dots + \beta_p \cdot x_p). \quad (3.2)$$

Относительно функции  $F(z)$  будем предполагать, что она удовлетворяет, к примеру, следующим требованиям:

$$\left. \begin{array}{ll} (a) & F(z) \text{ монотонно возрастает по } z; \\ (б) & 0 \leq F(z) \leq 1 \\ (в) & F(z) \rightarrow 1 \text{ при } z \rightarrow \infty \\ (г) & F(z) \rightarrow 0 \text{ при } z \rightarrow 0 \end{array} \right\} \quad (3.3)$$

Перечисленным свойствам удовлетворяют, например, такие функции, как:

1) функция распределения стандартной нормальной случайной величины

$$F(z) = \Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt = \int_{-\infty}^z \varphi(t) dt;$$

2) функция логистического распределения  $F(z) = \Lambda(z) = \frac{e^z}{1 + e^z}$ .

Для оценки параметров функции регрессии переходят к модели регрессии:

$$\tilde{p}_i = F(\tilde{x}_i \cdot \beta) + u_i, \quad i = 1, \dots, m, \quad (3.4)$$

где  $\tilde{p}_i$  – оценка вероятности события, состоящего в том, что для  $i$ -ой группы объектов резульативная переменная примет значение 1 (в качестве такой оценки естественно брать относительную частоту соответствующего события),

$\tilde{x}_i$  – вектор значений факторных признаков для  $i$ -ой группы объектов,

$u_i$  – случайная ошибка для  $i$ -ой группы объектов,

$m$  – количество групп объектов, характеризующихся одинаковыми значениями факторных признаков.

Подходы к оцениванию коэффициентов модели (3.4) по сгруппированным данным подробно описаны в [2]. Очевидно, что такие подходы требовательны к объему и структуре исходных данных: в выборке должно быть достаточное количество групп объектов, характеризующих разными наборами факторных признаков (чтобы обеспечить достаточно большую величину  $m$ ), и каждая из этих групп должна быть достаточно большой для получения приемлемой оценки вероятности  $\tilde{p}_i$ . Поэтому чаще оценка коэффициентов моделей с бинарными резульативными переменными осуществляется методом максимального правдоподобия (ММП), который описан ниже. Предварительно обоснуем модель бинарного выбора с помощью введения скрытой (латентной) переменной. Начнем с примера.

Рассмотрим модель принятия индивидуумом решения: согласиться на переезд в другой город в связи с повышением на работе или нет. Будем считать, что как потребительское, так и трудовое поведение индивидуума описывается некоторой функцией полезности, зависящей от многих параметров – дохода, свободного времени, наличия детей и родителей, их возраста, наличия собственного жилья и



т.д. Перед индивидуумом стоит выбор: согласиться на переезд в другой город и тем самым увеличить доход семьи, но столкнуться с необходимостью аренды жилья, смены школ детей, потерять возможность уделять больше времени родителям и старым друзьям, или не соглашаться. Каждой из этих ситуаций соответствует своя величина полезности [3].

Обозначим через  $U_i^1$  величину полезности, если индивидуум соглашается на переезд, и через  $U_i^0$  – если не соглашается. В случае, когда  $U_i^1 > U_i^0$ , индивидууму выгоднее согласиться на переезд, поскольку дополнительный доход от повышения в должности перевешивает денежные, временные и психологические затраты, связанные с переездом. Если же  $U_i^1 < U_i^0$ , то на переезд индивидуум не соглашается.

Предположим, что разность полезностей  $y_i^* = U_i^1 - U_i^0$  является линейной функцией от наблюдаемых характеристик: ставки заработной платы, возраста, наличия детей и других персональных характеристик. Тогда ее можно представить как функцию наблюдаемых и ненаблюдаемых характеристик:

$$y_i^* = X_i \cdot \beta + \varepsilon_i, \quad i = 1, \dots, n, \quad (3.5)$$

где  $n$  – количество объектов наблюдения (индивидуумов).

Переменная  $y_i^*$  является латентной и непрерывной, а случайная ошибка  $\varepsilon_i$  включает в себя влияние всех неучтенных в модели факторов.

Найдем вероятность того, что индивидуум  $i$  согласится на переезд, то есть результирующая переменная примет значение 1:

$$\begin{aligned} P(y_i = 1 | X_i) &= P(y_i^* \geq 0 | X_i) = P(X_i \cdot \beta + \varepsilon_i \geq 0) = \\ &= P(-\varepsilon_i \leq X_i \cdot \beta) = F_{-\varepsilon}(X_i \cdot \beta), \quad i = 1, \dots, n \end{aligned} \quad (3.6)$$

Получили, что искомая вероятность выражается через функцию распределения случайной величины  $(-\varepsilon_i)$ , которая в случае симметричного распределения

совпадает с функцией распределения  $F_{-\varepsilon}(z)$  случайной величины  $\varepsilon_i$  и далее для удобства будет обозначаться как  $F(z)$  [3].

Поскольку полезность не имеет стандартной единицы измерения, то для нее можно брать любой масштабирующий множитель, дисперсию ошибки не оценивать, а брать заранее заданной конкретной константе (это может быть единица или любая другая величина). Это называется нормализацией.

Во многих экономических задачах возможно обоснование модели дискретного выбора через разность значений функций полезности, но латентную переменную  $y_i^*$  часто вводят и напрямую:

$$y_i^* = X_i \cdot \beta + \varepsilon_{ip}, \quad i = 1, \dots, n, \quad (3.7)$$
$$y_i = \begin{cases} 1 - & \text{если } y_i^* > 0 \\ 0 - & \text{если } y_i^* \leq 0 \end{cases}$$

где  $n$  – количество объектов наблюдения.

При этом предполагается, что случайные ошибки  $\varepsilon_i$  независимы между собой и независимы с объясняющими переменными  $x = (x_1, x_2, \dots, x_p)^T$ .

Используя конкретные законы распределения случайной ошибки, можно получить различные варианты модели бинарного выбора, к примеру: пробит- модель в случае стандартного нормального распределения вероятностей и логит-модель в случае логистическом закона распределения вероятностей. Графики функции распределения нормального и логистического распределения при соответствующей нормировке достаточно близки. На интервале  $z \in [-1,2; 1,2]$  они практически одинаковы. Однако логистическая функция имеет более «тяжелые хвосты», т.е. медленнее стремится к нулю при  $z \rightarrow -\infty$  или единице при  $z \rightarrow \infty$ . Поэтому логит- и пробит модели дают похожий результат, если только изучаемая вероятность не слишком близка к нулю или единице.

Для оценки параметров  $\beta$  логит- и пробит моделей обычно используют метод максимального правдоподобия. В силу предположений относительно свойств случайных ошибок наблюдения  $y_1, \dots, y_n$  независимы. Поскольку  $y_i$  может принимать значения только 0 или 1, то задача поиска вектора оценок параметров  $\beta$  сводится к задаче вида:

$$L = \max_{\beta} \prod_{i=1}^n (F(X_i^T \beta))^{y_i} (1 - F(X_i^T \beta))^{1-y_i}. \quad (3.8)$$

Легко видеть, что точка максимума (оценка неизвестного вектора коэффициентов) для пробит-модели может быть найдена как решение системы нелинейных уравнений (3.9)

$$\sum_{i=1}^n \left[ \frac{y_i - \Phi(X_i \cdot \beta)}{\Phi(X_i \cdot \beta)(1 - \Phi(X_i \cdot \beta))} \varphi(X_i \cdot \beta) \right] X_i = 0 \quad (3.9)$$

и (3.10) для логит-модели

$$\sum_{i=1}^n (y_i - \Lambda(X_i \cdot \beta)) \cdot X_i = 0. \quad (3.10)$$

Оценка функции регрессии:

- для пробит-модели:  $\hat{P}(y_i = 1 | X_i) = \Phi(X_i \cdot \hat{\beta})$ ;
- для логит-модели:  $\hat{P}(y_i = 1 | X_i) = \Lambda(X_i \cdot \hat{\beta}) = \frac{\exp(X_i \cdot \hat{\beta})}{1 + \exp(X_i \cdot \hat{\beta})}$ .

Для проверки значимости уравнения необходимо убедиться, что коэффициенты  $\beta_1, \beta_2, \dots, \beta_p$  значимо отличаются от нуля.

Для проверки гипотезы  $H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$  использует критерий отношения правдоподобия:

$$LR = 2(\ln L - \ln L_0), \quad (3.11)$$

где  $\ln L$  – значение логарифма функции правдоподобия,

$\ln L_0$  – значение логарифма функции правдоподобия при нулевой гипотезе, т.е. для тривиальной модели.

При выполнении нулевой гипотезы величина  $LR$  имеет хи-квадрат распределение с  $p$  степенями свободы [3].

Для моделей бинарного выбора трудно предположить естественную меру качества аппроксимации, такую как коэффициент детерминации  $R^2$  для линейной модели. Чаще всего такие меры строятся путем прямого или косвенного сравнения текущей модели и тривиальной модели.

Очевидно,  $\ln L > \ln L_0$ . Чем больше разность между этими величинами, тем лучше полная модель по сравнению с тривиальной.

По аналогии с коэффициентами детерминации построен коэффициент

$$R_{pseudo}^2 = 1 - \frac{1}{1 + 2(\ln L - \ln L_0)/n}. \quad (3.12)$$

Альтернативная мера, называемая *индексом отношения правдоподобия*, предложена Макфадденом (McFadden):

$$R_{McFadden}^2 = LRI = 1 - \frac{\ln L}{\ln L_0}. \quad (3.13)$$

Если коэффициенты логит- или пробит- модели незначимы, т.е. все коэффициенты равны нулю, то  $\ln L = \ln L_0$  и псевдо-  $R^2$  и  $R^2$  Макфаддена равны нулю. Если модель совершенно точна, т.е. прогнозные вероятности совпадают с наблюдаемыми значениями, тогда все множители функции правдоподобия будут равны 1, а логарифм правдоподобия равен 0. Поэтому верхняя граница, равная 1, может достигаться для индекса отношения правдоподобия Макфаддена.

Альтернативный способ построения мер качества состоит в вычислении прогноза и сравнения его с фактическими значениями. Будем считать, что если предсказанная по модели вероятность больше 0,5, то прогнозное значение равно 1, если меньше 0,5, то 0. Так как плотность распределения нормального и логистического законов симметрична относительно нуля, это соответствует правилу:

$$\hat{y}_i = 1, \text{ если } X_i \cdot \hat{\beta} > 0$$

и

$$\hat{y}_i = 0, \text{ если } X_i \cdot \hat{\beta} < 0.$$

Доля неправильных прогнозов задается формулой

$$wr_1 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (3.14)$$

Логично сравнивать ошибку прогноза текущей и тривиальной модели. Для тривиальной модели все предсказанные вероятности будут равны между собой и равны доле успехов в выборке  $p = n_1 / n$ . Прогноз для всех наблюдений будет одинаков (все нули или единицы, если  $p$  меньше (больше) 0,5). Число ошибок прогноза будет равно  $wr_0 = 1 - \hat{p}$ , если  $\hat{p} > 0,5$  и  $wr_0 = \hat{p}$ , если  $\hat{p} \leq 0,5$ . Тогда качество подгонки может быть оценено как

$$R_p^2 = 1 - \frac{wr_1}{wr_0}. \quad (3.15)$$

Этот коэффициент может принимать отрицательные значения, если число ошибок текущей модели окажется больше, чем для простейшего предсказания.

Интерпретация коэффициентов моделей бинарного выбора отличается от обычной интерпретации коэффициентов линейной модели. Если в линейной модели коэффициенты регрессии  $\beta_j$  представляют собой производные по независимым переменным, то есть уже представляют собой предельные эффекты, то в логит- и пробит-моделях коэффициенты не соответствуют предельному (маржинальному) эффекту  $k$ -ой независимой переменной. Этот эффект является функцией всех объясняющих переменных. Для пробит-модели предельный эффект выглядит следующим образом:

$$\frac{\partial \Phi(X_i \cdot \beta)}{\partial X_{ij}} = \varphi(X_i \cdot \beta) \cdot \beta_j. \quad (3.16)$$

Для логит-модели имеем:

$$\frac{\partial \Lambda(X_i \cdot \beta)}{\partial X_{ij}} = \frac{e^{X_i \cdot \beta}}{(1 + e^{X_i \cdot \beta})^2} \cdot \beta_j. \quad (3.17)$$

Знак предельного эффекта  $j$ -ой переменной соответствует знаку коэффициента  $\beta_j$  и легко интерпретируется.

Отметим, что в случае логит-модели может быть получена более наглядная интерпретация. Из линейности этой модели в отношении логита следует, что для интерпретации коэффициента  $\beta_j$  можно использовать понятие шанса: при не-

большом изменении объясняющей переменной  $x_j$  шансы на то, что  $y_i = 1$  против того, что  $y_i = 0$  возрастают (уменьшаются) приблизительно на  $100 \cdot \beta_j \cdot \Delta x_j$ .

Если объясняющая переменная  $d$  принимает только два значения 0 и 1, то малые изменения переменной невозможны. Поэтому предельный эффект бинарной объясняющей переменной  $d$  определяется как разность  $P\{y = 1|x^*, d = 1\} - P\{y = 1|x^*, d = 0\}$ .

Поскольку величина предельного эффекта зависит от значений объясняющих переменных, то при анализе модели речь часто идет о «среднем предельном эффекте», для оценки которого возможны два подхода. Первый основан на усреднении значений объясняющих переменных и оценке предельного эффекта для этого «наиболее типичного объекта». Второй подход заключается в усреднении индивидуальных предельных эффектов (рассчитанных для каждого наблюдения). Считается, что при большом объеме выборки оба подхода дают одинаковые результаты, а при малом объеме выборки предпочтительнее второй подход.

### 3.2 Модели множественного выбора

Естественным обобщением модели бинарного выбора являются модель упорядоченного множественного выбора. Действительно, для индивидуума между вариантами «быть безработным» и «иметь работу» есть вариант «иметь частичную занятость», особенно когда речь идет о сокращении рабочего дня/недели по инициативе работодателя. При исследовании рисков естественно моделировать не просто наличие или отсутствие риска, а его уровень – например, хотя бы как низкий, средний или высокий. Аналогичным образом можно подойти к моделированию любых рейтингов (рейтингов инвестиционной привлекательности, кредитных рейтингов и т.д.). Можно сказать, что значения зависимой переменной в рассмотренных примерах выражают отношения предпочтения среди альтернативных вариантов.

Используя понятие латентной переменной (ненаблюдаемой полезности), по аналогии с моделью бинарного выбора, запишем модель упорядоченного множественного выбора следующим образом:

$$y_i^* = \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, \dots, n. \quad (3.18)$$

$$y_i = \begin{cases} 1, & \text{если } y_i^* \leq \gamma_1, \\ \dots \\ k, & \text{если } \gamma_{k-1} < y_i^* \leq \gamma_k \\ \dots \\ K, & \text{если } y_i^* > \gamma_{K-1} \end{cases}$$

где  $y_i^*$  – значение латентной переменной для объекта  $i$ ,

$X_i = (x_{i1}, \dots, x_{ip})$  – вектор-строка значений объясняющих переменных для объекта  $i$ ,

$\varepsilon_i$  – случайная ошибка, отражающая влияние на значение  $y_i^*$  неучтенных дополнительных факторов,

$n$  – количество объектов наблюдения,

$\gamma_1 < \dots < \gamma_k < \dots < \gamma_{K-1}$  – пороговые значения, вообще говоря, также ненаблюдаемые [4].

Предполагая, что ошибки  $\varepsilon_1, \dots, \varepsilon_n$  – независимые в совокупности случайные величины, имеющие одинаковое нормальное распределение с математическим ожиданием нуль и дисперсией  $\sigma^2$ , мы получаем порядковую пробит-модель, а в случае логистического распределения – порядковую логит-модель. Функцию распределения ошибок обозначим через  $F(z)$ .

Тогда, полагая  $\gamma_0 = -\infty$  и  $\gamma_k = +\infty$ , можно записать общее выражение для вероятности отнесения объекта  $i$  к классу  $j \in \{1, \dots, K\}$ , или вероятности выбора индивидуумом  $i$  варианта  $j$ :



$$P\{y_i = j|X_i\} = P\{\gamma_{j-1} < y_i^* \leq \gamma_j|X_i\} = F(\gamma_j - X_i \cdot \beta) - F(\gamma_{j-1} - X_i \cdot \beta), \quad i = 1, \dots, n.$$

Неизвестные  $p + K - 1$  параметров модели (3.18) оцениваются методом максимального правдоподобия. Логарифмическая функция правдоподобия в предположении независимости наблюдений  $y_i$  имеет вид:

$$\ln L(\beta, \gamma) = \sum_{i=1}^n \sum_{j=1}^K s_{ij} \log[F(\gamma_j - X_i \cdot \beta) - F(\gamma_{j-1} - X_i \cdot \beta)], \quad (3.19)$$

где  $s_{i,j} = \begin{cases} 1, & \gamma_{j-1} < y_i^* < \gamma_j \\ 0, & y_i^* \notin (\gamma_{j-1}; \gamma_j) \end{cases}$ .

Пусть, например, ошибки нормально распределены, то есть рассматривается пробит-модель, функция распределения ошибок  $F(z) = \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}} dt$ . Тогда для случая  $K=4$  вероятности выбора индивидуумом каждого варианта рассчитываются следующим образом.

Вероятность выбора первой альтернативы:

$$\begin{aligned} P\{y_i = 1|X_i\} &= P\{y_i^* \leq \gamma_1|X_i\} = P\{\beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i \leq \gamma_1|X_i\} = \\ &= P\{\varepsilon_i \leq \gamma_1 - (\beta_1 x_{i1} + \dots + \beta_p x_{ip})|X_i\} = \Phi\left(\frac{\gamma_1 - (\beta_1 x_{i1} + \dots + \beta_p x_{ip})}{\sigma}\right). \end{aligned}$$

Вероятность выбора второй альтернативы:

$$\begin{aligned} P\{y_i = 2|X_i\} &= P\{\gamma_1 < y_i^* \leq \gamma_2|X_i\} = P\{\gamma_1 < \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i \leq \gamma_2|X_i\} = \\ &= P\{\gamma_1 - (\beta_1 x_{i1} + \dots + \beta_p x_{ip}) < \varepsilon_i \leq \gamma_2 - (\beta_1 x_{i1} + \dots + \beta_p x_{ip})|X_i\} = \\ &= \Phi\left(\frac{\gamma_2 - (\beta_1 x_{i1} + \dots + \beta_p x_{ip})}{\sigma}\right) - \Phi\left(\frac{\gamma_1 - (\beta_1 x_{i1} + \dots + \beta_p x_{ip})}{\sigma}\right) \end{aligned}$$

Вероятность выбора третьей альтернативы:

$$\begin{aligned} P\{y_i = 3|X_i\} &= P\{\gamma_2 < y_i^* \leq \gamma_3|X_i\} = P\{\gamma_2 < \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i \leq \gamma_3|X_i\} = \\ &= P\{\gamma_2 - (\beta_1 x_{i1} + \dots + \beta_p x_{ip}) < \varepsilon_i \leq \gamma_3 - (\beta_1 x_{i1} + \dots + \beta_p x_{ip})|X_i\} = \\ &= \Phi\left(\frac{\gamma_3 - (\beta_1 x_{i1} + \dots + \beta_p x_{ip})}{\sigma}\right) - \Phi\left(\frac{\gamma_2 - (\beta_1 x_{i1} + \dots + \beta_p x_{ip})}{\sigma}\right) \end{aligned}$$

Вероятность выбора четвертой альтернативы:

$$\begin{aligned} P\{y_i = 4|X_i\} &= P\{y_i^* > \gamma_3|X_i\} = P\{\beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i > \gamma_3|X_i\} = \\ &= P\{\varepsilon_i > \gamma_3 - (\beta_1 x_{i1} + \dots + \beta_p x_{ip})|X_i\} = 1 - \Phi\left(\frac{\gamma_3 - (\beta_1 x_{i1} + \dots + \beta_p x_{ip})}{\sigma}\right). \end{aligned}$$

Применяя метод максимального правдоподобия, мы, как и в случае пробит-модели бинарного выбора, не можем однозначно восстановить значения параметров  $\beta_1, \dots, \beta_p$ , если известны только значения  $y_i, x_{i1}, \dots, x_{ip}$  и неизвестны значения  $\sigma, \gamma_1, \dots, \gamma_K$ . Поэтому и здесь для однозначной идентификации коэффициентов  $\beta_1, \dots, \beta_p$  необходима нормализация. Так, часто предполагается, что  $\sigma = 1$  и  $\gamma_1 = 0$ , хотя возможны и другие нормализации [4]. С использованием такой стандартной нормализации мы получаем в модели с четырьмя исходами вероятности отнесения  $i$ -го объекта к одному из четырех классов (или выбора одной из четырех альтернатив):

$$\begin{aligned} P\{y_i = 1|X_i\} &= P\{y_i^* \leq 0|X_i\} = \Phi(-X_i \cdot \beta), \\ P\{y_i = 2|X_i\} &= P\{0 < y_i^* \leq \gamma_2|X_i\} = \Phi(\gamma_2 - X_i \cdot \beta) - \Phi(-X_i \cdot \beta), \\ P\{y_i = 3|X_i\} &= P\{\gamma_2 < y_i^* \leq \gamma_3|X_i\} = \Phi(\gamma_3 - X_i \cdot \beta) - \Phi(\gamma_2 - X_i \cdot \beta), \\ P\{y_i = 4|X_i\} &= P\{y_i^* > \gamma_3|X_i\} = 1 - \Phi(\gamma_3 - X_i \cdot \beta). \end{aligned}$$

Оценка качества моделей упорядоченного множественного выбора осуществляется аналогично оценке качества моделей бинарного выбора – например, на основе предложенного Макфадденом индекса отношения правдоподобия LRI. Про-

верка статистической значимости отдельных коэффициентов модели осуществляется на основе статистики Вальда.

Интерпретация коэффициентов не так прозрачна, как в случае обычной модели регрессии. Поскольку модель нелинейна, то коэффициенты модели не выражают предельных эффектов факторов. Например, для случая пробит-модели и трех альтернатив  $K=3$  дифференцирование по любому из факторов приводит к следующим соотношениям:

$$\frac{\partial P\{y_i = 1|X_i\}}{\partial x_{ij}} = -\varphi(\gamma_1 - X_i\beta) \cdot \beta_j, \quad (3.20a)$$

$$\frac{\partial P\{y_i = 2|X_i\}}{\partial x_{ij}} = [\varphi(\gamma_1 - X_i\beta) - \varphi(\gamma_2 - X_i\beta)] \cdot \beta_j, \quad (3.20б)$$

$$\frac{\partial P\{y_i = 3|X_i\}}{\partial x_{ij}} = \varphi(\gamma_2 - X_i\beta) \cdot \beta_j. \quad (3.20в)$$

Предельный эффект представляет собой величину, перераспределяемую между вероятностями полученного распределения, причем в сумме все изменения равны нулю. При  $\beta_j > 0$  вероятность события  $y_i = 1$  уменьшается на величину  $\varphi(\gamma_1 - X_i\beta) \cdot \beta_j$ , одновременно с этим вероятность события  $y_i = 2$  на эту же величину увеличивается, но также и уменьшается на величину  $\varphi(\gamma_2 - X_i\beta) \cdot \beta_j$ , а вероятность события  $y_i = 3$  на  $\varphi(\gamma_2 - X_i\beta) \cdot \beta_j$  увеличивается. Это означает, что при положительных  $\beta_j$  смещение вероятности происходит вправо, при отрицательных – влево. Поэтому увеличение значения  $j$ -й объясняющей переменной, когда коэффициент при ней положителен, приводит к увеличению вероятностей событий, получивших высокие ранги, и уменьшению вероятностей событий, получивших ранги низкие (если коэффициент отрицателен, то наоборот) [4, 5].

Прогнозирование по оцененной модели производится в соответствии со следующим соглашением:

$$\hat{P}\{y_i = k_0 | X_i\} = \max_{k=1, \dots, K} \hat{P}\{y_i = k | X_i\}, i = 1, 2, \dots, n, \dots \quad (3.21)$$

Таким образом, можно считать, что при выполнении этого условия объект  $i$  принадлежит классу  $k_0$ .

### **3.3 Системы одновременных моделей бинарного выбора (двумерные пробит-модели)**

При моделировании социально-экономических процессов на основе моделей бинарного и множественного выбора, как и в случае с обычными регрессионными моделями, исследователь может столкнуться с проблемой эндогенности. Например, при моделировании эколого-экономического риска в регионе в качестве влияющего на него фактора естественно проанализировать уровень промышленного и экономического развития, потому что в большинстве случаев развитие промышленности ведет к появлению новых источников опасности, увеличению загрязнения окружающей среды. Вместе с тем, уровень экономического развития также может зависеть от уровня эколого-экономического риска, поскольку высокие риски ведут к появлению дополнительных затрат и снижают инвестиционную привлекательность региона. Другой пример – моделирование взаимосвязи инновационной активности предприятий и их эффективного функционирования. Внедрение инноваций, очевидно, должно повышать эффективность предприятия. В свою очередь, более эффективно функционирующим предприятиям легче выделять ресурсы на инновационное развитие.

Объединение моделей бинарного выбора в систему позволяет моделировать вероятности выбора в рамках многомерных альтернативных вариантов. В этом случае каждая из пробит-моделей используется для описания эксперимента с двумя исходами, а система из  $k$  пробит-моделей позволит моделировать вероятности уже  $2^k$  исходов. Например, система из двух пробит-моделей может быть исполь-

зована для моделирования вероятности переезда (непереезда) на новое место жительства и аренды (покупки) жилья [1]. Или для моделирования вероятностей принятия решений о голосовании за или против введения налога на имущество и о посещении детьми частной или государственной школы. Основная идея заключается в том, что такие решения должны быть связаны друг с другом. Так, возможность приобретения жилья на новом месте может способствовать принятию решения о переезде или, наоборот, переезд обуславливает необходимость аренды жилья.

Для случая двух бинарных зависимых переменных модель может быть представлена как система вида:

$$\begin{aligned}
 y_{i1}^* &= X_{i1} \cdot \beta_1 + \varepsilon_{i1} \\
 y_{i1} &= \begin{cases} 1, & y_{i1}^* > 0 \\ 0, & y_{i1}^* \leq 0 \end{cases}
 \end{aligned}
 \tag{3.22}$$

$$\begin{aligned}
 y_{i2}^* &= X_{i2} \cdot \beta_2 + \varepsilon_{i2} \\
 y_{i2} &= \begin{cases} 1, & y_{i2}^* > 0 \\ 0, & y_{i2}^* \leq 0 \end{cases}
 \end{aligned}$$

$$i = 1, \dots, n$$

Это запись системы из двух моделей бинарного выбора в терминах латентных переменных  $y_{i1}^*$  и  $y_{i2}^*$ , которые интерпретируются как ненаблюдаемые полезности (выгоды) принятых решений. Векторы  $X_{1i}$  и  $X_{2i}$  – это векторы-строки значений факторов для  $i$ -го объекта, а векторы  $\beta_1$  и  $\beta_2$  – это векторы-столбцы коэффициентов первого и второго уравнений. Ошибки  $\varepsilon_{i1}$  и  $\varepsilon_{i2}$  чаще всего предполагаются распределенными нормально, и тогда получаем двумерную пробит-модель.

Две альтернативы в каждой из пробит-моделей системы дают следующие четыре возможные комбинации решений:

$$\begin{aligned}
y_1 &= y_2 = 1 \\
y_1 &= 1, y_2 = 0 \\
y_1 &= 0, y_2 = 1 \\
y_1 &= y_2 = 0
\end{aligned}$$

Если ошибки независимы, то уравнения системы (3.22) могут быть оценены по отдельности.

Если же есть основания предполагать, что  $\text{cov}(\varepsilon_1, \varepsilon_2) \neq 0$ , то раздельное оценивание недопустимо. В этом общем случае закон совместного распределения ошибок модели  $\varepsilon_1$  и  $\varepsilon_2$  характеризуется следующими параметрами:

$$\begin{aligned}
M(\varepsilon_1 | x^1, x^2) &= M(\varepsilon_2 | x^1, x^2) = 0, \\
D(\varepsilon_1 | x^1, x^2) &= D(\varepsilon_2 | x^1, x^2) = 1, \\
\text{cov}(\varepsilon_1, \varepsilon_2 | x^1, x^2) &= \rho(\varepsilon_1, \varepsilon_2 | x^1, x^2) = \rho,
\end{aligned}$$

где  $x^1, x^2$  – это векторы факторных (объясняющих) переменных первого и второго уравнений соответственно.

Введем дополнительные обозначения, которые позволят более компактно записать закон распределения и, соответственно, функцию правдоподобия. Система записи и вывод предложены в [6]. Пусть

$$\begin{aligned}
q_{i1} &= 2 \cdot y_{i1} - 1 \\
q_{i2} &= 2 \cdot y_{i2} - 1, \quad i = 1, \dots, n
\end{aligned}$$

Тогда, очевидно,  $q_{i1}$  и  $q_{i2}$  будут принимать только два значения 1 и -1:

$$q_{ij} = \begin{cases} 1, & y_{ij} = 1 \\ -1, & y_{ij} = 0 \end{cases}, \quad i = 1, \dots, n; \quad j = 1, 2$$

Индекс  $j$  используется здесь для обозначения первого и второго уравнений системы.

Введем также в рассмотрение следующие переменные:

$$\begin{aligned} z_{ij} &= x_{ij} \cdot \beta_j \\ w_{ij} &= q_{ij} \cdot z_{ij}, \quad i = 1, \dots, n; \quad j = 1, 2 \end{aligned}$$

И, наконец, обозначим:  $\rho_{i^*} = q_{i1} \cdot q_{i2} \cdot \rho$

Тогда вероятности, которые будут входить в функцию правдоподобия, и логарифм самой функции правдоподобия имеют вид

$$\begin{aligned} P(Y_1 = y_{i1}, Y_2 = y_{i2} | x_1, x_2) &= \Phi_2(w_{i1}, w_{i2}, \rho_{i^*}) \\ \ln L &= \sum_{i=1}^n \ln \Phi_2(w_{i1}, w_{i2}, \rho_{i^*}) \end{aligned}$$

Через  $\varphi_2(\cdot)$  и  $\Phi_2(\cdot)$  будем обозначать плотность распределения вероятностей и функцию распределения вероятностей двумерного нормально распределенного случайного вектора. Обозначения без индекса 2  $\varphi(\cdot)$  и  $\Phi(\cdot)$  как обычно используются для плотности распределения вероятностей и функции распределения вероятностей одномерной случайной величины.

Оценка коэффициентов модели в силу необходимого условия существования экстремума ищется из следующей системы:

$$\begin{aligned} \frac{\partial \ln L}{\partial \beta_1} &= \sum_{i=1}^n \left( \frac{q_{i1} \cdot g_{i1}}{\Phi_2} \right) x_{i1} = 0 \\ \frac{\partial \ln L}{\partial \beta_2} &= \sum_{i=1}^n \left( \frac{q_{i2} \cdot g_{i2}}{\Phi_2} \right) x_{i2} = 0 \\ \frac{\partial \ln L}{\partial \rho} &= \sum_{i=1}^n \frac{q_{i1} \cdot q_{i2} \cdot \varphi_2}{\Phi_2} = 0 \end{aligned} \tag{3.23}$$

где  $i = 1, 2, \dots, n$ ,

$$g_{i1} = \varphi(w_{i1}) \cdot \Phi\left(\frac{w_{i2} - \rho_{i^*} \cdot w_{i1}}{\sqrt{1 - \rho_{i^*}^2}}\right),$$

$$g_{i2} = \varphi(w_{i2}) \cdot \Phi\left(\frac{w_{i1} - \rho_{i^*} \cdot w_{i2}}{\sqrt{1 - \rho_{i^*}^2}}\right).$$

Для исследования системы уравнений (3.23) строится матрица Гессе с элементами:

$$\frac{\partial^2 \ln L}{\partial \beta_1 \partial \beta_1^T} = \sum_{i=1}^n x_{i1} \cdot x_{i1}^T \cdot \left[ \frac{-w_{i1} \cdot g_{i1}}{\Phi_2} - \frac{\rho_{i^*} \cdot \varphi_2}{\Phi_2} - \frac{g_{i1}^2}{\Phi_2^2} \right]$$

$$\frac{\partial^2 \ln L}{\partial \beta_2 \partial \beta_2^T} = \sum_{i=1}^n x_{i2} \cdot x_{i2}^T \cdot \left[ \frac{-w_{i2} \cdot g_{i2}}{\Phi_2} - \frac{\rho_{i^*} \cdot \varphi_2}{\Phi_2} - \frac{g_{i2}^2}{\Phi_2^2} \right]$$

$$\frac{\partial^2 \ln L}{\partial \beta_1 \partial \beta_2^T} = \sum_{i=1}^n q_{i1} \cdot q_{i2} \cdot x_{i1} \cdot x_{i2}^T \cdot \left[ \frac{\varphi_2}{\Phi_2} - \frac{g_{i1} \cdot g_{i2}}{\Phi_2} \right]$$

$$\frac{\partial^2 \ln L}{\partial \beta_1 \partial \rho} = \sum_{i=1}^n q_{i2} \cdot x_{i1} \cdot \frac{\varphi_2}{\Phi_2} \cdot \left[ \rho_{i^*} \cdot \delta_i \cdot v_{i1} - w_{i1} - \frac{g_{i1}}{\Phi_2} \right]$$

$$\frac{\partial^2 \ln L}{\partial \beta_2 \partial \rho} = \sum_{i=1}^n q_{i1} \cdot x_{i2} \cdot \frac{\varphi_2}{\Phi_2} \cdot \left[ \rho_{i^*} \cdot \delta_i \cdot v_{i2} - w_{i2} - \frac{g_{i2}}{\Phi_2} \right]$$

$$\frac{\partial^2 \ln L}{\partial \rho^2} = \sum_{i=1}^n \frac{\varphi_2}{\Phi_2} \cdot \left[ \delta_i^2 \cdot \rho_{i^*} \cdot (1 - \delta_i^2 \cdot (w_{i1}^2 + w_{i2}^2 - 2 \cdot \rho_{i^*} \cdot w_{i1} \cdot w_{i2})) + \delta_i^2 \cdot w_{i1} \cdot w_{i2} - \frac{\varphi_2}{\Phi_2} \right]$$

где  $\delta_i = \frac{1}{\sqrt{1 - \rho_{i^*}^2}}$ ,

$$v_{i1} = \delta_i \cdot (w_{i2} - \rho_{i^*} \cdot w_{i1}) \Rightarrow g_{i1} = \varphi(w_{i1}) \cdot \Phi(v_{i1}),$$

$$v_{i2} = \delta_i \cdot (w_{i1} - \rho_{i^*} \cdot w_{i2}) \Rightarrow g_{i2} = \varphi(w_{i2}) \cdot \Phi(v_{i2}).$$



Анализ оцененной двумерной пробит-модели помимо проверки значимости модели в целом и проверки значимости коэффициентов модели включает в себя проверку равенства нулю коэффициента корреляции ошибок и оценку предельных эффектов.

Проверка равенства нулю коэффициента корреляции ошибок часто основана на использовании LM-теста (теста множителей Лагранжа):

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

При справедливости нулевой гипотезы о незначимости коэффициента двумерная пробит-модель распадается на две независимых пробит-модели, которые могут быть оценены по отдельности. Статистика критерия имеет вид:

$$LM = \frac{\left[ \sum_{i=1}^n q_{i1} q_{i2} \frac{\varphi(w_{i1}) \cdot \varphi(w_{i2})}{\Phi(w_{i1}) \cdot \Phi(w_{i2})} \right]^2}{\sum_{i=1}^n \frac{[\varphi(w_{i1}) \varphi(w_{i2})]^2}{\Phi(w_{i1}) \cdot \Phi(-w_{i1}) \cdot \Phi(w_{i2}) \cdot \Phi(-w_{i2})}}. \quad (3.24)$$

При справедливой нулевой гипотезе для двумерной модели статистика распределена по закону хи-квадрат с числом степеней свободы, равным 1.

Для определения предельных эффектов введем в рассмотрение вектор  $X_i$ , содержащий значения независимых переменных, характеризующих объект  $i$ , для обоих уравнений, то есть являющийся объединением векторов  $X_{i1} \cup X_{i2}$ .

Нам понадобятся также такие векторы коэффициентов  $\gamma_1$  и  $\gamma_2$ , что  $X_{i2} \cdot \beta_1 = X_i \cdot \gamma_1$  и  $X_{i2} \cdot \beta_2 = X_i \cdot \gamma_2$ , то есть  $\gamma_1$  составляется из элементов вектора коэффициентов  $\beta_1$  и нулей, стоящих на позициях, которые соответствуют переменным второго уравнения, а  $\gamma_2$  - из элементов вектора коэффициентов  $\beta_2$  и нулей, стоящих на позициях, которые соответствуют переменным первого уравнения.

Тогда вероятность того, что значения  $y_1$  и  $y_2$  одновременно будут равны единице, определяется как:

$$P(y_{i1} = 1, y_{i2} = 1 | X_i) = \Phi_2[X_i \cdot \gamma_1, X_i \cdot \gamma_2, \rho_{i*}]. \quad (3.25)$$

Предельные эффекты факторов для  $P(y_{i1} = 1, y_{i2} = 1 | X_i)$  могут быть определены согласно выражению (3.26):

$$\frac{\partial \Phi_2}{\partial X_i} = g_{1i} \cdot \gamma_1 + g_{2i} \cdot \gamma_2. \quad (3.26)$$

Условные математические ожидания зависимых переменных  $y_{i1}$  и  $y_{i2}$  для конкретных наборов независимых переменных  $x_i$  определяются по формуле (3.27):

$$\begin{aligned} M[y_{i1} | X_i] &= \Phi(X_i \cdot \gamma_1) \\ M[y_{i2} | X_i] &= \Phi(X_i \cdot \gamma_2) \end{aligned} \quad (3.27)$$

Для модели (3.22) можно также определить условное математическое ожидание зависимой переменной первого уравнения при условии, что  $y_{i2} = 1$ :

$$M[y_{i1} | y_{i2} = 1, X_i] = P[y_{i1} = 1 | y_{i2} = 1, X_i] = \frac{P[y_{i1} = 1, y_{i2} = 1 | X_i]}{P[y_{i2} = 1 | X_i]} = \frac{\Phi_2(X_i \cdot \gamma_1, X_i \cdot \gamma_2, \rho_{i*})}{\Phi(X_i \cdot \gamma_2)} \quad (3.28a)$$

Аналогично определяется условное математическое ожидание зависимой переменной второго уравнения при условии, что  $y_{i1} = 1$ :

$$M[y_{i2} | y_{i1} = 1, X_i] = P[y_{i2} = 1 | y_{i1} = 1, X_i] = \frac{P[y_{i1} = 1, y_{i2} = 1 | X_i]}{P[y_{i1} = 1 | X_i]} = \frac{\Phi_2(X_i \cdot \gamma_1, X_i \cdot \gamma_2, \rho_{i*})}{\Phi(X_i \cdot \gamma_1)} \quad (3.28b)$$

Оценивание двумерных пробит-моделей, проверку всех необходимых гипотез и оценку предельных эффектов можно выполнять в статистическом пакете Stata.

### **3.4 Ранжирование категорий безработных по степени их востребованности на рынке труда на основе моделей бинарного выбора**

Задача ранжирования объектов, характеризующихся многомерным вектором признаков, очень востребована на практике. Чаще всего, как отмечено в подразделе 2.3, эта задача решается на основе построения интегрального показателя эффективности функционирования (уровня развития) в форме работоспособной первой главной компоненты или экспертно-статистическим методом. Однако оба эти подхода традиционно предполагают, что исходные признаки являются количественными. Большинство же признаков, характеризующих безработных, зарегистрированных в Центре занятости населения в целях поиска работы, являются качественными. Так, например, к ним относятся следующие показатели:

- пол: мужской, женский;
- возраст: до 29 лет, от 30 до 49 лет, от 50 лет и старше;
- образование: высшее, среднее, начальное, отсутствие профессионального образования;
- профессия: рабочий, служащий, нет профессии;
- наличие опыта работы: есть опыт работы, нет опыта работы.

Рассмотренный в этом параграфе подход к ранжированию категорий безработных по степени их востребованности на рынке труда основывается на модели бинарного выбора, позволяющей учесть различные качественные характеристики безработного [7].

Ставится задача провести моделирование вероятности трудоустройства безработного, зарегистрированного в Центре занятости населения в целях поиска работы, на основе модели бинарного выбора, в которой результирующий показатель

у принимает одно из двух возможных состояний: 1 – «трудоустроен в течение 30 дней», 0 – «не трудоустроен в течение 30 дней». Безработный, зарегистрированный в Центре занятости населения в целях поиска работы, характеризуется следующими признаками: пол, возраст, образование, профессия, наличие опыта работы. Поскольку все перечисленные признаки являются качественными, то для построения модели сформирован следующий набор индикаторных объясняющих переменных:

$x_1 = 1$ , если пол женский,  $x_1 = 0$ , если пол мужской;

$x_2 = 1$ , если возраст от 30 до 49 лет,  $x_2 = 0$ , если другой возраст;

$x_3 = 1$ , если возраст 50 лет и старше,  $x_3 = 0$ , если другой возраст;

$x_4 = 1$ , если высшее образование,  $x_4 = 0$ , если другое образование;

$x_5 = 1$ , если среднее образование,  $x_5 = 0$ , если другое образование;

$x_6 = 1$ , если нет профессионального образования,  $x_6 = 0$ , иначе;

$x_7 = 1$ , если профессия «служащий»,  $x_7 = 0$ , если другая профессия;

$x_8 = 1$ , если нет профессии,  $x_8 = 0$ , иначе;

$x_9 = 1$ , если нет опыта работы,  $x_9 = 0$ , если есть опыт работы.

С помощью пакета Statistica по данным Центра занятости населения города Оренбурга за 2000-2005 годы найдена оценка логит-модели бинарного выбора, имеющая вид:

$$\hat{P}(y_i = 1 | X_i) = \frac{1}{1 + \exp(-(1,31 - 0,45x_{i1} - 0,29x_{i2} - 0,99x_{i3} - 0,31x_{i4} - 0,23x_{i5} + 0,97x_{i6} - 0,79x_{i7} + 0,43x_{i8} + 0,64x_{i9}))}$$

Все коэффициенты модели по тесту Вальда оказались значимы. С помощью построенного уравнения правильно классифицированы 73% наблюдений, что говорит о достаточно высоком качестве построенной модели. Фрагмент результатов ранжирования категорий безработных, различающихся по полу, возрасту, образованию, профессии, наличию опыта работы, по снижению оценки вероятности трудоустройства представлен в таблице 3.1.

Таблица 3.1 – Фрагмент результатов ранжирования категорий безработных по снижению оценки вероятности трудоустройства

Категория безработных					Оценка вероятности трудоустройства
Пол	Возраст	Образование	Профессия	Опыт работы	
Мужской	До 29	Нет	Нет	Нет	0,9661
Мужской	От 30 до 49	Нет	Нет	Нет	0,9550
Мужской	До 29	Нет	Рабочая	Нет	0,9489
Женский	До 29	Нет	Нет	Нет	0,9476
Мужской	До 29	Нет	Нет	Есть	0,9375
Мужской	От 30 до 49	Нет	Рабочая	Нет	0,9327
Женский	От 30 до 49	Нет	Нет	Нет	0,9310
Женский	До 29	Нет	Рабочая	Нет	0,9219
Мужской	От 30 до 49	Нет	Нет	Есть	0,9180
Мужской	50 и старше	Нет	Нет	Нет	0,9134
.....					
Женский	До 29	Высшее	Служащий	Есть	0,4390
Женский	50 и старше	Среднее	Рабочая	Есть	0,4071
Женский	От 30 до 49	Среднее	Служащий	Нет	0,3858
Женский	50 и старше	Среднее	Служащий	Есть	0,3718
Женский	50 и старше	Среднее	Служащий	Нет	0,3718
Женский	От 30 до 49	Высшее	Служащий	Есть	0,3687
Мужской	50 и старше	Среднее	Служащий	Есть	0,3293
Мужской	50 лет и старше	Высшее	Служащий	Есть	0,3134
Женский	50 лет и старше	Среднее	Служащий	Есть	0,2378
Женский	50 лет и старше	Высшее	Служащий	Есть	0,2248

Анализ таблицы позволяет сделать следующие выводы. Наибольшая вероятность трудоустройства в течение 30 дней пребывания на учете в Центре занятости населения принадлежит преимущественно безработным мужчинам в возрасте до 49 лет, без образования, без профессии или с рабочей профессией, не имеющим опыта работы. Наименьшей вероятностью трудоустройства в течение 30 дней пребывания на учете в Центре занятости населения будет преимущественно для женщин старше 50 лет с высшим и средним профессиональным образованием, с профессией «служащий», с опытом работы.

Несмотря на отмеченное ранее преимущество реализованного подхода к ранжированию, следует отметить его недостаток, связанный с невозможностью непосредственной интерпретации количественных оценок параметров построенной нелинейной модели.

### 3.5 Построение интегрального показателя, характеризующего экономическую безопасность региона, на основе моделей бинарного выбора

Построение интегрального показателя, характеризующего уровень экономической безопасности, в случае, если результирующая переменная принимает только целочисленные значения, возможно на основе моделей множественного выбора. Многомерная классификация муниципальных образований позволяет получить такую дискретную результирующую переменную.

В результате классификации городов и районов Оренбургской области по показателям, характеризующим экономическую безопасность региона, выявили два кластера с разными уровнями исследуемой латентной категории.

Среди предложенных показателей целесообразным является выявить те, которые оказывают существенное влияние на уровень экономической безопасности.

В нашем случае, в качестве результирующего признака возьмем дихотомическую переменную, так как ранее разбили муниципальные образования региона на два класса. Поэтому для выявления показателей, влияющих на экономическую безопасность, используем логит - модель:

$$P(y_i = 1 | X_i) = \frac{e^{X_i^T \beta}}{1 + e^{X_i^T \beta}} = \Lambda(X_i^T \beta), \quad (3.29)$$

где  $y_i = \begin{cases} 1, & \text{если } i\text{-ый объект принадлежит к первому классу,} \\ 0, & \text{если } i\text{-ый объект принадлежит к второму классу.} \end{cases}$

$X_i = (1, x_{i1}, \dots, x_{ik})^T$  - вектор значений независимых переменных для  $i$ -го наблюдения;

Для построения логит-модели (3.29) использовалась процедура пошаговой регрессии, основанная на включении в модель наиболее существенных факторов

по тесту Вальда (аналог t- критерия Стьюдента). Таким образом, оценка модели бинарного выбора имеет вид:

$$\hat{P}(y_i = 1 | X_i) = \frac{\exp(291,01 + 0,101x_{i1} - 0,17x_{i2} + 0,066x_{i4} + 0,045x_{i5} + 0,087x_{i6} + 0,091x_{i7} + 0,341x_{i78})}{1 + \exp(291,01 + 0,101x_{i1} - 0,17x_{i2} + 0,066x_{i4} + 0,045x_{i5} + 0,087x_{i6} + 0,091x_{i7} + 0,341x_{i78})}$$

pseudo- $R^2=0,813$ .

По аналогу коэффициента  $R^2$  можно говорить о хорошем качестве построенной модели.

По критерию Вальда проверяется гипотеза: существенно ли отличны от нуля коэффициенты регрессии при заданном уровне значимости 0,05, который подчиняется распределению хи-квадрат. Результаты проверки гипотезы сведем в таблицу 3.2.

Таблица 3.2 - Проверка гипотез о незначимости коэффициентов в логит-модели

Коэффициенты модели	Значение критерия Вальда	Вероятность принятия нулевой гипотезы	Вывод
$\beta_1$	4,41	0,00036	гипотеза отвергается
$\beta_2$	9,68	0,00002	гипотеза отвергается
$\beta_3$	4,49	0,00034	гипотеза отвергается
$\beta_4$	3,52	0,00081	гипотеза отвергается
$\beta_5$	4,45	0,00037	гипотеза отвергается
$\beta_6$	8,08	0,00004	гипотеза отвергается
$\beta_7$	5,43	0,00014	гипотеза отвергается

Из таблицы видно, что коэффициенты модели являются статистически значимыми.

Для проверки адекватности подобранной модели имеющимся данным имеется ряд статистических критериев согласия; одним из них является критерий Хосмера–Лемешоу. Отметим, что этот критерий основан на сравнении предсказываемых моделью и действительно наблюдаемых количеств случаев с  $y_i = 1$  в нескольких группах, на которые разбивается множество наблюдений. Значение ста-

тистики Хосмера–Лемешоу (рассчитанный по асимптотическому распределению хи-квадрат с соответствующим числом степеней свободы) равно 1,632,  $p = 0,938$ .

О качестве построенной модели также можно судить по таблице классификации (таблица 3.3). Из таблицы видно, что модель правильно прогнозирует вероятность того, что 36 городов и районов области относятся к классу с более низким уровнем экономической безопасности (100%). Модель правильно также предсказала вероятность того, что 6 объектов из 7 (85,7%) относятся к классу с более высоким уровнем экономической безопасности. Элементы, расположенные вне главной диагонали, показывают количество неверно классифицированных объектов. В данном случае таких объектов 1. Всего было правильно классифицировано 42 объекта из 43 (97,7%).

Таблица 3.3 – Отнесение объекта к классу с более высоким или низким уровнем экономической безопасностью региона в 2007 г.

	Предсказание отнесения объекта к классу с более высоким уровнем экон. безопасности	Предсказание отнесения объекта к классу с более низким уровнем экон. безопасности	Процент корректных предсказаний, %
Отнесение объекта к классу с более высоким уровнем экон. безопасности	6	1	85,7
Отнесение объекта к классу с более низким уровнем экон. безопасности	0	36	100
Общий процент			97,7

Из построенной модели следует, что на экономическую безопасность существенное влияние оказывают все рассмотренные показатели, кроме темпов роста числа предприятий и организаций, темпов роста численности населения в трудоспособном возрасте и удельного веса убыточных организаций. При этом, положительное влияние оказывают: показатель, характеризующий размеры внешней трудовой миграции, темп роста среднесписочной численности работников, кредиторская задолженность на душу населения, оборот розничной торговли на душу населения, средненоминальная заработная плата работников и инвестиции в основной капитал, а негативное влияние оказывает уровень безработицы. К примеру, при



увеличении числа зарегистрированных иностранных работников на 1 человека, вероятность того, что объект будет принадлежать к классу с сравнительно более высоким уровнем экономической безопасности, увеличится на 0,0227, т.е. составит 0,4833. Вклад иностранной рабочей силы в экономическое развитие региона в целом определяется как объем отгруженной продукции (работ, услуг) по фактическим видам экономической деятельности, умноженный на долю иностранной рабочей силы в численности занятых в экономике. Эффективность использования иностранной рабочей силы за последние три года увеличилась в примерно в 5 раз и в 2007 г. составила 3549,27 млн. руб., или 0,98% стоимости всех отгруженных товаров собственного производства, выполненных работ и оказанных услуг собственными силами. Влияние внешней трудовой миграции на экономическую безопасность опосредованно осуществляется и через остальные объясняющие переменные, вошедшие в модель, в силу их тесной корреляционной зависимости с показателем зарегистрированных иностранных работников. В частности, существенное влияние на уровень экономической безопасности оказывает оборот розничной торговли, который тесно связан с показателем, отражающим размеры внешней трудовой миграции.

Аналогичные модели, построенные за 2004-2006 гг. (приложение П), показали разную степень влияния на уровень экономической безопасности одного и того же набора показателей.

Помимо выявления факторов, оказывающих влияние на экономическую безопасность, на основе модели бинарного выбора получили интегральный показатель, позволяющий осуществить ранжирование муниципальных образований региона по вычисленным значениям вероятности отнесения объекта к классу с сравнительно более высоким уровнем экономической безопасности (таблица 3.4).

Все города области, кроме г. Бугуруслана, г. Медногорска и г. Сорочинска, относятся к классу со сравнительно более высоким уровнем экономической безопасности. Наибольшая вероятность отнесения объекта к классу со сравнительно более высоким уровнем экономической безопасности принадлежит г. Оренбургу и он занимает первую рейтинговую позицию.

Таблица 3.4 – Ранжирование объектов по вычисленным значениям вероятности отнесения объекта к классу с сравнительно более высоким уровнем экономической безопасности в 2007 г.

Города и районы	Вероятность отнесения объекта к классу с более высоким уровнем экономической безопасности	Ранг
г. Оренбург	0,99254	1
г. Орск	0,94557	2
г. Новотроицк	0,92496	3
г. Бузулук	0,83136	4
Оренбургский район	0,77263	5
г. Медногорск	0,59419	6
г. Бугуруслан	0,27229	7
г. Ясный и район	0,20409	8
г. Гай	0,19281	9
г. Сорочинск	0,13265	10
г. Соль - Илецк и район	0,05145	11
г. Кувандык и район	0,02355	12
Абдулинский район	0,02035	13
Новосергеевский район	0,01265	14
Новоорский район	0,01257	15
Тоцкий район	0,00952	16
Первомайский район	0,00914	17
Переволоцкий район	0,00830	18
Курманаевский район	0,00797	19
Пономаревский район	0,00784	20
Сакмарский район	0,00647	21
Адамовский район	0,00623	22
Краснагвардейский район	0,00561	23
Октябрьский район	0,00529	24
Саракташский район	0,00497	25
Шарлыкский район	0,00496	26
Ташлинский район	0,00492	27
Тюльганский район	0,00414	28
Северный район	0,00400	29
Акбулакский район	0,00310	30
Светлинский район	0,00284	31
Кваркенский район	0,00281	32
Грачевский район	0,00255	33
Александровский район	0,00225	34
Илекский район	0,00210	35
Беляевский район	0,00204	36
Бузулукский район	0,00202	37
Матвеевский район	0,00179	38
Асекеевский район	0,00134	39
Бугурусланский район	0,00117	40
Домбаровский район	0,00116	41
Сорочинский район	0,00103	42
Гайский район	0,00078	43

Сравнительно более высокий уровень экономической безопасности имеют города области и Оренбургский район. По уровню экономической безопасности города области существенно опережают сельскую местность: здесь сосредоточены предприятия, предоставляющие торговые, социальные и финансовые услуги. Легко объяснить, почему сельский Оренбургский район имеет более высокий уровень экономической безопасности, по сравнению с экономической безопасностью других сельских районов: примыкая к областному центру, он использует преимущество своего положения. Сравнительно более низкий уровень экономической безопасности наблюдается практически у всех сельских районов области. Сокращение производства сельскохозяйственной продукции и убыточность 60-70 % сельскохозяйственных предприятий объясняется медленной адаптацией сельского населения к рыночным отношениям, сокращением инвестиций в аграрный сектор экономики и ухудшением материально – технической базы сельского хозяйства [8].

### **3.6 Моделирование и прогнозирование эколого-экономического риска на основе моделей бинарного и множественного выбора**

В связи с существенным различием городов и районов области по величине эколого-экономического риска и невозможностью объяснения данных различий только изменениями антропогенной нагрузки на окружающую среду, возникает задача построения модели, связывающей изменения уровня эколого-экономического риска с изменением показателей социально-экономического развития муниципальных образований области с целью прогнозирования уровня риска.

В качестве показателей, предположительно оказывающих влияние на уровень эколого-экономического риска, предлагается взять:

– оборот розничной торговли (ROZN, тыс.руб.) и среднюю заработную плату (LOGZP, тыс. руб.) для характеристики материального благосостояния населения города/района;

- стоимость выпущенной продукции обрабатывающих отраслей (OBR, млн.руб.) для характеристики производства;
- долю убыточных предприятий (DUB,%), исходя из предположения о том, что предприятия, находящиеся в тяжелом финансовом положении, менее щепетильны в вопросах заботы об окружающей среде;
- новообразования (NOV1), осложнения беременности и родов (BER1) и детскую инвалидность (INVD, на 10 000 детей в возрасте до 17 лет) для характеристики заболеваемости населения, причем NOV1 и BER1 представить бинарными переменными, принимающими значение 1, если уровень соответствующей заболеваемости новообразованиями (осложнений беременности) в  $i$ -ом городе/районе превышает среднеобластной, и 0 – если нет;
- коэффициент миграционного прироста/снижения (MIGR, на 1000 человек) для характеристики реакции населения на условия проживания в городе/районе;
- фиктивную переменную CITY, принимающая значение 1 для городов или районов, центром которых является крупный город для учета пространственной неоднородности данных.

Зависимость между уровнем эколого-экономического риска и отобранными показателями будем искать с помощью модели:

$$y_i = \begin{cases} 1, & y_i^* \leq \gamma_1 \\ 2, & \gamma_1 < y_i^* \leq \gamma_2, \\ 3, & y_i^* > \gamma_2 \end{cases} \quad (3.30)$$

где  $i = 1, \dots, n$ ,

$n$  - количество анализируемых городов и районов,

$$y_i = \begin{cases} 1, & \text{если уровень эколого - экон. риска в городе/районе } i \text{ высокий} \\ 2, & \text{если уровень эколого - экон. риска в городе/районе } i \text{ средний} \\ 3, & \text{если уровень эколого - экон. риска в городе/районе } i \text{ низкий} \end{cases}$$

$y^*$  - ненаблюдаемая переменная  $y_i^* = \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i$

$$x = (ROZN, LOGZP, OBR, DUB, NOV1, BER1, INVD, MIGR, CITY)^T ;$$

$\gamma_1 < \dots < \gamma_j < \dots < \gamma_{k-1}$  - подлежащие оценке пороговые значения;

$\varepsilon_i$  - независимо и одинаково распределенные случайные величины с функцией

распределения  $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}} dt$ , учитывающие влияние неучтенных факторов.

Результаты оценивания моделей в пакете Eviews приведены в таблице 3.5, вероятности нахождения  $i$ -го объекта в классе высокого, среднего и низкого уровня эколого-экономического риска представлены формулой 3.30а.

Таблица 3.5 – Результаты оценивания порядковой пробит-модели, характеризующей взаимосвязь эколого-экономического риска и показателей социально-экономического состояния города/района

Переменная	Оценка	Станд. ошибка	z-стат.	p-значение
Коэффициенты				
CITY	-1,0130	0,3745	-2,7052	0,0068
ROZN	-0,2583	0,1421	-1,8180	0,0691
OBR	-0,0002	0,0001	-1,7868	0,0740
LOG(ZP)	-1,0988	0,5654	-1,9436	0,0519
1/INVD	48,7686	20,5219	2,3764	0,0175
1/DUB	9,1641	5,6394	1,6250	0,1042
1/MIGR	0,0262	0,0146	1,7927	0,0730
NOV1	-0,3361	0,3046	-1,1034	0,2698
BER1	-0,7551	0,3073	-2,4570	0,0140
Пороговые значения				
$\gamma_1$	-2,3163	0,5157	-4,4916	0,0000
$\gamma_2$	-0,8344	0,4517	-1,8471	0,0647

$$\begin{aligned}
\hat{P}\{y_i = 1|X_i\} &= \Phi(-2,32 - 1,01 \cdot CITY_i - 0,26 \cdot ROZN_i - 0,0002 \cdot OBR_i - 1,10 \cdot LOGZP_i + \\
&\quad + \frac{48,77}{INVD_i} + \frac{9,16}{DUB_i} + \frac{0,03}{MIRG_i} - 0,34 \cdot NOV1 - 0,76 \cdot BER1) \\
\hat{P}\{y_i = 2|X_i\} &= \Phi(-0,83 - 1,01 \cdot CITY_i - 0,26 \cdot ROZN_i - 0,0002 \cdot OBR_i - 1,10 \cdot LOGZP_i + \\
&\quad + \frac{48,77}{INVD_i} + \frac{9,16}{DUB_i} + \frac{0,03}{MIRG_i} - 0,34 \cdot NOV1 - 0,76 \cdot BER1) - \\
&\quad - \Phi(-2,32 - 1,01 \cdot CITY_i - 0,26 \cdot ROZN_i - 0,0002 \cdot OBR_i - 1,10 \cdot LOGZP_i + \\
&\quad + \frac{48,77}{INVD_i} + \frac{9,16}{DUB_i} + \frac{0,03}{MIRG_i} - 0,34 \cdot NOV1 - 0,76 \cdot BER1) \\
\hat{P}\{y_i = 3|X_i\} &= 1 - \Phi(-0,83 - 1,01 \cdot CITY_i - 0,26 \cdot ROZN_i - 0,0002 \cdot OBR_i - 1,10 \cdot LOGZP_i + \\
&\quad + \frac{48,77}{INVD_i} + \frac{9,16}{DUB_i} + \frac{0,03}{MIRG_i} - 0,34 \cdot NOV1 - 0,76 \cdot BER1)
\end{aligned} \tag{3.30a}$$

Таким образом, для городов, а также районов с относительно большим розничным товарооборотом, высокой долей убыточных предприятий характерен высокий уровень эколого-экономического риска населению. В муниципалитетах с высоким уровнем эколого-экономического риска наблюдается повышенный по сравнению со среднеобластным уровень заболеваемости новообразованиями, осложнений беременности и высокий уровень детской инвалидности. Повышенный миграционный прирост, высокий уровень заработной платы также влекут за собой повышение уровня эколого-экономического риска, что объясняется устремлением людей в более экономически активные города/ районы, развитие которых, однако, пока идет по экстенсивному пути хозяйствования.

Поскольку лицам, принимающим решения, часто необходимо только различать объекты, для которых характерно наличие риска, и те, для которых риска нет или он минимален, то имеет смысл при моделировании взаимосвязи между эколого-экономическим риском и социально-экономическим состоянием города или района в качестве объясняемой переменной в модели (3.30) рассматривать бинарную переменную  $y_i$ , принимающую значение 1, если объект  $i$  характеризуется низким уровнем эколого-экономического риска, и 0 – если иначе (высоким или средним). Исследования показали целесообразность введения в модель переменной TRAVMPR (травматизм на производстве, чел), а также, для учета пространст-

венной и временной неоднородности, фиктивных переменных OSf2 (принимает значение 1 для городов и районов со средним уровнем нагрузки на окружающую среду и 0 – в остальных случаях) и T04 (принимает значение 1 для 2005-2007 гг. и 0 для 2003-2004 гг., поскольку в 2004-2005 гг. произошло резкое увеличение количества муниципалитетов, население которых подвергается высокому эколого-экономическому риску).

Результаты оценивания в пакете EViews приведены в таблице 3.6, вероятность нахождения  $i$ -го объекта в классе высокого или среднего уровня эколого-экономического риска имеет вид формулы 3.30б. Для интерпретации коэффициентов модели оценены предельные эффекты всех вошедших в модель факторов, измеренных в количественной шкале (таблицы 3.7 - 3.8) и номинальной шкале (таблица 3.9). При этом для расчета предельных эффектов каждой из четырех вошедших бинарных переменных сначала рассчитывалась вероятность того, что зависимая переменная принимает значение 1 для всех наблюдений, где исследуемая фиктивная переменная принимает значение 1, и находилось среднее. Аналогично рассчитывалась вероятность для наблюдений, где исследуемая фиктивная переменная принимает значение 0. Разница между этими средними бралась в качестве предельного эффекта исследуемой бинарной переменной.

Таблица 3.6 – Результаты оценивания бинарной пробит-модели, характеризующей взаимосвязь эколого-экономического риска и показателей социально-экономического состояния города/района

Переменная	Оценка	Станд. ошибка	z-стат.	p
CONST	2,5893	0,7179	3,6068	0,0003
ROZN	-0,3674	0,1417	-2,5931	0,0095
LOGZP	-1,9801	0,5935	-3,3365	0,0008
DUB	-0,0151	0,0065	-2,3396	0,0193
TRAVMPR	-0,0891	0,0464	-1,9206	0,0548
1/NOV	1,5614	1,4642	1,0664	0,2862
BER1	-0,4911	0,2290	-2,1444	0,0320
OSf2	0,5937	0,3357	1,7685	0,0770
CITY	-1,0979	0,3772	-2,9108	0,0036
T02	-0,4402	0,2828	-1,5567	0,1195

Вероятность нахождения  $i$ -го объекта в классе высокого или среднего уровня эколого-экономического риска:

$$\hat{P}\{y_i = 1|X_i\} = \Phi(2,59 - 1,01 \cdot CITY_i - 0,37 \cdot ROZN_i - 1,98 \cdot LOGZP_i - 0,02 \cdot DUB_i - 0,09 \cdot TRAMPR_i + \frac{1,56}{NOV_i} - 0,49 \cdot BER1 + 0,59 \cdot OSf2 - 0,44 \cdot T02) \quad (3.306)$$

Таблица 3.7 – Оценки предельных эффектов непрерывных объясняющих переменных, рассчитанные за период 2003-2007 гг.

Переменная	Для объектов класса высокого и среднего риска		Для объектов класса низкого риска
	для объектов класса высокого риска	для объектов класса среднего риска	
ROZN	-0,082		-0,107
	-0,060	-0,098	
LOG(ZP)	-0,440		-0,577
	-0,323	-0,527	
DUB	-0,003		-0,004
	-0,002	-0,004	
TRAVMPR	-0,020		-0,026
	-0,015	-0,024	
1/NOV	0,347		0,455
	0,254	0,416	

Таблица 3.8 – Оценки предельных эффектов непрерывных объясняющих переменных, рассчитанные за периоды 2003-2004 гг и 2005-2007 гг

Переменная	Период					
	2003-2004 гг (T04=0)		Для объектов класса низкого риска	2005-2007 гг (T04=1)		Для объектов класса низкого риска
	для объектов класса высокого и среднего риска	для объектов класса среднего риска		для объектов класса высокого и среднего риска	для объектов класса среднего риска	
ROZN	-0,093		-0,097	-0,076		-0,117
	-0,077	-0,104		-0,053	-0,095	
LOG(ZP)	-0,502		-0,523	-0,411		-0,632
	-0,413	-0,560		-0,284	-0,511	
DUB	-0,004		-0,040	-0,003		-0,005
	-0,003	-0,004		-0,002	-0,004	
TRAVMPR	-0,023		-0,024	-0,018		-0,028
	-0,019	-0,025		-0,013	-0,023	
1/NOV	0,396		0,413	0,324		0,498
	0,326	0,442		0,224	0,403	



Из таблиц видно, что в среднем за период с 2003 по 2007 год влияние всех исследуемых переменных приблизительно в 1,3 раза больше для городов/районов с низким уровнем эколого-экономического риска, чем для городов/районов с более высоким уровнем риска (таблица 3.7), причем со временем эта разница увеличивается: если до 2004 г. влияние практически одинаково, то после 2004 года разница достигает полутора раз (таблица 3.8). Так, повышение логарифма среднемесячной заработной платы (LOGZP) в городе/районе с высоким или средним уровнем эколого-экономического риска уменьшает вероятность перехода в класс с более низким уровнем риска на 44%, а для объекта с низким уровнем риска - уменьшает вероятность остаться в этом статусе почти на 58% , причем до 2004 года на 52%, а после 2004 года – на 63%. Рост розничного товарооборота (ROZN) уменьшает вероятность улучшения экологической ситуации (переход в класс низкого риска) на 8-10%, рост травматизма на производстве (TRAVMPR) – на 20 и 26 % для объектов из класса высокого (среднего) и низкого риска соответственно, рост доли убыточных предприятий (DUB) – на 0,3-0,4%. Повышение в городе или районе заболеваемости новообразованиями (NOV1) также снижает вероятность для этого муниципального образования оказаться в классе низкого эколого-экономического риска.

Если среди городов/районов с высоким и средним уровнем эколого-экономического риска выделить отдельно города/районы высокого и среднего риска, то видно, что большее влияние исследуемые переменные оказывают на класс среднего риска. Например, рост логарифма заработной платы в районе среднего риска уменьшает для этого района вероятность перейти в класс низкого риска на 53%, а в районе высокого риска – на 32%. Это объясняется тем, что при сложившейся экологической ситуации в городах и районах с высоким уровне эколого-экономического риска изменения в лучшую сторону требуют более длительных мероприятий, серьезной и долгосрочной политики.

Таблица 3.9 – Предельные эффекты бинарных объясняющих переменных

Класс объектов	Переменная		
	BER1	CITY	T02
с высоким и средним уровнем эколого-экономического риска	0,065	-0,338	-0,230
с низким уровнем эколого-экономического риска	-0,123	-0,478	-0,179

Если мы видим, что для некоторого города/района с низким уровнем эколого-экономического риска, уровень осложнений беременности и родов превысил среднеобластной (BER1), то можно говорить, что для этого административно-территориального образования вероятность повышения эколого-экономического риска до высокого уровня на 12,3% больше, чем для административно-территориального образования, в котором уровень осложнений не превышает среднего по области. Появление в районе, характеризующемся низким уровнем риска, крупного города-центра (CITY) означало бы для него резкое увеличение вероятности ухудшения экологической ситуации (перехода в класс более высокого эколого-экономического риска) - почти на 48%. Влияние времени (T02) более выражено для муниципальных образований, относящихся к классу высокого и среднего эколого-экономического риска [9].

Анализ динамики розничного товарооборота, средней заработной платы и других показателей в целом за исследуемый период показывает наличие тенденции к их росту. В последние несколько лет после определенного снижения во многих муниципалитетах области происходит рост заболеваемости новообразованиями. Вместе с этим наметилась положительная тенденция сокращения доли убыточных предприятий. Для ответа на вопрос, приведут ли сложившиеся тенденции к снижению или росту эколого-экономического риска, осуществим прогнозирование его уровня, используя оцененные модели бинарного выбора.

Для построения прогнозов всех необходимых нам показателей, измеренных в количественной шкале, будем использовать один из адаптивных методов прогнозирования – экспоненциальное сглаживание. Суть метода состоит в том, прогнозируемое значение временного ряда представляет собой взвешенную сумму значений ряда во все предыдущие моменты времени, причем веса уменьшаются

экспоненциально по мере увеличения «возраста» наблюдения. Общая рекуррентная формула расчета простой экспоненциальной средней имеет вид [10]:

$$S_t = \alpha y_t + (1 - \alpha)S_{t-1}, \quad (3.31)$$

где  $S_t$  - значение экспоненциальной средней в момент времени  $t$ ;

$\alpha$  - параметр сглаживания (адаптации),  $0 < \alpha < 1$ ;

$y_t$  - значение временного ряда в момент времени  $t$ .

Модель (3.31) дает хорошие результаты при сглаживании и прогнозировании временных рядов, не имеющих тренда и сезонности, в противном случае ее использование приводит к смещенным прогнозам (при наличии тренда) или «выворачиванию» сезонной волны (при наличии сезонности). Для таких случаев разработаны модификации модели (3.31), также использующие принцип дисконтирования информации. Например, при наличии в ряду тренда и отсутствии сезонности, можно использовать модели линейного роста, в которых предполагается, что прогноз на  $\tau$  моментов времени вперед может быть получен в соответствии с формулой (3.32)

$$\hat{y}_t(\tau) = \hat{a}_{1,t} + \tau \hat{a}_{2,t}, \quad (3.32)$$

где  $\hat{a}_{1,t}, \hat{a}_{2,t}$  - текущие оценки коэффициентов адаптивного полинома первого порядка.

В частности, в двухпараметрической модели Хольта адаптация указанных коэффициентов производится по формуле (3.33):

$$\begin{aligned} \hat{a}_{1,t} &= \alpha_1 y_t + (1 - \alpha_1)(\hat{a}_{1,t-1} + \hat{a}_{2,t-1}) \\ \hat{a}_{2,t} &= \alpha_2 (\hat{a}_{1,t} - \hat{a}_{1,t-1}) + (1 - \alpha_2) \hat{a}_{2,t-1} \end{aligned}, \quad (3.33)$$

где  $\alpha_1, \alpha_2$  - параметры адаптации,  $0 < \alpha_1, \alpha_2 < 1$  [10].

Оценка моделей (3.31) и (3.32)-(3.33) сводится к поиску оптимальных значений параметров адаптации. Для их выбора существует ряд методов: от эвристических, предлагающих выбирать его исходя из горизонта прогнозирования и длины временного ряда, до поиска по сетке исходя из таких критериев, как минимум среднеквадратической или средней ошибки прогноза. Так, близость  $\alpha$  к единице означает, что при построении прогноза в основном учитывается влияние последних уровней временного ряда – это хорошо для конъюнктурных прогнозов; для долгосрочных прогнозов рекомендуется брать меньшую величину  $\alpha$ . Нами использовалась совокупность критериев; для поиска по сетке шаг был принят равным 0,1. Моделирование и прогнозирование осуществлялось в пакете Statistica, результаты приведены в таблице 3.10.

Таблица 3.10 – Прогнозные значения социально-экономических показателей на 2008-2010 гг. для отдельных районов Оренбургской области

Переменная	Прогноз		
	2008 год	2009 год	2010 год
1	2	3	4
<b>Беляевский район</b>			
Розничный товарооборот (тыс.руб., в сопоставимых ценах)	2,57	2,73	2,88
Среднемесячная заработная плата (тыс.руб., в сопоставимых ценах)	1,74	1,87	2,00
Доля убыточных предприятий (%)	45,65	43,88	42,11
Травматизм на производстве (на 1000 чел.)	7,30	7,27	7,24
Заболеваемость новообразованиями (на 1000 чел.)	8,42	7,69	6,96
Объем продукции обрабатывающих отраслей (млн.руб.)	30,65	34,07	37,49
Детская инвалидность (на 10000 чел. до 17 лет)	218,16	216,67	215,54
Коэффициент миграционного прироста	-2,57	-2,82	-3,07
<b>Новосергиевский район</b>			
Розничный товарооборот (тыс.руб., в сопоставимых ценах)	6,59	6,99	7,39
Среднемесячная заработная плата (руб., в сопоставимых ценах)	2,05	2,21	2,37
Доля убыточных предприятий (%)	24,10	25,27	26,45
Травматизм на производстве (на 1000 чел.)	3,45	3,56	3,66
Заболеваемость новообразованиями (на 1000 чел.)	20,72	22,25	23,77
Объем продукции обрабатывающих отраслей (млн.руб.)	322,2	340,04	357,88
Детская инвалидность (на 10000 чел до 17 лет)	213,42	214,71	220,1
Коэффициент миграционного прироста	2,90	3,54	4,18
<b>Ташлинский район</b>			
Розничный товарооборот (тыс.руб., в сопоставимых ценах)	3,85	3,99	4,13
Среднемесячная заработная плата (руб., в сопоставимых ценах)	1,40	1,50	1,60
Доля убыточных предприятий (%)	27,72	25,89	24,07
Травматизм на производстве (на 1000 чел.)	3,18	2,41	1,83

Продолжение таблицы 3.10

1	2	3	4
Заболеваемость новообразованиями (на 1000 чел.)	13,24	13,57	13,91
Объем продукции обрабатывающих отраслей (млн.руб.)	577,6	628,7	679,8
Детская инвалидность (на 10000 чел. до 17 лет)	230,53	227,59	210,64
Коэффициент миграционного прироста	-1,56	-1,87	-2,19

Используем полученные прогнозные значения социально-экономических показателей для прогнозирования уровня эколого-экономического риска в рассматриваемых муниципальных образованиях Оренбургской области по модели бинарного выбора 3.30б. Результаты оценки вероятности того, что город/район будет относиться к классу низкого риска, приведены в таблице 3.11.

Таблица 3.11 – Вероятности низкого уровня эколого-экономического риска в 2008-2010 гг. для отдельных районов Оренбургской области

Район	Вероятность $\hat{P}\{y_i = 1 X_i\}$		
	2008 год	2009 год	2010 год
Беляевский	0,0621	0,0455	0,0336
Новосергиевский	0,0109	0,0044	0,0017
Ташлинский	0,1567	0,1358	0,1152

Таким образом, ни в одном из рассмотренных районов внешние условия не способствуют снижению уровня эколого-экономического риска. Для уточнения прогноза воспользуемся моделью упорядоченного множественного выбора 3.30а, где результативная переменная измерена в порядковой шкале. Результаты прогнозирования класса риска по модели представлены в таблице 3.12.

Таблица 3.12 - Вероятности на 2008-2010 гг. низкого, среднего и высокого уровня эколого-экономического риска в отдельных районах Оренбургской области

Уровень эколого-экономического риска	Вероятность		
	2008 год	2009 год	2010 год
1	2	3	4
<b>Беляевский район</b>			
высокий	0,2430	0,2786	0,3146
средний	0,5408	0,5360	0,5265
низкий	0,2162	0,1855	0,1589
<b>Новосергиевский район</b>			
высокий	0,4842	0,5670	0,6457
средний	0,4412	0,3836	0,3225
низкий	0,0746	0,0494	0,0318

Продолжение таблицы 3.12

1	2	3	4
<b>Ташлинский район</b>			
высокий	0,3935	0,4291	0,4566
средний	0,4937	0,4747	0,4585
низкий	0,1128	0,0963	0,0849

В 2007 году для Беляевского и Новосергиевского районов был характерен средний уровень эколого-экономического риска, причем ранг Новосергиевского района превышал ранг Беляевского. В 2008 году Новосергиевский район перемещается в класс высокого уровня эколого-экономического риска. Отметим, что в 2008 году для данного района еще близки вероятности его отнесения в класс высокого и среднего уровня риска, однако со временем разрыв существенно увеличивается до двух раз. Для Новосергиевского района можно прогнозировать снижение его ранга с 14 до 12 в общем рейтинге 2007 года.

Тенденции, сложившиеся в Беляевском районе, позволяют сделать более благоприятный прогноз – в период с 2008 по 2010 год не прогнозируется его переход в класс высокого уровня эколого-экономического риска. Вместе с тем, несмотря на то, что район находится в классе среднего риска, происходит уменьшение вероятностей отнесения его как к этому классу, так и к более благополучному классу низкого риска за счет роста вероятности отнесения его в класс риска высокого. Это отражает тот факт, что данный район находится в классе среднего риска, но для него наметилась значимая тенденция ухудшения рейтинга.

Согласно полученному результату, в 2008 году Ташлинский район перемещается в класс среднего уровня эколого-экономического риска, причем в 2010 году вероятности его принадлежности к классу среднего и высокого риска практически одинаковы.

Таким образом, использование моделей бинарного и множественного выбора позволяет проводить достаточно глубокий анализ факторов, влияющих на уровень эколого-экономического риска. А сочетание аппарата моделей бинарного и множественного выбора с аппаратом обработки временных рядов дает вполне работоспособный инструмент для прогнозирования уровня риска.

### **3.7 Построение интегральных показателей антропогенной нагрузки и эколого-экономического риска на основе моделей множественного выбора**

Рассмотрим пример использования моделей упорядоченного множественного выбора в качестве инструмента построения интегрального показателя. Будем моделировать две латентные категории – антропогенную нагрузку на окружающую среду и эколого-экономический риск.

Информационной базой для построения интегральных показателей будут выступать значения показателей, характеризующих города и районы Оренбургской области за период 2000-2007 гг., полученные из официальных статистических сборников Территориального органа Федеральной службы государственной статистики по Оренбургской области и г. Оренбургу и информация о содержании химических веществ в атмосферном воздухе и питьевой воде, полученная от ФГУЗ «Центр гигиены и эпидемиологии по Оренбургской области».

Для характеристики антропогенной нагрузки будем использовать четыре показателя: сброс сточных вод в поверхностные водоемы на единицу территории (SB, млн. куб. м. на 1 тыс. км<sup>2</sup>); выбросы загрязняющих веществ от стационарных источников на единицу территории (ST, тыс. тонн на 1 км<sup>2</sup>); количество автомобилей, находящихся в личной собственности граждан, на единицу территории (NEST, штук на 1 км<sup>2</sup>); объем токсичных отходов на единицу территории (TOX, тыс. тонн на 1 км<sup>2</sup>).

Для характеристики эколого-экономического риска на основе данных о среднегодовых концентрациях основных химических веществ, содержащихся в питьевой воде и/или атмосферном воздухе 35 районов и 4 городов Оренбургской области в 2000-2007 гг. были рассчитаны оценки рисков возникновения канцерогенных (для веществ, относящихся к канцерогенам по классификации МАИР) и неканцерогенных эффектов для человека от систематического воздействия. Для характеристики риска развития неканцерогенных эффектов при комбинированном и комплексном воздействии был осуществлен переход от коэффициентов опасности HQ к индексам опасности HI различным системам организма человека.

Нами предлагается для построения интегрального показателя эколого-экономического риска использовать модели упорядоченного множественного выбора. Для этого на основе полученных в главе 2 классификаций городов и районов области введем новые переменные:

а)  $y_i^{nagr}$  для характеристики уровня антропогенной нагрузки на среду в  $i$ -м городе/районе, принимающую три значения: 1, если антропогенная нагрузка на окружающую среду в городе/районе высокая, 2 – если средняя и 3 – низкая;

б)  $y_i^{risk}$  для характеристики степени эколого-экономического риска здоровью населения в  $i$ -м городе/районе, принимающую три значения: 1, если степень риска в городе/районе высокая, 2 – если средняя и 3 – если низкая.

Тогда осуществим моделирование интегральных показателей по указанным выше частным критериям в виде:

$$y_i = \begin{cases} 1, & y_i^* \leq \gamma_1, \\ \dots & \\ j, & \gamma_{j-1} < y_i^* \leq \gamma_j, \\ \dots & \\ k, & y_i^* > \gamma_{k-1} \end{cases} \quad (3.34)$$

где  $y_i = y_i^{nagr}$  или  $y_i = y_i^{risk}$  в зависимости от моделируемого интегрального показателя;

$$y_i^{nagr} = \begin{cases} 1, & \text{если антропогенная нагрузка на ОС в городе/районе } i \text{ высокая} \\ 2, & \text{если антропогенная нагрузка на ОС в городе/районе } i \text{ средняя;} \\ 3, & \text{если антропогенная нагрузка на ОС в городе/районе } i \text{ низкая} \end{cases}$$

$$y_i^{risk} = \begin{cases} 1, & \text{если уровень эколого - экон. риска в городе/районе } i \text{ высокий} \\ 2, & \text{если уровень эколого - экон. риска в городе/районе } i \text{ средний;} \\ 3, & \text{если уровень эколого - экон. риска в городе/районе } i \text{ низкий} \end{cases}$$

$\gamma_1 < \dots < \gamma_j < \dots < \gamma_{k-1}$  - пороговые значения (или границы разделения) – неизвестные и подлежащие оцениванию параметры;



$y^*$  - ненаблюдаемая вспомогательная переменная  $y_i^* = \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i$ ;

$x = (x_1, x_2, \dots, x_p)^T$  - набор переменных, используемых для построения соответствующего интегрального показателя,

$\varepsilon_i$  - независимо и одинаково распределенные случайные величины с функцией распределения  $F(u)$ , учитывающие влияние неучтенных факторов.

Поскольку распределение ошибок  $\varepsilon_i$  не удастся хорошо описать с помощью одного из распространенных законов распределения (нормального, логистического или др.), то можно использовать аппроксимацию  $F_M(u)$ , например, предложенную Стюартом, приводящую к так называемому «полу-непараметрическому» методу оценивания – от англ. «semi-nonparametric», или SNP и, соответственно, к расширенной SNP-модели [11, 12, 13].

Суть метода заключается в поиске оценок параметров модели путем максимизации функции  $F_M(u)$ :

$$F_M(u) = \frac{\int_{-\infty}^u \left( \sum_{m=0}^M \lambda_m x^m \right)^2 \varphi(x) dx}{\int_{-\infty}^{\infty} \left( \sum_{m=0}^M \lambda_m x^m \right)^2 \varphi(x) dx}, \quad (3.35)$$

где  $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ ,

$\lambda = (\lambda_0, \lambda_1, \dots, \lambda_M)$  - вектор дополнительных, подлежащих оцениванию, так называемых SNP-параметров (весовых коэффициентов),

$M$  – количество SNP-параметров, определяемое путем перебора моделей с различным  $M$  на основе использования статистики отношения правдоподобия.

При этом удобно в качестве  $\lambda_0$  брать 1, а в качестве оценки порогового значения  $\gamma_1$  в SNP-модели - его оценку, полученную по обычной пробит-модели. Выражение (3.35) при различных  $M$  задает семейство «полунепараметрических» рас-

пределений, которое может использоваться для аппроксимации широкого класса распределений, в том числе островершинных с тяжелыми хвостами.

Таблица 3.13 – Результаты оценивания порядковой модели, характеризующей антропогенную нагрузку

Переменная	Метод оценивания							
	макс. правдоподобия, ММП				«полу-непараметрический», SNP			
	Оценка	Ст. ош.	z-стат.	p	Оценка	Ст. ош.	z-стат.	p
Коэффициенты								
SB	-0,03391	0,00775	-4,38	0,00	-0,03657	0,00716	-5,11	0,00
NEST	-0,00033	0,00007	-4,74	0,00	-0,00038	0,00007	-5,42	0,00
ST	-0,02223	0,00451	-4,93	0,00	-0,03297	0,00370	-8,91	0,00
TOX	-0,00013	0,00002	-5,95	0,00	-0,00016	0,00004	-4,15	0,00
Пороговые значения								
$\gamma_1$	-2,63274	0,22044	-11,94	0,00	-2,63274	-	-	-
$\gamma_2$	-1,35970	0,15206	-8,94	0,00	-0,47459	0,12279	-3,71	0,00
SNP-коэффициенты								
1	-	-	-	-	1,25647	0,34147	3,68	0,00
2	-	-	-	-	-0,30065	0,65474	-4,59	0,00
3	-	-	-	-	-0,29746	0,05707	-5,21	0,00

Таблица 3.14 – Результаты оценивания порядковой модели, характеризующей эколого-экономический риск

Переменная (среднегод. концентр. в-ва)	Метод оценивания							
	макс. правдоподобия, ММП				«полу-непараметрический», SNP			
	Оценка	Ст. ош.	z-стат.	p	Оценка	Ст. ош.	z-стат.	p
Коэффициенты								
железо	-0,988	0,429	-2,30	0,02	-0,681	0,479	-1,42	0,16
мышьяк	-200,080	48,156	-4,15	0,00	-179,722	27,057	-6,64	0,00
нитраты	-0,029	0,011	-2,57	0,01	-0,017	0,009	-1,92	0,05
свинец	-2,377	2,001	-1,19	0,24	-1,903	3,049	-0,62	0,53
хром	-131,081	19,194	-6,83	0,00	-114,012	14,257	-8,00	0,00
медь	-2,420	0,937	-2,58	0,01	-1,699	0,931	-1,82	0,07
марганец	-5,644	2,559	-2,21	0,03	-3,750	3,914	-0,96	0,33
взвеш.в-ва	-16,152	2,474	-6,52	0,00	-15,251	0,842	-18,11	0,00
диоксид азота	-16,230	4,361	-3,72	0,00	-14,055	2,408	-5,84	0,00
сероводород	-447,308	102,105	-4,38	0,00	-389,725	83,003	-4,70	0,00
оксид углерода	-0,251	0,068	-3,69	0,00	-0,178	0,064	-2,76	0,01
формальдегид	-75,049	30,517	-2,46	0,01	-84,642	7,898	-10,72	0,00
аммиак	-6,851	1,092524	-6,27	0,00	-5,919	0,448	-12,99	0,00
Пороговые значения								
$\gamma_1$	-8,487	1,246	-6,811	0,00	-8,487	-	-	-
$\gamma_2$	-4,045	0,496	-8,163	0,00	-4,316	0,207	-20,85	0,00
SNP-коэффициенты								
1	-	-	-	-	-1,870	0,579	-3,23	0,00
2	-	-	-	-	0,091	0,189	0,48	0,63
3	-	-	-	-	0,245	0,078	3,16	0,00

В таблицах 3.13 и 3.14 приведены оценки параметров моделей интегральных показателей антропогенной нагрузки и эколого-экономического риска. Оценка мо-

делей методом максимального правдоподобия произведена в пакете EViews, полунепараметрическим методом - в пакете Stata. В качестве объясняющих переменных в модели антропогенной нагрузки использованы значения введенных выше переменных выбросов от стационарных и нестационарных источников, сбросов сточных вод и количества токсичных отходов на единицу территории; в модели эколого-экономического риска – среднегодовые концентрации загрязняющих веществ, содержащихся в атмосферном воздухе или водопроводной воде.

В таблице 3.15 приведены характеристики качества моделей. По величине индексов отношения правдоподобия Макфаддена делаем вывод об адекватности построенных моделей.

Таблица 3.15 – Характеристики качества моделей

LRI	Модель антропогенной нагрузки		Модель эколого-экономического риска здоровью населения	
	ММП	SNP	ММП	SNP
	0,44	0,51	0,75	0,80

Построенные модели позволили произвести окончательное отнесение городов и районов Оренбургской области к классам с той или иной степенью антропогенной нагрузки на окружающую среду и эколого-экономического риска [14,15].

Считая, что чем выше степень риска, тем меньше номер класса муниципального образования, можно утверждать, что все объекты 1 класса обладают худшим рейтингом по сравнению с любым объектом из класса 2 и тем более из класса 3. В рамках же каждого класса объекты считаются неразличимыми. Между тем, часто такой информации недостаточно, требуется в рамках каждого класса выделить объекты, наиболее приоритетные, например, для реализации управленческих решений по снижению риска, или те, положение которых неустойчиво и также требует внимания со стороны органов власти. Кроме того, как было показано ранее при анализе результатов многомерного шкалирования, выделенные классы административно-территориальных образований не являются в полном смысле однородными.

Для решения указанной задачи предлагается двухэтапная процедура ранжирования городов и районов области по степени эколого-экономического риска (или нагрузки на среду).

На **первом этапе** оценивается описанная выше модель упорядоченного множественного выбора, на основе которой каждый город/район относится к одному из  $k$  классов, каждой группе объектов присваивается групповой рейтинг 1, 2, ...  $k$  в зависимости от степени риска. В нашем случае  $k$  было принято равным 3.

При упорядочении объектов в рамках одного класса (рейтинга) для объектов с рейтингами 1 и  $k$  можно использовать ранжирование по убыванию/возрастанию вероятности принадлежности соответствующему классу, поскольку это будет соответствовать уменьшению степени проявления анализируемой латентной категории (в нашем случае риска).

Для объектов с рейтингами 2,...,  $k-1$  такой подход неправомерен. Рассмотрим объекты  $l$  и  $m$  с одинаковым рейтингом  $j$ . Обозначим векторы вероятностей отнесения объектов в каждый из классов через  $P_l = (P_l^1, \dots, P_l^{j-1}, P_l^j, P_l^{j+1}, \dots, P_l^k)$  и  $P_m = (P_m^1, \dots, P_m^{j-1}, P_m^j, P_m^{j+1}, \dots, P_m^k)$  соответственно. Пусть, например,  $P_l^{j-1} > P_m^{j-1}$ , а  $P_l^{j+1} < P_m^{j+1}$ . Это означает, что для объекта  $l$  можно говорить о возможности перехода в класс с худшим рейтингом  $j-1$ , а для объекта  $m$  существует реальная возможность улучшения своего состояния и перехода в класс с рейтингом  $j+1$ . Такие объекты, безусловно, следует различать. Это можно сделать введением класса-эталона, увеличение вероятности принадлежности к которому будет свидетельствовать об уменьшении риска. В ситуациях равных вероятностей отнесения объекта к классу-эталону берется второй эталон. Для каждого  $j$ -го класса естественными эталонами являются соседние  $j+1$  и  $j-1$  классы, хотя их выбор может определяться содержательной постановкой задачи.

Таким образом, на **втором этапе** предлагаемой процедуры ранжирования для каждого класса осуществляется выбор одного или двух классов-эталонов и производится ранжирование объектов по возрастанию/убыванию величины  $P_i^{et}$  - вероятности отнесения объекта  $i$  к классу-эталону. При необходимости может

быть составлен общий рейтинг, однако при его использовании следует учитывать качественное различие объектов с рейтингами  $R_{n_j}$  и  $R_{n_{j+1}}$ ,  $j = 1, 2, \dots, k-1$ , где  $n_j$  - количество объектов с рейтингом  $j$  (при отсутствии связанных рангов).

Схема процедуры ранжирования приведена на рисунке 3.1



Рисунок 3.1 - Схема двухэтапной процедуры ранжирования

Полученные с использованием указанной процедуры результаты ранжирования городов и районов Оренбургской области по степени эколого-экономического риска (за 2007 год) приведены в таблице 3.16, в качестве эталонов для 2-го класса были выбраны классы 3 и 1; для 1-го класса – классы 2 и 1, а для 3-го – классы 3 и 2 (вероятности отнесения объекта к эталонному классу выделены жирным шрифтом).

Таблица 3.16 – Результаты ранжирования городов и районов Оренбургской области по степени эколого-экономического риска (за 2007 год)

Общий рейтинг	Рейтинг внутри класса	Город/район	Вероятность отнесения к классу $\hat{P}\{y_i = j X_i\}$		
			$j = 1$	$j = 2$	$j = 3$
<b>1 класс – высокий эколого-экономический риск</b>					
4,5	4,5	Абдулинский	1,000	<b>0,000</b>	0,000
4,5	4,5	Бузулукский	1,000	<b>0,000</b>	0,000
4,5	4,5	Грачевский	1,000	<b>0,000</b>	0,000
4,5	4,5	Соль-Илецкий	1,000	<b>0,000</b>	0,000
4,5	4,5	г.Медногорск	1,000	<b>0,000</b>	0,000
4,5	4,5	г.Новотроицк	1,000	<b>0,000</b>	0,000
4,5	4,5	г.Оренбург	1,000	<b>0,000</b>	0,000
4,5	4,5	г.Орск	1,000	<b>0,000</b>	0,000
9	9	Пономаревский	0,983	<b>0,017</b>	0,000
10	10	Тоцкий	0,956	<b>0,044</b>	0,000
11	11	Бугурусланский	0,890	<b>0,107</b>	0,002
<b>2 класс – средний эколого-экономический риск</b>					
12	1	Сакмарский	0,313	0,673	<b>0,014</b>
13	2	Беляевский	0,282	0,704	<b>0,015</b>
14	3	Новосергиевский	0,178	0,805	<b>0,016</b>
15	4	Гайский	0,157	0,826	<b>0,017</b>
16	5	Матвеевский	0,016	0,966	<b>0,018</b>
17	6	Новоорский	0,011	0,969	<b>0,019</b>
18	7	Шарлыкский	0,002	0,967	<b>0,031</b>
19	8	Кувандыкский	0,002	0,964	<b>0,034</b>
20	9	Первомайский	0,001	0,955	<b>0,044</b>
21	10	Оренбургский	0,001	0,931	<b>0,068</b>
22	11	Курманаевский	0,000	0,901	<b>0,098</b>
23	12	Светлинский	0,000	0,843	<b>0,157</b>
24	13	Сорочинский	0,000	0,810	<b>0,190</b>
25	14	Ясненский	0,000	0,779	<b>0,221</b>
<b>3 класс – низкий эколого-экономический риск</b>					
26	1	Домбаровский	0,000	<b>0,426</b>	0,574
27	2	Красногвардейский	0,000	<b>0,408</b>	0,592
28	3	Кваркенский	0,000	<b>0,335</b>	0,665
29	4	Октябрьский	0,000	<b>0,325</b>	0,675
30	5	Саракташский	0,000	<b>0,250</b>	0,750
31	6	Акбулакский	0,000	<b>0,176</b>	0,824
32	7	Адамовский	0,000	<b>0,070</b>	0,930
33,5	8,5	Северный	0,000	<b>0,050</b>	0,950
33,5	8,5	Переволоцкий	0,000	<b>0,050</b>	0,950
35	10	Асекеевский	0,000	<b>0,012</b>	0,988
36	11	Александровский	0,000	<b>0,003</b>	0,997
37,5	12,5	Тюльганский	0,000	<b>0,002</b>	0,998
37,5	12,5	Ташлинский	0,000	<b>0,002</b>	0,998
39	14	Илекский	0,000	<b>0,001</b>	0,999

Таблица 3.17 содержит сводные результаты ранжирования городов и районов Оренбургской области по уровню эколого-экономического риска за весь исследуемый период.

Таблица 3.17 – Ранжирование городов и районов Оренбургской области по уровню эколого-экономического риска за период 2000-2007 гг.

Город/район	Ранг							
	2000	2001	2002	2003	2004	2005	2006	2007
Абдулинский	14	16	17	18	17	12	5,5	4,5
Адамовский	27,5	18	35,5	35,5	37	31	29	32
Акбулакский	29	33	31	31,5	32	35	32	31
Александровский	36,5	37	35,5	35,5	37	37,5	37,5	36
Асекеевский	19	37	35,5	35,5	30	36	34	35
Беляевский	15	15	18	20	19	18	4	13
Бугурусланский	2	2	3	4	15	23	19	11
Бузулукский	10	8	7	14	11	3	2	4,5
Гайский	30	27	35,5	35,5	37	21	22	15
Грачевский	12	9	16	15	14	15	20	4,5
Домбаровский	36,5	25	21	24	31	29	28	26
Илекский	39	39	39	39	39	39	39	39
Кваркенский	27,5	34	29	28	25	27	15	28
Красногвардейский	33	26	26	27	20	33	33	27
Кувандыкский	6	3	10	2	10	11	14	19
Курманаевский	33	22	27	29	21	24	23	22
Матвеевский	25	35	35,5	30	24	28	26	16
Новоорский	9	14	8	6,5	4	13	16	17
Новосергиевский	16	20	14	13	12	16	11	14
Октябрьский	36,5	31	31	31,5	29	26	24	29
Оренбургский	21	23	22	17	9	14	12	21
Первомайский	24	17	19	6,5	26	25	35	20
Переволоцкий	11	10	15	19	22	34	25	33,5
Пономаревский	5	5,5	4	9	7	6,5	10	9
Сакмарский	7	12	12	10	3	9	13	12
Саракташский	20	13	11	35,5	37	37,5	36	30
Светлинский	36,5	37	35,5	35,5	37	32	31	23
Северный	31	32	31	6,5	33	20	30	33,5
Соль-Илецкий	3	5,5	2	2	1,5	1,5	2	4,5
Сорочинский	33	24	20	22	18	18	18	24
Ташлинский	26	29	28	25	23	6,5	37,5	37,5
Тоцкий	18	21	23	23	6	10	9	10
Тюльганский	22	28	13	16	16	18	17	37,5
Шарлыкский	23	30	25	26	27	30	27	18
Ясненский	17	19	24	21	28	22	21	25
г.Медногорск	1	1	1	2	1,5	1,5	8	4,5
г.Новотроицк	8	11	9	11	13	5	5,5	4,5
г.Оренбург	13	4	6	12	8	8	7	4,5
г.Орск	4	7	5	6,5	5	4	2	4,5

Из таблицы 3.17 видно, что за исследуемый период ранг большинства административно-территориальных образований изменялся, поэтому целесообразно провести анализ динамики их рангов. Известный подход к анализу динамики, предложенный Айвазяном в работах, посвященных качеству жизни, не позволяет судить о наличии положительных или отрицательных изменений в состоянии самого объекта, поскольку ориентирован на исследование динамики положения объекта в ряду остальных объектов, то есть на величину  $\delta_{i,t} = R_{i,t} - R_{i,t-1}$ , где  $R_{i,t}$  - ранг  $i$ -го объекта в момент времени  $t$ . Положительные значения  $\delta_{i,t}$  должны свидетельствовать о положительной межтерриториальной динамике объекта  $i$ . Он хорошо работает в случае улучшения ситуации по анализируемому интегральному свойству – действительно, отрицательная величина  $\delta_{i,t}$  свидетельствует если не об ухудшении состояния объекта, то, по крайней мере, о более медленном по сравнению с остальными объектами, улучшении этого состояния, обусловленным, возможно, недоиспользованием своих возможностей. В нашем случае в исследуемом периоде наблюдается рост показателей, характеризующих риски, имеет место общее ухудшение ситуации, и говорить о повышении рейтинга  $i$ -го объекта при положительном значении  $\delta_{i,t}$ , наблюдающемся одновременно со снижением значения  $X_i \hat{\beta}$ , на наш взгляд, неправомерно.

Сначала проанализируем изменение рангов каждого административно-территориального образования. Построив ранги во времени, мы не получаем, однако, никакого представления об ухудшении или улучшении ситуации для каждого конкретного объекта. В определенной степени о положительной или отрицательной динамике в поведении  $i$ -го объекта можно судить, определив его ранг в момент времени  $t$  по показателям, характеризующим этот объект в момент времени  $t + j$ , причем в нашем случае  $j=1, \dots, 7$ . Результаты такого ранжирования приведены в таблице 3.18.



Таблица 3.18 – Ранжирование каждого города/района Оренбургской области по уровню эколого-экономического риска за период 2000-2007 гг.

Город/район	Ранг							
	2000	2001	2002	2003	2004	2005	2006	2007
Абдулинский	14	14	16	16	15	6	3	1,5
Адамовский	27,5	15	36	36	36	24	23	21
Акбулакский	29	27	32,5	32,5	30,5	29	25	19
Александровский	36,5	36,5	36,5	36,5	36,5	36,5	36,5	33,5
Асекеевский	19	36	36	36	27	29,5	25	25
Беляевский	15	13	16	16	16	11	3	6
Бугурусланский	2	2	5	2	3	16	13	5
Бузулукский	10	6	8	9	9	3	1,5	1,5
Гайский	30	20	36	36	36	15	17	7
Грачевский	12	6	14	11	13	9	14	1,5
Домбаровский	36,5	17	19	21	30,5	22	221	17
Илекский	39	39	39	39	39	39	39	39
Кваркенский	27,5	27,5	29,5	28	24	22	13	18
Красногвардейский	33	19	25	24	19	28	26,5	17
Кувандыкский	6	3	9	1,5	7	6	11	11
Курманаевский	33	16	26,5	28,5	21	18	18	14
Матвеевский	25	33	36	29,5	23	23	20	14
Новоорский	9	10,5	9	4	6	7	13	8
Новосергиевский	16	16	14	8,5	12	12	6	7
Октябрьский	36,5	26	34	34	28	221	19	19
Оренбургский	21	17	21	16	6	8	8	14
Первомайский	24	14	20	5	24	20	27	12
Переволоцкий	11	7	13	16	20	27	19	23
Пономаревский	4,5	4,5	5	5	5	3	5	3
Сакмарский	7	7,5	12	6	6	6	7	6
Саракташский	20	1	13	36	36	36	30,5	20
Светлинский	36,5	36,5	36,5	36,5	36,5	26	25	14
Северный	31	28	33	5	33	14	24	24
Соль-Илецкий	3	3,5	3	1,5	1,5	1,5	1,5	1,5
Сорочинский	33	17	20	20	17	13	14	14
Ташлинский	26	22	27	24	21	3	36	36
Тоцкий	18	16	23	21	6	6	6	5
Тюльганский	22	21	15	14	34	14	15	36
Шарлыкский	23	26	24	31,5	24	23,5	22	28
Ясенский	17	16	24	19	25	16	18	16
г.Медногорск	1	1	1	1	1	1	1	1
г.Новотроицк	8	9	9	8	12	3	3	1,5
г.Оренбург	13	3	9	9	7	6	3	1,5
г.Орск	4,5	7	5	6,5	5	4	2	4,5

Обозначим полученные ранги  $rang_{i,t}$ ,  $t=1,2,\dots,T$ ;  $T=8$  - временной ряд рангов  $i$ -го объекта,  $i=1,2,\dots,n$ ,  $n=39$ . В этом случае проверка гипотезы о несущественности ухудшения/улучшения ситуации в  $i$ -ом городе/районе сводится к

проверке гипотезы об отсутствии в ряду  $rang_i$  тренда. Одним из непараметрических критериев проверки такой гипотезы является критерий, основанный на проверке значимости коэффициента ранговой корреляции Спирмена между значениями исследуемого временного ряда и значениями последовательности  $t = 1, 2, \dots, T$ . Для тех городов/районов области, в которых происходило ухудшение ситуации (уменьшение значений  $rang_i$ ), коэффициент ранговой корреляции будет значим, близок к минус единице; в которых имело место улучшение ситуации – значим и близок к единице; случаи стабильной обстановки по риску или наличия в течение исследуемого периода существенных колебаний, не позволяющие сделать вывод об улучшении/ухудшении ситуации, дадут незначимое небольшое по абсолютной величине положительное или отрицательное значение коэффициента. Использование такой процедуры позволит выяснить тенденцию изменения уровня риска для каждого исследуемого объекта, что представляется более обоснованным, нежели выводы по анализу изменений за отдельные годы. Для объектов, изменения рангов которых отсутствуют, коэффициент ранговой корреляции Спирмена принят равным нулю. Из-за малой длины временных рядов при проверке значимости коэффициентов имело смысл использовать более высокий, чем обычно, уровень значимости. Таким образом, выявлены города/районы, сходные по тенденции изменения уровня эколого-экономического риска (таблица 3.19).

Для того чтобы учесть также «величину» эколого-экономического риска, будем рассматривать показатель  $rang_i$  совместно со значением линейной комбинации  $X_i \hat{\beta}$ . Это объясняется тем, что если значение  $X_i \hat{\beta}$  больше  $\hat{\gamma}_1$ , то согласно модели (3.30) интегральный показатель указывает на принадлежность объекта к классу низкого риска, если меньше  $\hat{\gamma}_1$  и больше  $\hat{\gamma}_2$ , то к классу среднего риска и т.д. Построим график взаимного расположения исследуемых городов и районов на плоскости  $rang_i - X_i \hat{\beta}$ , где  $X_i \hat{\beta}$  относится к последнему году исследуемого периода (рисунок 3.2).

Таблица 3.19 – Города и районы Оренбургской области, ранжированные по степени проявления тенденции роста интегрального показателя эколого-экономического риска

Название города/района	Оценка коэфф. Спирмена	р-знач.	Название города/района	Оценка коэфф. Спирмена	р-знач.
<b>Существенный рост уровня эколого-экономического риска</b>					
Новосергиевский	-0,89	0,00	г.Оренбург	-0,70	0,05
Светлинский	-0,87	0,00	г.Новотроицк	-0,64	0,09
Октябрьский	-0,84	0,01	Абдулинский	-0,63	0,10
Матвеевский	-0,83	0,01	Курманаевский	-0,61	0,11
Соль-Илецкий	-0,82	0,01	Беляевский	-0,59	0,13
Сорочинский	-0,81	0,01	Гайский	-0,59	0,13
Тоцкий	-0,78	0,02	Александровский	-0,58	0,13
Кваркенский	-0,78	0,02	Сакмарский	-0,57	0,14
Бузулукский	-0,77	0,03	г.Орск	-0,57	0,14
Оренбургский	-0,73	0,04	Акбулакский	-0,48	0,23
<b>Несущественное изменение уровня эколого-экономического риска</b>					
Северный	-0,42	0,30	Новоорский	-0,14	0,73
Красногвардейский	-0,32	0,43	Первомайский	-0,01	0,98
Асекеевский	-0,26	0,54	Илекский	0,00	1,00
Адамовский	-0,24	0,56	г.Медногорск	0,00	1,00
Домбаровский	-0,23	0,59	Шарлыкский	0,02	0,96
Пономаревский	-0,18	0,67	Тюльганский	0,07	0,86
Ясненский	-0,15	0,73	Ташлинский	0,28	0,51
Грачевский	-0,14	0,73	Новоорский	-0,14	0,73
<b>Уменьшение уровня эколого-экономического риска</b>					
Саракташский	0,43	0,29	Бугурусланский	0,68	0,06
Кувандыкский	0,63	0,10	Переволоцкий	0,86	0,01

Административно-территориальными образованиями, население которых стабильно подвергается высокому эколого-экономическому риску, являются объекты VII условного квадранта – город Медногорск и районы Пономаревский и Грачевский. Для объектов VI условного квадранта характерно наличие высокого уровня эколого-экономического риска и, более того, значимая тенденция к его росту. Это основные промышленные и торговые центры Оренбургской области: города Новотроицк, Орск и Оренбург, а также Абдулинский, Бузулукский и Соль-Илецкий районы. Обращает на себя внимание почти пустота IX квадранта, в котором должны были находиться города и районы, характеризующиеся высоким эколого-экономическим риском, для которых наметилась тенденция к снижению уровня риска.

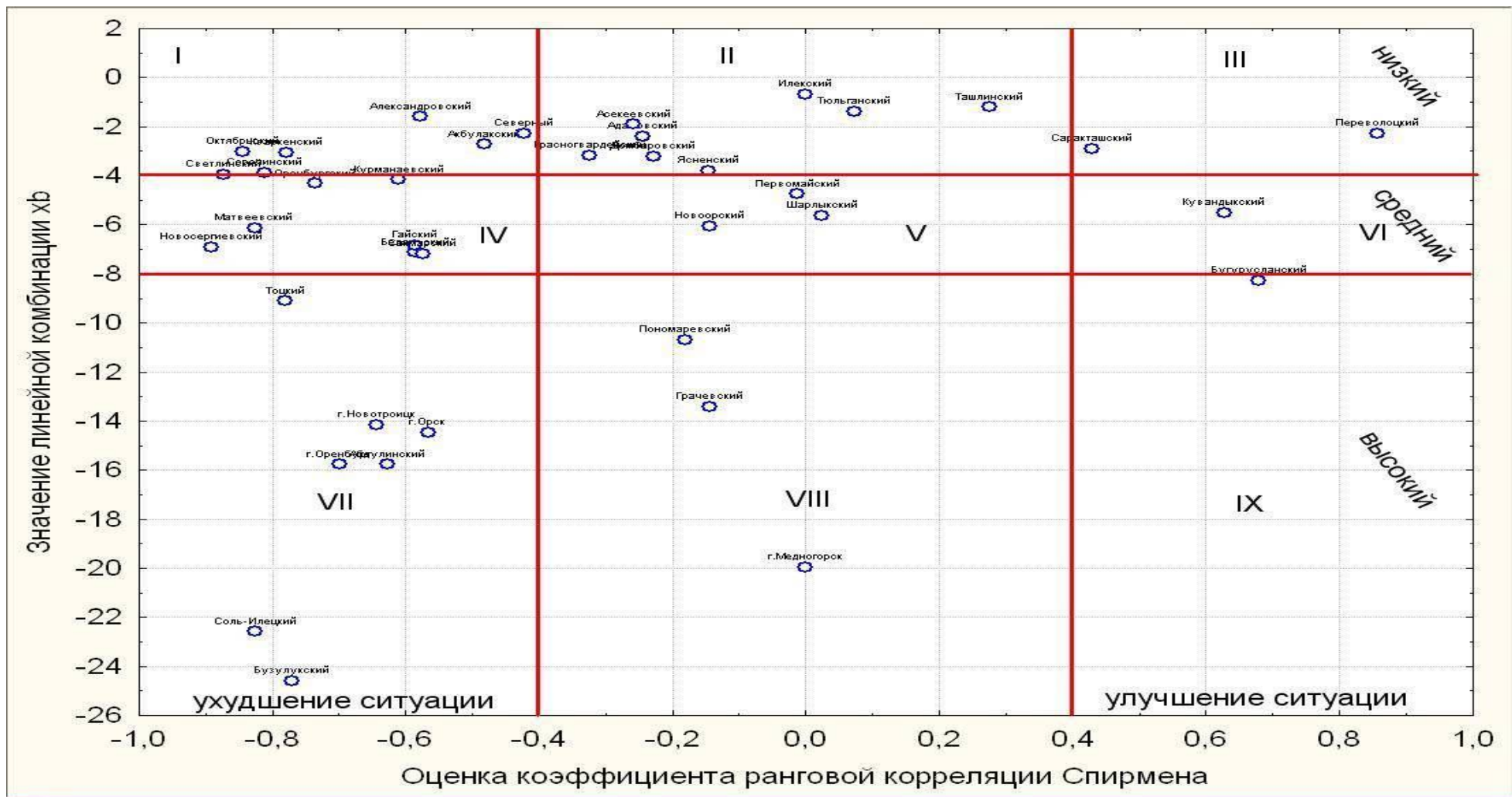


Рисунок 3.2 – Расположение городов и районов Оренбургской области на плоскости  $rang_i - X_i \hat{\beta}$

Большая часть городов и районов области находится в I, IV и VII условных квадрантах, для которых характерно наличие значимой тенденции к росту риска, даже если для некоторых городов и районов такие изменения пока не привели к переходу в другой класс. Это означает, что ухудшение экологической ситуации происходит во всех классах, в том числе и классах низкого эколого-экономического риска.

### **3.8 Моделирование взаимного влияния эколого-экономического риска и показателей социально-экономического развития муниципалитетов на основе двумерных пробит-моделей**

Рассмотрим пример использования систем моделей бинарного выбора для оценки взаимного влияния эколого-экономического риска, характерного для муниципалитета региона, и показателей его социально-экономического развития. Необходимость привлечения аппарата моделей бинарного выбора объясняется номинальной шкалой, в которой доступнее всего получить как оценку эколого-экономического риска, так и оценку уровня социально-экономического развития конкретного муниципалитета. Необходимость объединения моделей бинарного выбора в систему объясняется тем, что не только уровень социально-экономического развития муниципалитета позволяет прогнозировать эколого-экономический риск, но, несомненно, может выступать одним из факторов замедления или ускорения развития муниципальных образований региона. Это связано, например, с тем, что наличие высокого эколого-экономического риска означает для предприятий появление дополнительных расходов по его снижению и управлению им, использование ресурсов (трудовых, природных) более низкого качества и т.д. В долгосрочной перспективе высокий уровень эколого-экономического риска может стать барьером для привлечения инвестиций и реализации кредитуемых проектов.

В частности, эколого-экономический риск связан с ущербом для населения, который может проявляться в росте его заболеваемости. Моделирование заболе-

ваемости населения позволяет характеризовать динамику экономических потерь как на уровне отдельных предприятий, домохозяйств, так и на региональном уровне.

Перейдем к построению модели в виде системы одновременных уравнений, описывающей взаимосвязи эколого-экономического риска и заболеваемости населения от показателей социально-экономического состояния муниципалитетов.

Информационной базой для построения модели будут выступать значения социально-экономических показателей, характеризующие города и районы Оренбургской области за период 2000-2007 гг., полученные из официальных статистических сборников Территориального органа Федеральной службы государственной статистики по Оренбургской области и г. Оренбургу и Министерства здравоохранения по Оренбургской области, а также информация о содержании химических веществ в атмосферном воздухе и питьевой воде, полученная от ФГУЗ «Центр гигиены и эпидемиологии по Оренбургской области», на основе которых оценивался уровень эколого-экономического риска.

Введем следующие обозначения:

RN – бинарная переменная, принимающая значение 0, если город или район характеризуется низким эколого-экономическим риском, значение 1 – высоким или средним риском;

UOZ – бинарная переменная, принимающая значение 0, если уровень заболеваемости в городе или районе ниже среднеобластного, значение 1 – уровень заболеваемости выше среднеобластного; назовем ее показателем превышения среднеобластной заболеваемости.

Высокий уровень эколого-экономического риска может быть ответной реакцией среды на увеличение антропогенной нагрузки (IPER). Также в качестве факторов, предположительно влияющих на повышение эколого-экономического риска, нами были выбраны объем платных услуг населению (OPUN, на душу населения, тыс. руб.) и коэффициент миграционного прироста (MIGR, на 1 тыс. чел).

В качестве факторов, влияющих на превышение среднеобластной заболеваемости, исходя из содержательных соображений, нами были выбраны эколого-экономический риск (RN, описанная выше бинарная переменная), задолженность предприятий по заработной плате (ZZP, на 1 работ., тыс. руб.), объем инвестиций в основной капитал (IOK, на душу насел., тыс. руб.), объем инвестиций в жилищно-коммунальное строительство (IZKS, на душу насел., тыс. руб.).

Естественно предположить, что наличие в  $i$ -м городе/районе эколого-экономических рисков для населения и превышение в этом городе/районе среднеобластной заболеваемости, являются зависимыми между собой событиями. Так, существование эколого-экономического риска для населения может способствовать более высокому (по сравнению со средним) уровню заболеваемости населения, а превышение среднеобластного уровня заболеваемости может чаще наблюдаться в районах с повышенным уровнем эколого-экономического риска. Поэтому моделирование вероятности повышенного эколого-экономического риска и вероятности повышенной заболеваемости необходимо осуществлять совместно:

$$\begin{cases} P\{RN_i = 1 | X_{i1}\} = \Phi\{\beta_{10} + \beta_{11} \cdot IPER_i + \beta_{12} \cdot OPUN_i + \beta_{13} \cdot MIGR_i\} \\ P\{UOZ_i = 1 | X_{i2}\} = \Phi\{\beta_{20} + \beta_{21} \cdot RN_i + \beta_{22} \cdot IOK_i + \beta_{23} \cdot ZZP_i + \beta_{24} \cdot IZKS_i\} \end{cases}$$

где  $X_{i1} = (IPER_i, OPUN_i, MIGR_i)$ ,  $X_{i2} = (RN_i, IOK_i, ZZP_i, IZKS_i)$ ,

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt.$$

Для оценки параметров  $\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{20}, \beta_{21}, \beta_{22}, \beta_{23}, \beta_{24}$  построим систему одновременных пробит-моделей, учитывающую существование ненулевой ковариационной связи между соответствующими ошибками:

$$\begin{cases} RN_i = 1, y_{i,RN}^* > 0 \\ RN_i = 0, y_{i,RN}^* \leq 0 \end{cases}, \begin{cases} UOZ_i = 1, y_{i,UOZ}^* > 0 \\ UOZ_i = 0, y_{i,UOZ}^* \leq 0 \end{cases}, \quad (3.36)$$

где  $y_{i,RN}^*$  и  $y_{i,UOZ}^*$  – латентные переменные, причем

$$\begin{cases} y_{i,RN}^* = \beta_{10} + \beta_{11} \cdot IPER_i + \beta_{12} \cdot OPUN_i + \beta_{13} \cdot MIGR_i + \varepsilon_{i,RN} \\ y_{i,UOZ}^* = \beta_{20} + \beta_{21} \cdot RN_i + \beta_{22} \cdot IOK_i + \beta_{23} \cdot ZZP_i + \beta_{24} \cdot IZKS_i + \varepsilon_{i,UOZ} \end{cases}$$

$$\begin{pmatrix} \varepsilon_{i,RN} \\ \varepsilon_{i,UOZ} \end{pmatrix} \sim N(\mu, \Sigma), \mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \rho - \text{ковариация ошибок } \varepsilon_{i,RN} \text{ и } \varepsilon_{i,UOZ},$$

$$X_{i1} = (IPER_i, OPUN_i, MIGR_i),$$

$$X_{i2} = (RN_i, IOK_i, ZZP_i, IZKS_i).$$

Возможны следующие комбинации событий, образующие массив зависимых переменных модели: 1) отсутствие эколого-экономического риска и отсутствие повышенной заболеваемости (RN=UOZ=0); 2) наличие эколого-экономического риска и отсутствие повышенной заболеваемости (RN=1,UOZ=0); 3) отсутствие эколого-экономического риска и наличие повышенной заболеваемости (RN=0, UOZ=1); 4) наличие эколого-экономического риска и наличие повышенной заболеваемости (RN=UOZ=1).

Рекурсивная модель (3.36) идентифицируема. Состоятельные и асимптотически эффективные оценки получены в Stata (таблица 3.20).

Таблица 3.20 – Результаты оценивания двумерной модели, описывающей эколого-экономические риски и превышение среднеобластной заболеваемости населения

Показатель	Оценка коэфф.	Стандарт. отклон	z-стат	p	Доверительный интервал	
Уравнение для эколого-экономического риска, $\hat{R}_{pseudo}^2 = 0,58$						
IPER	-0,027	0,010	-2,67	0,008	-0,047	-0,007
OPUN	0,322	0,077	4,18	0,000	0,171	0,473
MIGR	0,023	0,011	2,09	0,037	0,001	0,045
cons	-0,656	0,141	-4,65	0,000	-0,933	-0,379
Уравнение для показателя превышения заболеваемости, $\hat{R}_{pseudo}^2 = 0,71$						
RN	1,028	0,306	3,36	0,001	0,429	1,627
IOK	0,037	0,016	2,38	0,018	0,007	0,068
ZZP	0,126	0,070	1,80	0,072	-0,011	0,263
IZKS	-0,255	0,090	-2,84	0,004	-0,431	-0,079
cons	-0,513	0,201	-2,55	0,011	-0,907	-0,119
Wald chi2(7) = 71,42			Prob > chi2 = 0,0000			
chi2(1) = 7,0329			Prob > chi2 = 0,0080			



Таким образом, снижение антропогенной нагрузки на окружающую среду действительно может привести к снижению эколого-экономических рисков для населения. Высокие эколого-экономические риски наблюдаются в городах и районах с высоким уровнем миграции и большим объемом платных услуг, оказываемых населению. В целом, на наличие эколого-экономического риска оказывает значительное влияние состояние окружающей среды и показатели, отражающие привлекательность муниципалитета. Превышение среднеобластной заболеваемости действительно связано с наличием высоких эколого-экономических рисков, увеличением объема инвестиций в основной капитал и задолженностью по заработной плате перед населением. Это можно объяснить тем, что инвестиции в основной капитал преимущественно направляются на расширение производственных возможностей предприятий, а не на технические усовершенствование оборудования и ввод очистных сооружений для снижения негативного воздействия на окружающую среду [16].

В таблице 3.21 представлены оценки условных и безусловных вероятностей событий, связанных с наличием/отсутствием эколого-экономического риска и наличием/отсутствием повышенной заболеваемости в административно-территориальном образовании.

Таблица 3.21 – Характеристики двумерной пробит-модели, описывающей взаимосвязь эколого-экономического риска и показателя превышения среднеобластной заболеваемости

Характеристика	Годы								Среднее значение
	2000	2001	2002	2003	2004	2005	2006	2007	
P (RN=1,UOZ=1)	0,14	0,14	0,16	0,21	0,20	0,27	0,27	0,30	0,21
P (RN=1,UOZ=0)	0,29	0,29	0,30	0,27	0,31	0,29	0,32	0,37	0,31
P (RN=0,UOZ=1)	0,36	0,36	0,35	0,34	0,33	0,32	0,27	0,22	0,32
P (RN=0,UOZ=0)	0,21	0,21	0,19	0,18	0,16	0,12	0,13	0,11	0,16
P (RN=1)	0,42	0,42	0,46	0,48	0,51	0,56	0,60	0,67	0,52
P (UOZ=1)	0,50	0,50	0,51	0,55	0,53	0,59	0,55	0,52	0,53
P (RN=1 UOZ=1)	0,22	0,22	0,26	0,30	0,32	0,40	0,41	0,48	0,33
P (UOZ=1 RN=1)	0,27	0,27	0,29	0,33	0,33	0,41	0,38	0,38	0,33

Из таблицы видно, что среднее за период с 2000 по 2007 гг. значение вероятности того, что города и районы области одновременно характеризуются высоким уровнем эколого-экономического риска и превышением среднеобластной заболеваемости, составляет 0,21. За рассматриваемый промежуток времени эта величина увеличилась более чем в два раза. Анализируя величину вероятности этого события в разрезе городов и районов области, выявлено, что наибольшие вероятности возникновения такой ситуации наблюдаются в Бузулукском, Кувандыкском, Оренбургском и Новоорском районах, городах Медногорск, Новотроицк и Орск. Наиболее проблемными за последние годы стали Адамовский, Гайский, Светлинский, Октябрьский, Соль-Илецкий, Сорочинский, Тоцкий и Ясенский районы.

К концу исследуемого периода наиболее вероятной была ситуация, связанная с наличием значительного эколого-экономического риска и отсутствием превышения средней заболеваемости населения. Особенно она была характерна для Бугурусланского, Саракташского, Северного, Ташлинского, Тюльганского районов и города Оренбурга. Учитывая, что предельная вероятность превышения среднеобластной заболеваемости для этих административно-территориальных образований значительно снижается с течением времени, а предельная вероятность высокого эколого-экономического риска возрастает, можно с одной стороны предположить повышение в рассмотренных районах качества предоставляемых медицинских услуг и условий быта.

Средняя вероятность превышения среднеобластной заболеваемости в сочетании с низким уровнем эколого-экономического риска равна 0,32, причем в последние два года отмечается ее снижение почти на треть. При низких эколого-экономических рисках для населения различия в уровнях заболеваемости для городов и районов области могут возникать из-за различных климатических, ландшафтных и других особенностей рассматриваемых территорий.

Средняя вероятность того, что в административно-территориальных образованиях области отсутствуют эколого-экономические риски, а заболеваемость населения не превышает среднюю по области, составляет 0,16. По отношению к

2000 году вероятность возникновения такой ситуации в 2007 году существенно уменьшилась, что свидетельствует об ухудшении общей ситуации в регионе.

Если 2000 году вероятность того, что заболеваемость в *i*-ом городе/районе превысит среднюю по области в условиях высокого эколого-экономического риска составляла 0,22, то в 2007 году - уже 0,48. Это означает, что практически в каждом втором случае высокий эколого-экономический риск стал «реализовываться» в виде повышенной заболеваемости населения (против каждого пятого случая в 2000 году). Таким образом, за исследуемый период фактор эколого-экономического риска начал играть все возрастающую роль в повышении заболеваемости населения. Анализ модели также показал, что вероятность повышенной заболеваемости в некотором городе/районе при наличии эколого-экономического риска увеличивается на 0,39 по сравнению с ситуацией, когда эколого-экономического риска нет.

Построенная модель позволяет также разделить все административно-территориальных образований исследуемого региона на 4 группы (рисунок 3.3 и таблица 3.22).

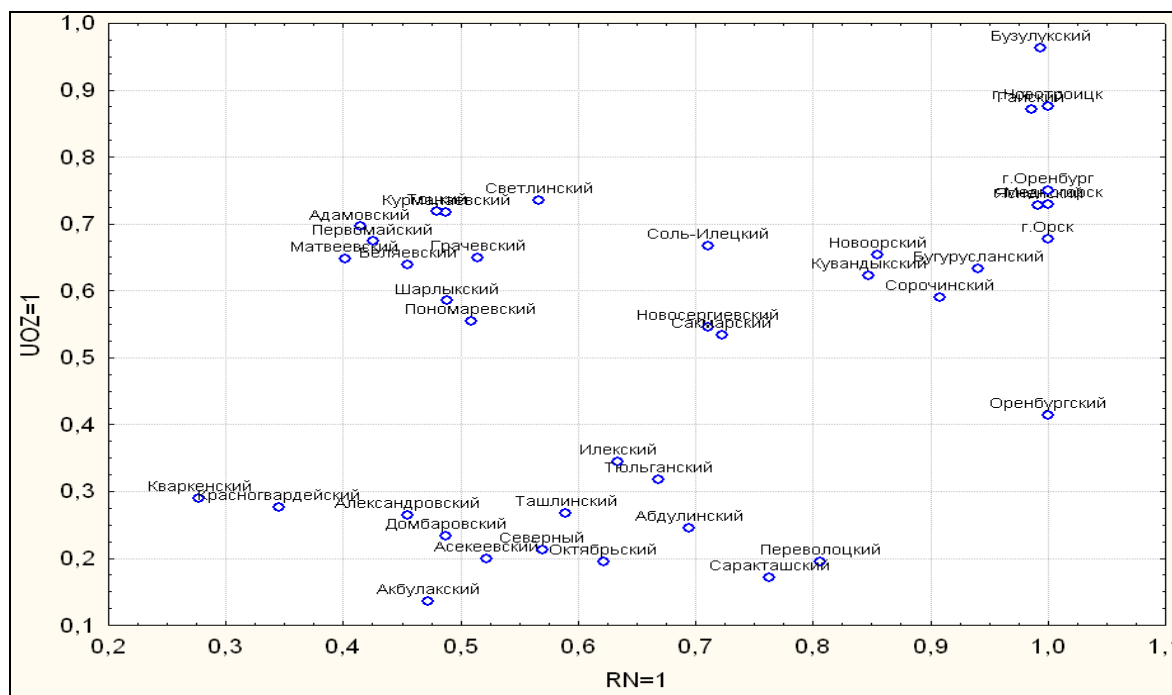


Рисунок 3.3 – График расположения городов и районов Оренбургской области в пространстве предельных вероятностей событий RN=1, UOZ=1 (2007 г.)

Таблица 3.22 – Характеристики групп административно-территориальных образований

<p>IV - повышенная заболеваемость без риска C1=9,371; C2=10,043</p>	<p>I- повышенная заболеваемость и риск C1=9,497; C2=10,385</p>
<p>III-невысокая заболеваемость без риска C1=8,867 ; C2=9,800</p>	<p>II - невысокая заболеваемость и риск C1=9,189 ; C2=10,344</p>

Согласно Программе государственных гарантий оказания гражданам Российской Федерации бесплатной медицинской помощи в 2007 году норматив объема амбулаторно-поликлинической помощи составляет 9,198 посещение. В выделенной на основе модели группе I городов и районов, для которых характерен повышенный по сравнению со среднеобластным уровень заболеваемости населения и высокий эколого-экономический риск, среднее число посещений (C1, на одного жителя в год) максимально превышает норматив. Также наибольшей является средняя длительность пребывания в круглосуточном стационаре (C2, дней) для населения муниципалитетов, где наблюдался высокий уровень эколого-экономического риска (группы I и II).

С учетом того, что согласно Программе норматив финансовых затрат на одно посещение амбулаторно-поликлинической организации составлял 116,9 рубля, а один койко-день в круглосуточном стационаре - 674,3 рубля, получаем:

1) для муниципалитетов I группы дополнительные расходы на обслуживание амбулаторно-поликлиническое обслуживание населения составляют 35 рублей на одного жителя в год, то есть дополнительно почти 30% от норматива;

2) средняя длительность пребывания в круглосуточном стационаре 1 пациента муниципалитета группы I на 0,585 дня дольше, чем пациента муниципалитета группы III; в стоимостном выражении это означает необходимость дополнительных расходов в размере 394,7 рубля на 1 госпитализированного или с учетом среднего уровня госпитализации на 100 человек населения - 7845,9 рубля;

3) средняя длительность пребывания в круглосуточном стационаре 1 пациентов муниципалитетов групп I и II, где присутствует высокий риск, на

0,443 дня дольше, чем пациентов муниципалитетов групп III и IV, где риска нет; в стоимостном выражении это означает необходимость дополнительных расходов в размере 299 рубля на 1 госпитализированного или с учетом среднего уровня госпитализации на 100 человек населения - 6070 рублей.

Учитывая выявленную по модели отрицательную динамику ситуации имеет смысл рассчитать дополнительные расходы, взяв нормативы финансовых затрат 2009 года: 218,1 рубля за одно посещение амбулаторно-поликлинической организации и 1380,6 рубля за один койко-день в круглосуточном стационаре. Получаем, что:

1) для муниципалитетов I группы дополнительные расходы на обслуживание населения составляют 65,2 рубля на одного жителя в год;

2) для муниципалитетов с повышенной заболеваемостью и высоким уровнем эколого-экономического риска в стоимостном выражении с учетом среднего уровня госпитализации дополнительные расходы составят 16064,7 рубля на 100 человек населения по сравнению с муниципалитетами, где нет риска, а заболеваемость ниже средней по области;

3) для муниципалитетов с высоким уровнем эколого-экономического риска в стоимостном выражении с учетом среднего уровня госпитализации дополнительные расходы составят 12428 рубля на 100 человек населения по сравнению с муниципалитетами, где уровень эколого-экономического риска низок.

Как отмечалось нами выше, в настоящее время не вызывает сомнения взаимное влияние уровня социально-экономического развития территорий региона и наличия эколого-экономического риска. Несмотря на то, что в долгосрочной перспективе экологический фактор признается важным фактором как при реализации стратегии экономического роста, так и при реализации стратегии устойчивого развития, в краткосрочной перспективе часто выбирается тактика «сначала экономическое развитие, а решение экологических проблем потом». Так, высокий уровень социально-экономического развития территорий региона часто сопровождается повышением эколого-экономических рисков для населения, а вы-

сокие эколого-экономические риски пока чаще наблюдаются в городах и районах с высоким уровнем социально-экономического развития.

Для моделирования взаимосвязи уровня социально-экономического развития и эколого-экономических рисков введем в рассмотрение переменную CER, характеризующая уровень социально-экономического развития города или района и принимающую значение 0, если город/район имеет низкий уровень развития, значение 1 – высокий уровень развития. Для выделения однородных по уровню социально-экономического развития групп городов и районов области была проведена их многомерная классификация по таким показателям, как оборот розничной торговли, инвестиции в основной капитал, инвестиции в жилищно-коммунальное строительство, удельный вес убыточных предприятий и задолженность организаций по заработной плате за 2000–2007 гг; в качестве расстояния между объектами использовалось обычное евклидово расстояние. Тогда для описания взаимосвязи вероятности повышенного эколого-экономического риска и вероятности высокого уровня социально-экономического развития региона предлагается система:

$$\begin{cases} P\{RN_i = 1 | X_{i1}\} = \Phi\{\beta_{11} \cdot IPER_i + \beta_{12} \cdot LOGZP_i + \beta_{13} \cdot MIGR_i + \beta_{14} \cdot CER_{t-1i}\} \\ P\{CER_i = 1 | X_{i2}\} = \Phi\{\beta_{20} + \beta_{21} \cdot RN_i + \beta_{22} \cdot SOZ_i + \beta_{23} \cdot VZD_i + \beta_{24} \cdot KB_i\} \end{cases}$$

где  $X_{i1} = (IPER_i, LOGZP_i, MIGR_i, CER_{t-1i})$ ,  $X_{i2} = (RN_i, SOZ_i, VZD_i, KB_i)$ ,

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt,$$

оценка параметров которой производится на основе двумерной пробит-модели вида:

$$\begin{cases} RN_i = 1, y_{i,RN}^* > 0 \\ RN_i = 0, y_{i,RN}^* \leq 0 \end{cases}, \begin{cases} CER_i = 1, y_{i,CER}^* > 0 \\ CER_i = 0, y_{i,CER}^* \leq 0 \end{cases}$$

где  $y_{i,RN}^*$  и  $y_{i,CER}^*$  – латентные переменные, причем

$$\begin{cases} y_{i,RN}^* = \beta_{10} + \beta_{11} \cdot IPER_i + \beta_{12} \cdot LOGZP_i + \beta_{13} \cdot MIGR_i + \beta_{14} \cdot CER_{t-1i} + \varepsilon_{i,RN} \\ y_{i,CER}^* = \beta_{20} + \beta_{21} \cdot RN_i + \beta_{22} \cdot SOZ_i + \beta_{23} \cdot VZD_i + \beta_{24} \cdot KB_i + \varepsilon_{i,UOZ} \end{cases}$$

$$\begin{pmatrix} \varepsilon_{i,RN} \\ \varepsilon_{i,CER} \end{pmatrix} \sim N(\mu, \Sigma), \mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix},$$

$\rho$  – ковариация ошибок  $\varepsilon_{i,RN}$  и  $\varepsilon_{i,CER}$ .

LOGZP – логарифм средней заработной платы (тыс.руб.),

SOZ – средняя обеспеченность жильем (на душу насел., кв. м.),

VZD – ввод в действие жилых домов (на душу насел., руб.),

KB – количество безработных (на 1 тыс. чел., чел.),

$$X_{i,1} = (IPER_i, LOGZP_i, MIGR_i, CER_{t-1i}),$$

$$X_{i,2} = (RN_i, SOZ_i, VZD_i, KB_i).$$

Результаты оценивания модели в статистическом пакете Stata приведены в таблицах (3.23) и (3.24).

Таблица 3.23 – Результаты оценивания двумерной модели, описывающей взаимосвязь эколого-экономического риска и показателей, характеризующих уровень социально-экономического развития территорий региона

Показатель	Оценка ко-эфф.	Станд. отклон.	z-стат.	Уровень знач.	Доверительный интервал	
Уравнение для эколого-экономического риска, $\hat{R}_{pseudo}^2 = 0,41$						
CER <sub>t-1</sub>	0,494	0,154	3,21	0,000	0,192	0,795
IPER	-0,025	0,010	-2,63	0,008	-0,044	-0,006
LOGZP	0,710	0,298	2,38	0,017	0,125	1,295
MIGR	0,039	0,012	3,15	0,002	0,015	0,063
cons	-0,873	0,291	-3,00	0,003	-1,443	-0,303
Уравнение для уровня социально-экономического развития региона, $\hat{R}_{pseudo}^2 = 0,70$						
RN	1,434	0,295	4,86	0,000	0,856	2,013
SOZ	-0,177	0,049	-3,62	0,000	-0,273	-0,081
VZD	0,002	0,001	2,84	0,004	0,001	0,003
KB	-0,074	0,043	-1,72	0,085	-0,158	0,010
cons	2,166	1,021	2,12	0,034	0,165	4,167
Wald chi2(7) = 159,80		Prob > chi2 = 0,0000				
chi2(1) = 11,7124		Prob > chi2 = 0,0006				

На основании таблицы 3.23 можно сделать вывод, что высокий уровень социально-экономического развития соответствует городам и районам области с высоким значением показателя ввода в действие жилых домов. Низкий уровень социально-экономического развития характерен для городов и районов Оренбургской области с высоким уровнем безработицы. Кроме того, для муниципальных образований Оренбургской области с высоким уровнем социально-экономического развития характерно наличие высоких эколого-экономических рисков для населения, при этом доказано влияние лага: повышение уровня социально-экономического развития в период  $t$  приводит в следующий  $t+1$  период к повышению эколого-экономического риска.

Таблица 3.24 – Характеристики двумерной модели, описывающей взаимосвязь эколого-экономического риска и уровня социально-экономического развития муниципалитетов

Характеристика	Годы							Среднее значение
	2001	2002	2003	2004	2005	2006	2007	
P (RN=1, CER=1)	0,15	0,15	0,15	0,16	0,21	0,22	0,17	0,17
P (RN=1, CER=0)	0,28	0,28	0,32	0,35	0,36	0,43	0,43	0,35
P (RN=0, CER=1)	0,26	0,26	0,25	0,21	0,24	0,21	0,23	0,24
P (RN=0, CER=0)	0,31	0,31	0,27	0,28	0,19	0,15	0,18	0,24
P (RN=1)	0,43	0,43	0,47	0,51	0,57	0,64	0,59	0,52
P (CER=1)	0,41	0,41	0,40	0,37	0,45	0,42	0,40	0,41
P (RN=1   CER=1)	0,22	0,22	0,22	0,24	0,32	0,34	0,28	0,26
P (CER=1   RN=1)	0,23	0,23	0,21	0,21	0,27	0,27	0,21	0,23

Среднее значение вероятности того, что в городах и районах области с высоким уровнем социально-экономического развития присутствуют высокие эколого-экономические риски для населения, составляет 0,17. Наличие такой ситуации наиболее вероятно в Бузулукском, Новоорском, Оренбургском и Сакмарском районах, в городах Новотроицк и Оренбург.

Среднее значение вероятности того, что города и районы области одновременно характеризуются низким эколого-экономическим риском и высоким уровнем социально-экономического развития составляет 0,24. Наибольшие вероятно-



сти возникновения такой ситуации наблюдаются в Адамовском, Беляевском, Кувандыкском и Соль-Илецком районах.

В последнее время для Оренбургской области наиболее вероятной является ситуация, связанная с наличием значительного эколого-экономического риска для населения и низким уровнем социально-экономического развития территорий региона. Такая ситуация наиболее характерна для Абдулинского, Гайского, Переволоцкого, Саракташского и Сорочинского районов, городов Медногорск и Орск.

Анализ рассмотренных характеристик за 2001–2007 гг. показывает, что по сравнению с 2001 г. на конец рассматриваемого периода наблюдается снижение уровня социально-экономического развития территорий региона и увеличение эколого-экономических рисков для населения.

Таким образом, построены модели, связывающие уровень эколого-экономического риска с показателями, характеризующими социально-экономическое состояние муниципальных образований региона. Выяснено, что сложившиеся в большинстве муниципалитетов области тенденции при отсутствии мер со стороны органов власти приведут к дальнейшему росту эколого-экономического риска, следствием чего в краткосрочной перспективе будет, как минимум, повышение заболеваемости населения, а в долгосрочной перспективе – снижение инвестиционной привлекательности региона. Среди факторов, связанных с высоким уровнем эколого-экономического риска, выделяются заработная плата и розничный товароборот - на текущем этапе развития Оренбургской области экономический рост приводит к повышению эколого-экономического риска. Выяснено, что на наличие эколого-экономического риска оказывают значительное влияние состояние окружающей среды и показатели, отражающие экономическую привлекательность муниципального образования.

### Вопросы и задания по теоретическому материалу раздела 3

1. Приведите примеры социально-экономических проблем, для исследования которых подходящим инструментарием были бы модели бинарного и множественного выбора. Для каждого приведенного примера проанализируйте возможность и целесообразность увеличения (уменьшения) количества рассматриваемых градаций результативной переменной. В каких из приведенных Вами примеров Вы можете столкнуться с проблемой эндогенности?

2. Приведите примеры социально-экономических проблем, при исследовании которых обязательно возникает проблема эндогенности (случаи с бинарными результативными переменными).

3. Дайте классификацию моделей с дискретной зависимой переменной.

4. Запишите функцию регрессии для дискретной зависимой переменной, принимающей только два возможных значения, как функцию от некоторого набора объясняющих (факторных) переменных. Укажите известные Вам подходы к оценке ее коэффициентов.

5. Поясните смысл латентной переменной, вводимой в моделях бинарного и множественного выбора. Приведите примеры.

6. Объясните, почему в качестве функции распределения остатков в модели бинарного выбора может быть использована функция стандартного нормального распределения/функция логистического распределения. Предложите другие подходящие функции.

7. Дайте понятие предельного эффекта. Покажите, почему в моделях бинарного и множественного выбора коэффициенты не могут интерпретироваться аналогично коэффициентам линейной модели регрессии. Запишите предельный эффект для логит- и пробит-модели бинарного выбора, для логит- и пробит-модели упорядоченного множественного выбора.

8. Объясните, зачем нужна нормализация в моделях бинарного и множественного выбора. Какие виды нормализации Вам известны?

9. Сколько неизвестных параметров необходимо оценивать в модели бинарного выбора с 5 факторными признаками? в модели упорядоченного множественного выбора с 3 альтернативами? Уточните, как повлияет введение нормализации на количество оцениваемых параметров.

10. Опишите схему оценивания коэффициентов модели бинарного/множественного выбора на основе метода максимального правдоподобия.

11. Опишите схему статистического анализа оцененной модели бинарного/множественного выбора.

12. Назовите известные Вам характеристики качества моделей бинарного и множественного выбора. Поясните, какая идея лежит в основе построения каждой из них.

13. Как осуществляется прогнозирование по модели бинарного выбора? по модели упорядоченного множественного выбора?

14. Опишите, какая информация необходима для построения интегрального показателя на основе моделей бинарного/множественного выбора. Как можно получить наблюдаемые значения результативной переменной для построения такой модели?

15. В чем, на Ваш взгляд, преимущества и недостатки построения интегрального показателя на основе моделей бинарного/множественного выбора?

16. В каком случае допускается раздельное оценивание коэффициентов уравнений двумерной пробит-модели?

17. Прокомментируйте результаты оценивания логит-модели бинарного выбора для результативного показателя  $y_1$  и факторных признаков  $x_1, x_2, x_3, x_4$  с точки зрения значимости модели и ее коэффициентов модели, качества модели:

```
. logit y1 x1 x2 x3 x4
Iteration 0:  log likelihood = -117.47043
Iteration 1:  log likelihood = -55.996481
Iteration 2:  log likelihood = -55.118215
Iteration 3:  log likelihood = -55.110327
Iteration 4:  log likelihood = -55.110323

Logistic regression              Number of obs   =       171
LR chi2(4)                       =       124.72
Prob > chi2                       =       0.0000
Pseudo R2                        =       0.5309

Log likelihood = -55.110323
```

	y1	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
	x1	-.2747377	.049597	-5.54	0.000	-.371946 - .1775294
	x2	-.0124799	.0147956	-0.84	0.399	-.0414787 .0165189
	x3	.0281605	.0114542	2.46	0.014	.0057108 .0506103
	x4	.205434	.0514054	4.00	0.000	.1046812 .3061869
	_cons	-6.393775	3.082215	-2.07	0.038	-12.4348 - .3527447

18. Прокомментируйте результаты оценивания логит-модели бинарного выбора для факторных признаков  $x_1$ ,  $x_2$ ,  $x_3$  и  $x_4$  используя понятие «отношение шансов» (Odds Ratio):

```
Logistic regression
```

	Number of obs	=	171
	LR chi2(4)	=	124.72
	Prob > chi2	=	0.0000
	Pseudo R2	=	0.5309

Log likelihood = -55.110323

	y1	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
	x1	.7597714	.0376824	-5.54	0.000	.6893914 .8373364
	x2	.9875977	.0146121	-0.84	0.399	.9593698 1.016656
	x3	1.028561	.0117813	2.46	0.014	1.005727 1.051913
	x4	1.228058	.0631289	4.00	0.000	1.110357 1.358236

19. Прокомментируйте результаты оценивания пробит-модели бинарного выбора для результативного показателя  $y_1$  и факторных признаков  $x_1$ ,  $x_2$ ,  $x_3$ ,  $x_4$  с точки зрения значимости модели, значимости коэффициентов модели и ее качества:

```
. probit y2 x1 x2 x3 x4
```

```
Iteration 0: log likelihood = -99.567747
Iteration 1: log likelihood = -92.976978
Iteration 2: log likelihood = -92.944457
Iteration 3: log likelihood = -92.944453
```

```
Probit regression
```

	Number of obs	=	171
	LR chi2(4)	=	13.25
	Prob > chi2	=	0.0101
	Pseudo R2	=	0.0665

Log likelihood = -92.944453

	y2	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
	x1	.0476077	.0195349	2.44	0.015	.00932 .0858954
	x2	.0032423	.0057031	0.57	0.570	-.0079355 .0144201
	x3	.0004066	.0047878	0.08	0.932	-.0089773 .0097905
	x4	.0707483	.0208176	3.40	0.001	.0299466 .11155
	_cons	-3.549767	1.327356	-2.67	0.007	-6.151338 -.9481972

20. Прокомментируйте результаты оценки предельных эффектов для результативного показателя факторных признаков  $x_1$ ,  $x_2$ ,  $x_3$  и  $x_4$  пробит-модели бинарного выбора:

```
. dprobit y2 x1 x2 x3 x4
```

```
Iteration 0: log likelihood = -99.567747
Iteration 1: log likelihood = -92.976978
Iteration 2: log likelihood = -92.944457
Iteration 3: log likelihood = -92.944453
```

```
Probit regression, reporting marginal effects
```

	Number of obs	=	171
	LR chi2(4)	=	13.25
	Prob > chi2	=	0.0101
	Pseudo R2	=	0.0665

Log likelihood = -92.944453

	y2	dF/dx	Std. Err.	z	P> z	x-bar [ 95% C.I. ]
	x1	.0152823	.0062196	2.44	0.015	21.5965 .003092 .027472
	x2	.0010408	.0018302	0.57	0.570	66.807 -.002546 .004628
	x3	.0001305	.001537	0.08	0.932	150.444 -.002882 .003143
	x4	.0227105	.0066204	3.40	0.001	41.0351 .009735 .035686
	obs. P	.7309942				
	pred. P	.7451599 (at x-bar)				

z and P>|z| correspond to the test of the underlying coefficient being 0

21. Прокомментируйте результаты оценивания пробит-модели упорядоченного множественного выбора для результативного показателя  $y_1$  и факторных признаков  $x_1, x_2, x_3, x_4$  с точки зрения значимости модели, значимости коэффициентов модели и ее качества:

```
. oprobit y3 x1 x2 x3 x4
```

```
Iteration 0: log likelihood = -170.94867
Iteration 1: log likelihood = -117.71542
Iteration 2: log likelihood = -115.03282
Iteration 3: log likelihood = -115.02811
Iteration 4: log likelihood = -115.02811
```

```
Ordered probit regression
```

Log likelihood = -115.02811	Number of obs	=	171
	LR chi2(4)	=	111.84
	Prob > chi2	=	0.0000
	Pseudo R2	=	0.3271

	y3	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	x1	.1185439	.0210523	5.63	0.000	.0772821	.1598057
	x2	.011036	.0054222	2.04	0.042	.0004088	.0216632
	x3	-.013312	.0044901	-2.96	0.003	-.0221125	-.0045115
	x4	-.0817393	.019885	-4.11	0.000	-.1207133	-.0427654
	/cut1	-3.59281	1.231868			-6.007228	-1.178393
	/cut2	-1.071642	1.211793			-3.446713	1.30343

22. Прокомментируйте результаты оценивания двумерной пробит-модели для результативных показателей  $y_1$  и  $y_2$  и факторных признаков  $x_1, x_2, x_3, x_4$  с точки зрения возможности раздельного оценивания уравнений системы:

```
Seemingly unrelated bivariate probit
```

Log likelihood = -145.73631	Number of obs	=	171
	Wald chi2(7)	=	134.87
	Prob > chi2	=	0.0000

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]		
y1	y2	-1.336901	.3021572	-4.42	0.000	-1.929119	-.744684
	x1	-.1098343	.0211931	-5.18	0.000	-.151372	-.0682965
	x2	-.002735	.0054081	-0.51	0.613	-.0133347	.0078646
	_cons	3.341709	.5588713	5.98	0.000	2.246341	4.437076
	y2	-1.554636	.292526	-5.31	0.000	-2.127976	-.9812954
y2	x1	-.0012216	.0055125	-0.22	0.825	-.0120259	.0095828
	x2	.0024319	.0044923	0.54	0.588	-.0063728	.0112367
	x3	.062315	.0207124	3.01	0.003	.0217195	.1029105
	x4	-.062315	.0207124	-3.01	0.003	-.0217195	-.1029105
	_cons	-1.581454	1.003101	-1.58	0.115	-3.547496	.3845881
/athrho	15.41864	1429.607	0.01	0.991	-2786.56	2817.397	
rho	1	2.31e-10			-1	1	

Likelihood-ratio test of rho=0:      chi2(1) = 32.4935      Prob > chi2 = 0.0000

## Задания к исследовательской работе

1. На основе показателей по субъектам РФ или по муниципальным образованиям области (края, республики) постройте интегральный показатель, характеризующий экономическую безопасность на основе моделей бинарного или множественного выбора.

2. На основе информации о продолжительности поиска работы безработными и их индивидуальных характеристик выявите факторы, оказывающие влияние на продолжительность пребывания в статусе безработного. Ранжируйте факторы по степени их влияния на резульативный показатель.

Безработные, зарегистрированные в Центре занятости населения, характеризуются следующими признаками:

- пол: мужской, женский;
- возраст: до 29 лет, от 30 до 49 лет, от 50 лет и старше;
- уровень профессионального образования: высшее, среднее, начальное, нет профессионального образования;
- профессия: рабочий, служащий;
- наличие опыта работы: нет опыта работы, есть опыт работы.

Тестовые исходные данные приведены в файле bezrub.xls. Резульативную переменную сформировать самостоятельно: например, 1 – «трудоустроен в течение  $t_1$  дней», 0 – «не трудоустроен в течение  $t_1$  дней»; 1 – «трудоустроен в течение  $t_1$  дней», 2 – «трудоустроен трудоустроен в течение  $t_2$  дней» и 3 – «не трудоустроен в течение  $t_2$  дней» ( $t_1 < t_2$ ).

Все ли объясняющие переменные в модели можно рассматривать как экзогенные? При необходимости осуществите учет эндогенности, оценив двумерную модель бинарного выбора.

3. На основе информации об уровне эколого-экономического риска в административно-территориальных образованиях Оренбургской области выявите факторы, оказывающие на него влияние.

Тестовые данные представлены в файле Risk.xls. Переменная «класс» принимает значение 1, если уровень эколого-экономического риска в административно-территориальном образовании высокий, 2 – если средний, 3 – если низкий. Переменная «нагрузка на окружающую среду» принимает значение 1, если уровень антропогенной нагрузки на среду в административно-территориальном образовании высокий, 2 – если средний, 3 – если низкий.

По возможности расширьте перечень анализируемых факторов. Используйте как аппарат моделей множественного, так и бинарного выбора. Проведите содержательный и статистический анализ эндогенности объясняющих переменных (факторов) модели. При необходимости выполните учет эндогенности.

### **Список использованных источников к разделу 3**

- 1 Тихомиров, Н. П. Эконометрика / Н.П. Тихомиров, Е. Ю. Дорохина – М.: Издательство «Экзамен», 2003. – 512 с.
- 2 Айвазян, С. А. Прикладная статистика. Основа эконометрики : в 2 т.: учеб. для вузов / С. А. Айвазян, В. С. Мхитарян. - М. : ЮНИТИ-ДАНА, 2001. – 432 с.
- 3 Эконометрика: учеб. / под ред. д-ра экон. наук, проф. В.С. Мхитаряна. – Проспект, 2009. – 384 с.
- 4 Носко, В. П. Эконометрика: учебник / В.П. Носко. – Кн. 2 Ч. 3, 4. - М.: Издательский дом «Дело» РАНХиГС, 2011. – 576 с.
- 5 Давнис, В. В. Прогнозные модели экспертных предпочтений: монография / В.В Давнис, В.В. Тинякова. – Воронеж: Изд-во Воронеж. гос. ун-та, 2005. – 248 с.
- 6 Greene, W. H. Econometric Analysis / W. H. Greene. – 5<sup>th</sup> ed. – Prentice Hall, 2002. – 1026 p.
- 7 Чудинова, О. С. Ранжирование категорий безработных по степени их востребованности на рынке труда / О.С. Чудинова // Вестник ОГУ. – 2011. – №8. – С. 205–208.
- 8 Васянина, В. И. Сравнительный анализ административно-территориальных образований региона по уровню экономической безопасности / В.И. Васянина // Вестник ОГУ. – 2008. – № 10. – С. 76-80.
- 9 Седова, Е. Н. Анализ предельных эффектов влияния социальных и экономических факторов на уровень риска для населения от воздействия окру-

жающей среды / Е.Н. Седова // Материалы Международной научно-практической конференции «Современные проблемы региональной экономики, управления и юриспруденции». – Мурманск: МГТУ, 2008 (электронный ресурс), с.881-886.

10 Лукашин, Ю. П. Адаптивные методы краткосрочного прогнозирования временных рядов: учебное пособие / Ю. П. Лукашин. – М.: Финансы и статистика, 2003. – 416 с.

11 Stewart, M. B. A Comparison of Semiparametric estimators for the Ordered Response Model// Stata Journal – 2004, vol. 4(1), pages 27-39

12 Stewart, M. B. Semi-nonparametric estimation of extended ordered probit models// Preprint submitted to Elsevier Science, 2004, 23 p.

13 Stewart, M. B. Pensioner financial well-being, equivalence scales and ordered response models. // 2002. - 34 p.

14 Седова, Е.Н. Моделирование экологических рисков на основе порядковых пробит-моделей / Е.Н. Седова // Системное моделирование социально-экономических процессов и систем. Сборник трудов 30-й Юбилейной Международной школы-семинара им. академика С. С. Шаталина. – Воронеж, Издательско-полиграф. Центр Воронежского государственного университета. – 2007. – Ч. 2, - С. 243-248.

15 Седова, Е. Н. Модели бинарного и множественного выбора в задачах управления эколого-экономическими рисками / Е.Н. Седова // Вестник ОГУ. – 2008. - № 8. – С. 96-102.

16 Седова, Е. Н. Моделирование экологических рисков и заболеваемости на основе двумерной пробит-модели / Е.Н. Седова // Сборник статей IV Международной научно-практической конференции «Экономическое прогнозирование: методы и модели». – Воронеж: Издательско-полиграфический центр Воронежского государственного университета. - 2008. - Ч.1,- С. 286-289.



## **4 Моделирование социально-экономических процессов с учетом пространственно-временной неоднородности данных**

При моделировании социально-экономических процессов исследователь сталкивается с проблемой неоднородности данных в пространстве и во времени. Методы регрессионного анализа, применяемые к данным за разные годы или к объединенным данным за ряд лет (метод «заводо-лет»), не позволяют учесть неоднородность данных. Поэтому целесообразным представляется применять модели регрессии для панельных данных. Благодаря специальной структуре они позволяют строить более гибкие и содержательные модели, которые дают возможность учитывать и анализировать как индивидуальные отличия между рассматриваемыми объектами, так и временные эффекты [1, 2, 3, 4, 5, 6].

Еще одна ситуация связана с тем, что исходные данные являются пространственно зависимыми и измеряются на достаточно большой территории неоднородной по составу, например, при моделировании ценообразования недвижимости для рынков различных городов. Пространственно представленные данные включают явную географическую привязку и несут в себе информацию о местоположении, поэтому рекомендуется при их анализе применения метода географически взвешенной регрессии [7, 8, 9, 10, 11].

Раздел организован следующим образом. В подразделе 4.1 описываются вопросы построения и исследования регрессионных моделей панельных данных. В подразделе 4.2 по значениям индивидуальных эффектов модели панельных данных осуществлено ранжирование категорий безработных. В подразделе 4.3. на основе панельных данных осуществлено моделирование стоимости жилья на вторичном рынке г. Оренбурга. В подразделе 4.4 на основе панельных данных построен интегральный показатель, характеризующий уровень демографической безопасности, с учетом неоднородности данных во времени. В подразделе 4.5 для выявления характера взаимосвязей факторов, оказывающих существенное влияние на составляющие инвестиционной привлекательности с учетом про-

пространственной и временной неоднородности данных, предложены модели в форме рекурсивной системы одновременных уравнений на основе панельных данных. И, наконец, в подразделе 4.6 осуществлено моделирование зависимости уровня экологического риска от показателей социально-экономического состояния муниципалитета с учетом пространственной и временной неоднородности данных с помощью бинарных моделей.

#### **4.1 Модели линейной регрессии для панельных данных**

Эконометрические модели, опирающиеся на данные пространственных выборок или временных рядов, носят агрегированный характер и описывают поведение лишь усредненных объектов. При оценивании моделей по таким данным может возникать «смещение агрегирования», приводящее к не согласующимся с экономической теорией значениям и даже знакам коэффициентов моделей. Кроме того, возникают трудности, связанные с невозможностью учесть влияние на результативный показатель (например, уровень эколого-экономического риска) некоторых специфических трудноизмеряемых факторов, таких как особенности климата и географического положения муниципальных образований, степень деградации, устойчивости или ассимиляционный потенциал природной среды, субъективно воспринимаемая вероятность избежания ответственности за совершение экологических нарушений, качество управления природоохранной деятельностью и др. [1, 6]

Выходом из такой ситуации является сбор и использование данных иной структуры, получившей название панельной. Панельные данные (от англ. «panel») – это данные, содержащие сведения об одном и том же множестве объектов за ряд последовательных периодов времени. В качестве объектов могут выступать индивидуумы или их группы, предприятия, домохозяйства, регионы, страны и т.д.

Работая с панельными данными, мы одновременно рассматриваем и пространственный, и временной аспекты данных, что позволяет провести их более глубокий анализ и дает, как минимум, четыре преимущества.

Во-первых, большее количество наблюдений обеспечивает большую эффективность оценивания параметров эконометрической модели.

Во-вторых, модели панельных данных предоставляют разнообразные возможности учета неоднородности, причем как неоднородности объектов, так и временной неоднородности: они позволяют предотвратить смещение агрегированности, возникающее как при анализе временных рядов, где рассматривается эволюция усредненного объекта, так и при анализе перекрестных данных, где не учитываются ненаблюдаемые индивидуальные характеристики объектов.

В-третьи, панельные данные за счет последовательного наблюдения за одними и теми же объектами позволяют отделить межиндивидуальные различия от интраиндивидуальных. Понимание дополнительных источников вариации может дать весьма полезную информацию для отделения индивидуальной динамики от средней.

В-четвертых, имея панельные данные, можно уменьшить или даже избежать ошибок спецификации и смещенности оценок, обусловленных невключением в модель существенных переменных.

Кроме того, панельные данные позволяют строить и тестировать более сложные поведенческие модели, чем модели на основе временных рядов или модели по пространственным данным в отдельности, а также решают проблему поиска хороших инструментов при оценивании моделей с эндогенными регрессорами.

Для удобства дальнейшего изложения введены следующие обозначения:

$n$  – число объектов наблюдения;

$T$  – число периодов времени;

$k$  – число объясняющих переменных;

$y$  – вектор размерности  $nT$  с элементами  $y_{it}$  – значение зависимой переменной для  $i$ -ого объекта наблюдения в момент времени  $t$ ,  $i = \overline{1, \dots, n}$ ,  $t = \overline{1, \dots, T}$ ;

$X$  – матрица размерности  $nT \times k$  с элементами  $x_{j,it}$  – значение  $j$ -ой объясняющей переменной для  $i$ -ого объекта наблюдения в момент времени  $t$ ,  $j = \overline{1, \dots, k}$ ,  $i = \overline{1, \dots, n}$ ,  $t = \overline{1, \dots, T}$ .

#### 4.1.1 Модель линейной регрессии с фиксированными эффектами

Рассмотрим модель:

$$y_{it} = \alpha_i + \sum_{j=1}^k \beta_j x_{j,it} + \varepsilon_{it}, \quad (4.1)$$

где  $\alpha_i$  – индивидуальные эффекты;

$\beta_{j,it}$  – неизвестные коэффициенты, подлежащие оцениванию;

$\varepsilon_{it}$  – ошибка.

$$M(\varepsilon_{it}) = 0; \quad M(\varepsilon_{it}^2) = \sigma^2;$$

$$\text{cov}(\varepsilon_{it}, \varepsilon_{ps}) = 0, \text{ если } i \neq p \text{ или } t \neq s;$$

$$\text{cov}(\varepsilon_{it}, x_{j,ps}) = 0;$$

$$i, p = \overline{1, \dots, n}; \quad t, s = \overline{1, \dots, T}; \quad j = \overline{1, \dots, k},$$

параметры  $\alpha_i$  в модели (4.1) называются индивидуальными (фиксированными) эффектами. Они учитывают влияние на результирующий показатель всех (наблюдаемых или ненаблюдаемых) переменных, которые принимают разные значения для разных объектов, но не меняются во времени. Эти характеристики могут не только оказывать влияние на результирующий показатель, но и быть коррелированными с объясняющими переменными. Если в такой ситуации не учитывать панельную структуру данных и рассматривать обычную регрессию по объединенным данным, то это приведет к смещенным оценкам параметров.

Модель (4.1) идентична модели с фиктивными переменными с  $k + n$  неизвестными, для оценки которых можно использовать обычный метод наименьших квадратов. Однако большое количество  $n$  фиктивных переменных приводит к необходимости обращать матрицу большой размерности. Поэтому предпочтительнее для оценки коэффициентов модели (4.1) использовать двухступенчатую процедуру, при которой сначала с помощью метода наименьших квадратов вычисляются оценки  $\hat{\beta}_j \equiv b_j$  (within оценки или оценки с учетом вариации в рамках объекта наблюдения, внутригрупповые оценки) регрессии, построенной по отклонениям от групповых средних:

$$(y_{it} - \bar{y}_i) = \sum_{j=1}^k \beta_j (x_{j,it} - \bar{x}_{j,i}) + (\varepsilon_{it} - \bar{\varepsilon}_i), \quad i = \overline{1, \dots, n}, \quad t = \overline{1, \dots, T}, \quad (4.2)$$

а затем оценки индивидуальных эффектов с помощью формулы:

$$\hat{\alpha}_i \equiv a_i = \bar{y}_i - \sum_{j=1}^k b_j \bar{x}_{j,i}, \quad i = \overline{1, \dots, n}. \quad (4.3)$$

Таким образом, оценки коэффициентов при объясняющих переменных в модели с фиксированными эффектами можно получить с помощью стандартных пакетов статистического анализа [1]. Для этого необходимо найти для всех входящих в модель переменных средние значения для каждого объекта за весь период наблюдения и вычесть групповые средние из исходных данных (провести внутригрупповое преобразование). К преобразованным данным можно применять стандартные методы регрессионного анализа. Оценки коэффициентов модели (4.2) совпадут с оценками коэффициентов модели с фиксированными эффектами, а оценку остаточной дисперсии, стандартные ошибки коэффициентов, значения  $t$ -статистики необходимо будет скорректировать, поскольку модели (4.1) и (4.2) имеют разное число степеней свободы ( $nT - n - k$  и  $nT - k$  соответственно). Кроме того, различными будут и величины общих сумм квадратов, рассчитанные по

исходным и трансформированным данным [4, 5]. Следовательно, значение  $F$ -статистики и коэффициент детерминации, получающиеся в результате оценивания модели (4.2), не будут совпадать с аналогичными характеристиками модели (4.1).

Условия, наложенные на модель (4.1), гарантируют несмещенность и состоятельность оценок  $b_j$ . Оценки  $a_i$  являются несмещенными и состоятельными для фиксированного  $n$  и при  $T \rightarrow \infty$ . При  $n \rightarrow \infty$  или при  $T \rightarrow \infty$  оценки  $b_j$  являются асимптотически нормальными, поэтому можно пользоваться стандартными процедурами для проверки гипотез относительно параметров  $\beta_j$  [2, 3].

Статистика для проверки гипотезы  $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$  имеет вид:

$$F = \frac{(TSS_{FE} - RSS_{FE}) / k}{RSS_{FE} / (nT - n - k)}, \quad (4.4)$$

где  $RSS_{FE}$  – сумма квадратов остатков модели с фиксированными эффектами (4.2);

$$TSS_{FE} = \sum_{i=1}^n \sum_{t=1}^T (y_{it} - \bar{y}_i)^2.$$

Если справедлива гипотеза  $H_0$  и выполняется предпосылка о нормальном распределении ошибок (или в случае достаточно больших  $n$ ), статистика (4.5) имеет (приблизительно)  $F$ -распределения с  $k$  и  $(nT - n - k)$  степенями свободы.

Один из наиболее интересных вопросов в отношении параметров модели с фиксированными эффектами заключается в том, отличаются ли параметры  $\alpha_i$  для разных объектов наблюдения. Если эффектов, специфических для отдельных объектов наблюдения, нет, все данные могут быть объединены и вместо модели с фиксированными эффектами предпочтение отдается обычной регрессии с единственной константой [1, 2, 3]. Нулевая гипотеза формулируется  $H_0 : \alpha_i = \alpha_j = \alpha$  для любых  $i, j$ . Эта гипотеза соответствует объединенной модели регрессии:

$$y_{it} = \alpha + \sum_{j=1}^k \beta_j x_{j,it} + \varepsilon_{it}, \quad i = \overline{1, \dots, n}, \quad t = \overline{1, \dots, T}. \quad (4.5)$$

Альтернативная гипотеза  $H_1 : \alpha_i \neq \alpha_j$  хотя бы для одной пары  $i, j$ , что соответствует модели с фиксированными эффектами (4.2). Для проверки нулевой гипотезы используется  $F$ -статистика:

$$F = \frac{(RSS_{pool} - RSS_{FE}) / (n - 1)}{RSS_{FE} / (nT - n - k)}, \quad (4.6)$$

где  $RSS_{pool}$  – сумма квадратов остатков модели (4.5).

Если справедлива гипотеза  $H_0$  и выполняется предпосылка о нормальном распределении ошибок (или в случае достаточно больших  $n$ ), статистика (4.6) имеет (приближенно)  $F$ -распределения с  $(n - 1)$  и  $(nT - n - k)$  степенями свободы.

Фиктивные переменные могут использоваться и для учета временных эффектов [1, 2]. Коэффициенты при фиктивных переменных для каждого периода времени вберут в себя влияние всех наблюдаемых или ненаблюдаемых переменных, которые зависят только от времени, но одинаковы для всех единиц совокупности. Это позволит выявить и учесть в модели особенности характерные для каждого из рассматриваемых периодов времени. Двухнаправленная модель с фиксированными эффектами имеет вид:

$$y_{it} = \alpha_i + \gamma_t + \sum_{j=1}^k \beta_j x_{j,it} + \varepsilon_{it}, \quad i = \overline{1, \dots, n}, \quad t = \overline{1, \dots, T}, \quad (4.7)$$

где  $\gamma_t$  – эффект специфический для каждого периода.

Модель (4.7) получена из модели (4.1) включением дополнительных  $T - 1$  фиктивных переменных (не включен в модель из-за строгой коллинеарности, на-

пример, эффект  $\gamma_T$ ). Оценки модели (4.7) могут быть найдены с помощью двух-ступенчатой процедуры, описанной выше. Модель (4.2) в этом случае имеет вид:

$$(y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}) = \sum_{j=1}^k \beta_j (x_{j,it} - \bar{x}_{j,i} - \bar{x}_{j,t} + \bar{x}_j) + (\varepsilon_{it} - \bar{\varepsilon}_i - \bar{\varepsilon}_t + \bar{\varepsilon}), \quad i = \overline{1, \dots, n}, \quad t = \overline{1, \dots, T}. \quad (4.8)$$

Оценки временных эффектов рассчитываются по формуле:

$$\hat{\gamma}_t \equiv c_t = (\bar{y}_t - \bar{y}) - \sum_{j=1}^k b_j (\bar{x}_{j,t} - \bar{x}_j), \quad t = \overline{1, \dots, T}. \quad (4.9)$$

Статистический анализ модели (4.7) включает проверку следующих гипотез [1]:

а) о значимости индивидуальных и временных эффектов

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_n, \quad \gamma_1 = \gamma_2 = \dots = \gamma_{T-1} = 0$$

с помощью  $F$ -критерия  $F = \frac{(RSS_{pool} - RSS)/(n + T - 2)}{RSS/(nT - n - T - k + 1)}$ , где  $RSS$  – сумма

квадратов остатков модели (4.7);

б) о значимости временных эффектов  $H_0 : \gamma_1 = \gamma_2 = \dots = \gamma_{T-1} = 0$  с помощью  $F$ -критерия  $F = \frac{(RSS_{FE} - RSS)/(T - 1)}{RSS/(nT - n - T - k + 1)}$ ;

в) о значимости индивидуальных эффектов  $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_n$  с помощью  $F$ -критерия  $F = \frac{(RSS_{FET} - RSS)/(N - 1)}{RSS/(nT - n - T - k + 1)}$ , где  $RSS_{FET}$  – сумма квадратов

остатков модели регрессии с фиксированными временными эффектами

$$y_{it} = \gamma_t + \sum_{j=1}^k \beta_j x_{j,it} + \varepsilon_{it};$$

г) проверка значимости коэффициентов при объясняющих переменных

$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$  с помощью  $F$ -критерия  $F = \frac{(TSS - RSS)/k}{RSS/(nT - n - T - k + 1)}$ ,



$$\text{где } TSS = \sum_{i=1}^n \sum_{t=1}^T (y_{it} - \bar{y}_i - \bar{y}_t + \bar{y})^2.$$

#### 4.1.2 Модель линейной регрессии со случайными эффектами

Если предполагается, что ненаблюдаемые переменные не коррелированы с остальными регрессорами, то их влияние учитывается иначе – как компоненты ошибок наблюдения. В этом случае для панельных данных используется модель со случайными эффектами, имеющая вид:

$$y_{it} = \alpha + \sum_{j=1}^k \beta_j x_{j,it} + \varepsilon_{it}, \quad (4.10)$$

$$\varepsilon_{it} = u_i + v_{it};$$

$$M(u_i) = 0; M(u_i^2) = \sigma_u^2;$$

$$M(v_{it}) = 0; M(v_{it}^2) = \sigma_v^2;$$

$$\text{cov}(v_{it}, u_p) = 0;$$

$$\text{cov}(u_i, u_p) = 0, \text{ если } i \neq p;$$

$$\text{cov}(v_{it}, v_{ps}) = 0, \text{ если } i \neq p \text{ или } t \neq s;$$

$$\text{cov}(\varepsilon_{it}, x_{j,ps}) = 0, \text{ cov}(u_i, x_{j,ps}) = 0;$$

$$i, p = \overline{1, \dots, n}, t, s = \overline{1, \dots, T}, j = \overline{1, \dots, k}.$$

Суммарная ошибка в модели (4.10)  $\varepsilon_{it}$  складывается из ошибки, характерной для  $i$ -го объекта и не зависящая от времени  $u_i$  и случайной ошибки регистрации  $v_{it}$ . Модель (4.10) представляет собой модель линейной регрессии при гетероскедастичности ошибок, поэтому для получения эффективных оценок неизвестных параметров необходимо использовать обобщенный метод наименьших квадратов, требующий предварительного оценивания дисперсий  $\sigma_u^2$  и  $\sigma_v^2$ .

Оценки коэффициентов модели (4.10) могут быть также найдены как средневзвешенные внутри- и межгрупповые оценки [1, 2, 3].

Также как и в модели с фиксированными эффектами встает вопрос: различаются ли компоненты ошибки  $u_i$  у разных объектов наблюдения? Бреуш и Паган предложили метод множителей Лагранжа для проверки гипотезы о значимости случайных эффектов, основанный на остатках обычной регрессии, построенной по объединенным данным. Выдвигаются гипотезы:

$$H_0 : \sigma_u^2 = 0;$$

$$H_1 : \sigma_u^2 > 0.$$

Для проверки нулевой гипотезы используется статистика: [1]

$$LM = \frac{nT}{2(T-1)} \left[ \frac{\sum_{i=1}^n \left\{ \sum_{t=1}^T e_{it} \right\}^2}{\sum_{i=1}^n \sum_{t=1}^T e_{it}^2} - 1 \right]^2, \quad (4.11)$$

где  $e_{it}$  – остатки в объединенной регрессии (3.6).

Если верна гипотеза  $H_0$  и выполняется предпосылка о нормальном распределении ошибок, статистика (4.11) имеет асимптотическое распределение Хи-квадрат с одной степенью свободы. Если нулевая гипотеза не отвергается, то нет оснований предпочесть модель со случайными эффектами обычной регрессии (4.5).

Временные эффекты учитываются в модели со случайными эффектами добавлением еще одного компонента в структуру ошибок:  $\varepsilon_{it} = u_i + w_t + v_{it}$  с необходимыми предпосылками относительно распределения  $w_t$ .

Таким образом, предлагается рассмотреть две различные спецификации модели панельных данных, позволяющие учесть как неоднородность данных по объектам наблюдения, так и по периодам времени. Модель (4.1) по сравнению с моделью (4.10) содержит больше параметров, подлежащих оцениванию, что ве-

дет к потере степеней свободы, а также усугубляет проблемы коллинеарности. Модель со случайными эффектами позволяет устранить эти недостатки, но требует введения дополнительного предположения о некоррелированности специфического для объекта наблюдения слагаемого ошибки с регрессорами. Нарушение этого предположения приведет к расчету несостоятельных оценок параметров модели. В пользу модели с фиксированными эффектами говорит также небольшое число объектов наблюдения и интерес к построению прогноза для каждого объекта наблюдения. В моделях с панельными данными использовать коэффициент детерминации для того, чтобы определить, какой метод оценивания лучше, нецелесообразно. Его можно применять только для сравнения моделей, отличающихся набором регрессоров и оцениваемых одним и тем же методом [3].

Подтвердить или опровергнуть априорные предположения относительно спецификации модели можно с помощью теста Хаусмана проверки гипотезы об ортогональности случайных эффектов и регрессоров [1, 2, 3]. Его подход основан на том, что если гипотеза об отсутствии корреляции верна, то оценки моделей с фиксированными и случайными эффектами являются состоятельными, но оценки модели со случайными эффектами эффективными. Поэтому при выполнении нулевой гипотезы между оценками нет систематического смещения. При альтернативной гипотезе состоятельны лишь оценки модели с фиксированными эффектами. Для проверки гипотезы используется статистика: [1]

$$W = [b_{FE} - b_{RE}]^T [Cov(b_{FE}) - Cov(b_{RE})]^{-1} [b_{FE} - b_{RE}], \quad (4.12)$$

где  $b_{FE}$  – оценки параметров модели с фиксированными эффектами;

$b_{RE}$  – оценки параметров модели со случайными эффектами;

$Cov(b_{FE})$ ,  $Cov(b_{RE})$  – оценки ковариационных матриц для параметров моделей с фиксированными и случайными эффектами.

При выполнении нулевой гипотезы статистика (4.12) асимптотически подчиняется закону распределения Хи-квадрат с  $k$  степенями свободы. Если нулевая

гипотеза не отвергается, то необходимо выбрать модель со случайными эффектами, в противном случае – модель с фиксированными эффектами.

Для оценивания коэффициентов в моделях панельных данных можно использовать различные статистические пакеты, реализующие процедуры стандартного регрессионного анализа: EViews, Statistica, Stadia, SPSS и другие. При этом предварительно нужно определенным образом преобразовать исходные данные, а после реализации стандартных методов регрессионного анализа в полученные результаты внести коррективы, учитывающие различие степеней свободы в моделях, построенных по панельным и по преобразованным данным [2]. К этим неудобствам добавляется необходимость проведения самостоятельных расчетов для проверки специфических гипотез, сопровождающих только модели панельных данных. Поэтому для обработки панельных данных целесообразно использовать статистический пакет Stata 6.0, в котором с помощью команд определенного формата можно найти оценки моделей панельных данных, а также реализовать процедуры проверки различных гипотез [1].

#### **4.2 Моделирование числа трудоустроенных безработных с учетом пространственно-временной неоднородности данных**

Как известно, безработные, различающиеся по полу, возрасту, уровню образования и другим характеристикам, имеют разную продолжительность поиска работы [12]. Учесть особенности трудоустройства выделенных групп безработных невозможно только в рамках моделей, основанных на временных рядах или пространственных данных, поэтому предлагается моделирование трудоустройства безработных осуществить с помощью методов анализа панельных данных.

Ставится задача построить модель, описывающую зависимость числа трудоустроенных безработных при содействии Центра занятости населения  $y_i$  от объясняющих переменных:

$x_{1,t}$  – число безработных, зарегистрированных в Центре занятости населения в целях поиска работы в квартале  $t$ , чел;

$x_{2,t}$  – число безработных, состоявших на учете в Центре занятости населения в целях поиска работы на начало квартала  $t$ , чел.;

$x_{3,t}$  – число безработных, получавших пособие по безработице в квартале  $t$ , чел.;

$x_{4,t}$  – число безработных, состоявших на учете в Центре занятости населения, которым предлагались вакансии в квартале  $t$ , чел.

Объектами наблюдения являются категории безработных, полученные в результате группировки безработных по качественным признакам, оказывающим существенное влияние на процесс трудоустройства – пол, возраст, уровень профессионального образования. Каждая категория безработных характеризуется поквартальными данными за период с 2001 по 2006 годы по вышеописанным показателям.

На основе панельных данных (всего объектов 24), с помощью статистического пакета Stata 6.0, найдены оценки модели (4.2) с фиксированными эффектами и модели (4.11) со случайными эффектами. Результаты проверки гипотез о значимости параметров моделей представлены в таблице 4.1.

Таблица 4.1 – Результаты проверки гипотез о значимости параметров моделей с фиксированными и случайными эффектами для безработных, различающихся по полу, возрасту и уровню профессионального образования

Проверяемая гипотеза	Наблюдаемое значение статистики	Значимость нулевой гипотезы
Модель с фиксированными эффектами		
$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$	$F(4,548) = 441,87$	$p = 0,0000$
$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_{24}$	$F(23,548) = 4,16$	$p = 0,0000$
Модель со случайными эффектами		
$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$	$Wald(4) = 4244,87$	$p = 0,0000$
$H_0 : \sigma_u^2 = 0$	$LM(1) = 31,01$	$p = 0,0000$

По результатам проверки гипотез можно сделать вывод, что обе модели значимы, оправданным является использование в обоих случаях панельной структуры данных, что подтверждает предположение о неоднородности групп безработных, различающихся по полу, возрасту и уровню профессионального образования.

Результаты оценивания коэффициентов при объясняющих переменных и проверки их значимости представлены в таблице 4.2.

Таблица 4.2 – Оценки коэффициентов моделей с фиксированными и случайными эффектами для безработных, различающихся по полу, возрасту и уровню профессионального образования

Коэффициент	Оценка	Стандартная ошибка	Значение $t$ -статистики	Значимость	Доверительный интервал	
					Верхняя граница	Нижняя граница
Модель с фиксированными эффектами						
$\beta_1$	0,3444	0,0386	8,93	0,0000	0,2686	0,4201
$\beta_2$	0,3215	0,0427	7,53	0,0000	0,2376	0,4054
$\beta_3$	-0,6533	0,0341	-19,17	0,0000	-0,7202	-0,5863
$\beta_4$	0,4906	0,0432	11,36	0,0000	0,4058	0,5754
Модель со случайными эффектами						
$\beta_1$	0,3641	0,0359	10,15	0,0000	0,2938	0,4345
$\beta_2$	0,2266	0,0393	5,76	0,0000	0,1495	0,3037
$\beta_3$	-0,6922	0,0255	-27,14	0,0000	-0,7422	-0,6423
$\beta_4$	0,5216	0,0398	13,10	0,0000	0,4435	0,5996

Все коэффициенты при объясняющих переменных в моделях значимы. Оценки коэффициентов детерминации для моделей с фиксированными и случайными эффектами составили соответственно 0,88 и 0,90. По результатам проверки гипотез можно отметить, что обе модели обладают достаточно хорошими статистическими свойствами, однако, тест Хаусмана ( $W(4) = 43,57$ ,  $p = 0,0000$ ) подтвердил априорные доводы в пользу модели с фиксированными эффектами, оценка которой имеет вид:

$$\hat{y}_{it} = a_i + \underset{(0,04)}{0,34}x_{1,it} + \underset{(0,04)}{0,32}x_{2,it} - \underset{(0,02)}{0,65}x_{3,it} + \underset{(0,04)}{0,49}x_{4,it}, \quad i = \overline{1, \dots, 24}, \quad t = \overline{1, \dots, 24}. \quad (4.13)$$

Согласно модели (4.13), при увеличении числа безработных, зарегистрированных в Центре занятости населения в целях поиска работы в квартале  $t$  на 100 человек, число трудоустроенных Центром занятости безработных в данном квартале увеличивается на 34 человек. Такому соотношению числа зарегистрированных и числа трудоустроенных безработных можно дать следующее объяснение. Во-первых, не все безработные, зарегистрированные в текущем квартале, будут трудоустроены в этом квартале. Часть из них будет трудоустроена в последующие периоды времени. Об этом свидетельствует оценка коэффициента при второй объясняющей переменной, согласно которой при увеличении числа безработных, состоявших на учете в Центре занятости населения в целях поиска работы на начало квартала  $t$  на 100 человек, число трудоустроенных Центром занятости безработных в данном квартале увеличивается на 32 человек. Во-вторых, не все зарегистрированные безработные трудоустраиваются при содействии Центра занятости, многие из них трудоустраиваются самостоятельно или снимаются с учета по другим причинам: личное заявление, уход на пенсию, неявка в Центр занятости, перемена места жительства и др.

Отрицательный коэффициент при объясняющей переменной  $X_3$  свидетельствует о том, что чем больше число безработных, получавших пособие по безработице в квартале  $t$ , тем меньше число трудоустроенных безработных в данном квартале. Согласно пункту 1 статьи 31 Закона РФ «О занятости населения в Российской Федерации» пособие по безработице выплачивается гражданам, признанным в установленном порядке безработными. При этом, согласно «Порядку регистрации безработных граждан», не все безработные, зарегистрированные в целях поиска подходящей работы, признаются безработными, а только те, кто не трудоустроен в течение 10 дней со дня их регистрации в целях поиска работы. Таким образом, чем больше число безработных, получающих пособие по безработице, тем больше безработных, не пользующихся спросом на рын-

ке труда и как следствие тем меньше число трудоустроенных безработных. С другой стороны, выявленная закономерность косвенно свидетельствует о том, что выплата пособия по безработице является стимулом к более длительному пребыванию в статусе безработного.

Коэффициент при объясняющей переменной  $X_4$  можно интерпретировать следующим образом: из 100 безработных, которым предлагались вакансии в квартале  $t$ , трудоустроились в этом квартале только 49 человек. Это отражает, с одной стороны, несоответствие предлагаемой работы требованиям безработных, что проявляется в большом количестве отклоненных безработными вакансий, с другой стороны, несоответствие безработных требованиям работодателей, что проявляется в большом количестве отказов работодателей в трудоустройстве безработных.

Оценки индивидуальных эффектов модели (4.13):

$$\begin{aligned}
 a_1 = -7,42; & \quad a_5 = -11,12; & \quad a_9 = -4,68; & \quad a_{13} = -7,73; & \quad a_{17} = -2,61; & \quad a_{21} = -3,42; \\
 a_2 = -18,45; & \quad a_6 = -13,05; & \quad a_{10} = -21,14; & \quad a_{14} = -8,33; & \quad a_{18} = -4,44; & \quad a_{22} = -6,41; \\
 a_3 = -11,14; & \quad a_7 = -6,01; & \quad a_{11} = -5,88; & \quad a_{15} = -4,33; & \quad a_{19} = 8,99; & \quad a_{23} = -12,22; \\
 a_4 = -16,28; & \quad a_8 = -21,72; & \quad a_{12} = -20,46; & \quad a_{16} = -10,73; & \quad a_{20} = 6,75; & \quad a_{24} = -23,87.
 \end{aligned}$$

Предполагая наличие в модели с фиксированными эффектами временных эффектов, найдена оценка двунаправленной модели. Результаты проверки различных гипотез о значимости параметров модели представлены в таблице 4.3.

Все нулевые гипотезы отвергаются, следовательно, двунаправленная модель значима. Оценка коэффициента детерминации составила 0,94.

Таблица 4.3 – Результаты проверки гипотез о значимости параметров двунаправленной модели с фиксированными эффектами для безработных, различающихся по полу, возрасту и уровню профессионального образования

Проверяемая гипотеза	Наблюдаемое значение статистики	Критическое значение статистики
$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_{24},$ $\gamma_1 = \gamma_2 = \dots = \gamma_{23} = 0$	$F_{набл} = 8,26$	$F_{кр} (0,05;46;525) = 1,39$
$H_0 : \gamma_1 = \gamma_2 = \dots = \gamma_{23} = 0$	$F_{набл} = 10,61$	$F_{кр} (0,05;23;525) = 1,55$
$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_{24}$	$F_{набл} = 7,00$	$F_{кр} (0,05;23;525) = 1,55$
$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$	$F_{набл} = 413,93$	$F_{кр} (0,05;4;525) = 2,39$



Оценки коэффициентов при объясняющих переменных и результаты проверки гипотез об их значимости представлены в таблице 4.4.

Таблица 4.4 – Оценки коэффициентов двунаправленной модели с фиксированными эффектами для безработных, различающихся по полу, возрасту и уровню профессионального образования

Коэффициент	Оценка	Стандартная ошибка	Наблюдаемое значение $t$ -статистики	Доверительный интервал	
				Верхняя граница	Нижняя граница
$\beta_1$	0,3175	0,0422	7,53	0,2348	0,4002
$\beta_2$	0,4094	0,0426	9,61	0,3259	0,4929
$\beta_3$	-0,6936	0,0348	-19,95	-0,7618	-0,6254
$\beta_4$	0,4964	0,0469	10,59	0,4045	0,5883

Критическое значение  $t$ -статистики составляет:  $t_{кр}(0,05;525) = 1,96$ . Так как  $|t_{набл}| > t_{кр}$  для всех коэффициентов модели, то все они являются значимыми. Оценка двунаправленной модели имеет вид:

$$\hat{y}_{it} = a_i + c_t + 0,32x_{1,it} + 0,41x_{2,it} - 0,69x_{3,it} + 0,50x_{4,it}. \quad (4.14)$$

(0,04)            (0,04)            (0,03)            (0,05)

Сравнивая оценки параметров при объясняющих переменных моделей (4.13) и (4.14) можно отметить, что влияние последних трех показателей на число трудоустроенных безработных незначительно усилилось, а влияние показателя  $X_1$  ослабло. В остальном интерпретация коэффициентов моделей совпадает.

По формуле (4.3) найдены оценки индивидуальных эффектов модели (4.14):

$$\begin{aligned} a_1 = -6,75; & \quad a_5 = -12,33; \quad a_9 = -4,85; \quad a_{13} = -6,90; \quad a_{17} = -2,71; \quad a_{21} = -1,73; \\ a_2 = -16,61; & \quad a_6 = -15,14; \quad a_{10} = -23,80; \quad a_{14} = -7,93; \quad a_{18} = -4,96; \quad a_{22} = -6,36; \\ a_3 = -11,88; & \quad a_7 = -5,61; \quad a_{11} = -6,40; \quad a_{15} = -4,33; \quad a_{19} = 12,84; \quad a_{23} = -12,73; \\ a_4 = -18,60; & \quad a_8 = -20,75; \quad a_{12} = -23,42; \quad a_{16} = -11,45; \quad a_{20} = 9,18; \quad a_{24} = -26,09. \end{aligned}$$

По формуле (4.10) найдены оценки временных эффектов модели (4.14):

$$\begin{aligned}
c_1 &= -22,06; & c_5 &= 8,79; & c_9 &= -3,92; & c_{13} &= -6,02; & c_{17} &= 1,73; & c_{21} &= 2,39; \\
c_2 &= -5,45; & c_6 &= -1,02; & c_{10} &= 18,15; & c_{14} &= -1,72; & c_{18} &= 17,69; & c_{22} &= 4,69; \\
c_3 &= -5,04; & c_7 &= -2,21; & c_{11} &= -3,06; & c_{15} &= 0,94; & c_{19} &= 5,74; & c_{23} &= 0,97; \\
c_4 &= 0,20; & c_8 &= -8,26; & c_{12} &= 1,13; & c_{16} &= -4,08; & c_{20} &= -5,12; & c_{24} &= 3,53.
\end{aligned}$$

Временные эффекты отражают особенности трудоустройства безработных в разные периоды времени. Диаграмма значений временных эффектов представлена на рисунке 4.1.

Анализируя график, можно отметить, что значения временных эффектов в течение рассматриваемых периодов существенно колебались. Наибольшее число трудоустроенных безработных приходилось, как правило, на второй квартал года, а наименьшее – на четвертый. При прочих равных условиях наблюдается тенденция к увеличению числа безработных, трудоустроенных при содействии Центра занятости.

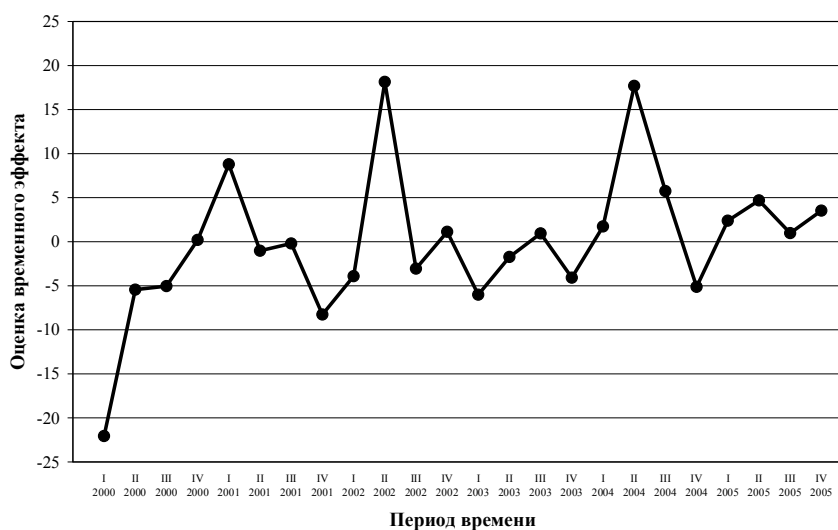


Рисунок 4.1 – Оценки временных эффектов

Ранжирование категорий безработных предлагается проводить по значениям индивидуальных эффектов модели панельных данных. Чем больше значение индивидуального эффекта, тем при прочих равных условиях большее число безработных данной категории трудоустраивается при содействии Центра занятости и, следовательно, тем более востребована данная категория безработных

на рынке труда. В таблице 4.5 все группы безработных, различающихся по полу, возрасту и уровню профессионального образования, упорядочены по уменьшению значений индивидуальных эффектов, входящих в модель (4.14).

Таблица 4.5 – Результаты ранжирования категорий безработных, различающихся по полу, возрасту и уровню профессионального образования

№ п /п	Оценка индивидуально-го эффекта	Категория безработных		
		Пол	Возраст	Образование
1	12,84	Мужской	До 29	Нет
2	9,18	Женский	До 29	Нет
3	-1,73	Мужской	От 30 до 49	нет
4	-2,71	Мужской	50 и старше	Начальное
5	-4,33	Мужской	От 30 до 49	Начальное
6	-4,85	Мужской	От 30 до 49	Среднее
7	-4,96	Женский	50 и старше	Начальное
8	-5,62	Мужской	До 29	Среднее
9	-6,36	Женский	От 30 до 49	Нет
10	-6,40	Мужской	50 и старше	Среднее
11	-6,75	Мужской	До 29	Высшее
12	-6,90	Мужской	До 29	Начальное
13	-7,93	Женский	До 29	Начальное
14	-11,45	Женский	От 30 до 49	Начальное
15	-11,88	Мужской	От 30 до 49	Высшее
16	-12,33	Мужской	50 и старше	Высшее
17	-12,73	Мужской	50 и старше	Нет
18	-15,14	Женский	50 и старше	Высшее
19	-16,61	Женский	До 29	Высшее
20	-18,60	Женский	От 30 до 49	Высшее
21	-20,75	Женский	До 29	Среднее
22	-23,42	Женский	50 и старше	Среднее
23	-23,80	Женский	От 30 до 49	Среднее
24	-26,06	Женский	50 и старше	Нет

По данным таблицы 4.5 можно сделать вывод, что наибольшим спросом на регистрируемом рынке труда города Оренбурга пользуются безработные мужчины и женщины в основном молодого и среднего возраста без профессионального образования и с начальным профессиональным образованием, наименее востребованы безработные женщины со средним и высшим профессиональным образованием. При этом женщинам труднее найти работу, чем мужчинам, а безработным пожилого возраста – труднее, чем безработным молодого и среднего возраста.

Благодаря использованию моделей с фиксированными эффектами, содержащими индивидуальные и временные эффекты, при оценке числа трудоустроенных безработных удалось учесть не только влияние объясняющих переменных, но и специфические характеристики различных категорий безработных, а также особенности трудоустройства безработных в разные периоды времени. На основе полученных моделей можно проводить имитацию различных сценариев развития ситуации на регистрируемом рынке труда. Оценки всех параметров моделей имеют большое практическое значение:

- оценки коэффициентов при объясняющих переменных характеризуют эффективность функционирования Центра занятости;
- оценки временных эффектов отражают динамику трудоустройства безработных;
- оценки индивидуальных эффектов позволяют судить о востребованности различных категорий безработных на рынке труда.

#### **4.3 Модели оценки стоимости квартир на вторичном рынке жилья с учетом пространственно-временной неоднородности**

Стоимость жилья формируется в результате взаимодействия многообразных факторов, необходимо выявить эти факторы, установить существующие между ними взаимосвязи и конкретную форму зависимости. Анализ данных риэлторских фирм, а также информация, полученная от специалистов в области недвижимости, позволили выделить следующие основные факторы, влияющие на стоимость квартиры ( $y$ , тыс. руб.) в городе [13]:

$x_1$  – общая площадь,  $m^2$ ;

$x_2$  – жилая площадь,  $m^2$ ;

$x_3$  – площадь кухни,  $m^2$ ;

$x_4$  – вспомогательная площадь,  $m^2$ ;

$x_5$  – этаж, на котором расположена квартира;

$x_6$  – наличие/отсутствие балкона;

$x_7$  – наличие/отсутствие телефона;

$x_8$  – тип дома;

$x_9$  – материал стен;

$x_{10}$  – месторасположение дома, в котором находится квартира;

$x_{11}$  – период продаж.

Как видно, если первые четыре показателя являются количественными и легко поддаются измерению, то другие показатели предлагается представить в виде «фиктивных» переменных.

Информационной базой послужили данные о ценах и свойствах объектов продажи на вторичном рынке жилья города Оренбурга (839 квартир, выставленных на продажу). Отобранный массив данных о ценах и свойствах квартир, предлагаемых на продажу на вторичном рынке жилья г. Оренбурга, подвергался развернутому статистическому анализу, целью которого является отбор подмножества показателей, влияние которых на цену квартиры наиболее существенно.

На основании экономического анализа имеющейся информации пришли к выводу о том, что стоимость квартиры меняется на определенный процент, в зависимости от типа дома, наличия балкона, этажа и района, следовательно, параметры модели должны быть коэффициентами эластичности, поэтому для моделирования стоимости жилья на вторичном рынке г. Оренбурга целесообразно использовать мультипликативную (степенную) функцию, как известно, переход к логарифмической форме позволяет привести такую модель к линейному виду. Построение регрессионных моделей стоимости квартир в зависимости от факторов, влияющих на нее, осуществлялось в ППП Statistica. Результаты расчетов представлены в таблице 4.6.

Коэффициент детерминации свидетельствует о том, что 63,9% - 73,5% вариации цены объясняется вошедшими в модели показателями. Средняя ошибка аппроксимации характеризует адекватность моделей. Эластичность цены однокомнатных квартир по жилой площади ( $x_2$ ) изменилась с 0,26% в 2004г. до 0,43%

в 2005г., а в 2006г. достигла уровня 2004г. Цена увеличилась в среднем на 6,0% - 7,0%, если квартира улучшенной планировки ( $x_8$ ). Квартира, проданная в первом полугодии, на 16% – 28% дешевле, чем аналогичная квартира, проданная во втором полугодии ( $x_{11}$ ).

Таблица 4.6 – Результаты регрессионного анализа стоимости квартир на вторичном рынке жилья г. Оренбурга

Тип квартиры	Годы	Результаты регрессионного анализа
однокомнатные	2004	$\hat{y} = e^{4,04} \cdot x_2^{0,26} \cdot (e^{x_8})^{0,06} \cdot x_{10}^{-0,05} \cdot (e^{x_{11}})^{-0,16}$ $\hat{R}^2 = 0,679$ ; A=5,4%; $F_{фак} = 51,82$ ; $F_{табл} (0,05;4;98)=2,46$
	2005	$\hat{y} = e^{4,934} \cdot x_2^{0,43} \cdot (e^{x_8})^{0,063} \cdot x_{10}^{-0,33} \cdot (e^{x_{11}})^{-0,28}$ $\hat{R}^2 = 0,639$ ; A=5,3%; $F_{фак} = 35,40$ ; $F_{табл} (0,05;4;59)=2,76$
	2006	$\hat{y} = e^{5,33} \cdot x_2^{0,20} \cdot (e^{x_8})^{0,07} \cdot x_{10}^{-0,35}$ $\hat{R}^2 = 0,735$ ; A=5,6%; $F_{фак} = 81,13$ ; $F_{табл} (0,05;3;118)=2,45$
двухкомнатные	2004	$\hat{y} = e^{4,45} \cdot x_2^{0,34} \cdot x_3^{0,14} \cdot (e^{x_5})^{0,12} \cdot (e^{x_7})^{0,05} \cdot x_{10}^{-0,07} \cdot (e^{x_{11}})^{-0,13}$ $\hat{R}^2 = 0,709$ ; A=4,6%; $F_{фак} = 49,10$ ; $F_{табл} (0,05;6;116)=2,18$
	2005	$\hat{y} = e^{4,539} \cdot x_2^{0,36} \cdot x_3^{0,15} \cdot (e^{x_5})^{0,11} \cdot (e^{x_7})^{0,06} \cdot x_{10}^{-0,04} \cdot (e^{x_{11}})^{-0,22}$ $\hat{R}^2 = 0,719$ ; A=5,8%; $F_{фак} = 23,45$ ; $F_{табл} (0,05;6;55)=2,27$
	2006	$\hat{y} = e^{4,48} \cdot x_2^{0,36} \cdot x_3^{0,20} \cdot (e^{x_5})^{0,12} \cdot (e^{x_7})^{0,07} \cdot x_{10}^{-0,44}$ $\hat{R}^2 = 0,618$ ; A=4,8%; $F_{фак} = 22,00$ ; $F_{табл} (0,05;5;68)=2,35$
трехкомнатные	2004	$\hat{y} = e^{5,49} \cdot x_2^{0,43} \cdot x_3^{0,09} \cdot (e^{x_5})^{0,08} \cdot (e^{x_6})^{0,09} \cdot x_{10}^{-0,04}$ $\hat{R}^2 = 0,985$ ; A=8,6%; $F_{фак} = 1221,4$ ; $F_{табл} (0,05;5;93)=2,31$
	2005	$\hat{y} = e^{5,67} \cdot x_2^{0,32} \cdot x_3^{0,20} \cdot (e^{x_5})^{0,18} \cdot (e^{x_6})^{0,10} \cdot x_{10}^{-0,59} \cdot (e^{x_{11}})^{-0,12}$ $\hat{R}^2 = 0,819$ ; A=4,3%; $F_{фак} = 40,72$ ; $F_{табл} (0,05;6;44)=2,42$
	2006	$\hat{y} = e^{5,12} \cdot x_2^{0,31} \cdot x_3^{0,16} \cdot (e^{x_5})^{0,03} \cdot (e^{x_7})^{0,07} \cdot x_{10}^{-0,89}$ $\hat{R}^2 = 0,947$ ; A=7,4%; $F_{фак} = 189,4$ ; $F_{табл} (0,05;5;53)=2,39$

За анализируемый период увеличивается влияние на стоимость двухкомнатных квартир фактора – жилая площадь ( $x_2$ ) с 0,34% до 0,36%. Увеличение площади кухни ( $x_3$ ) на 1% влечет повышение стоимости жилья в среднем на 0,14%–0,20%, наличие телефона увеличивает цену квартиры в среднем на 5,0%–

7,0% ( $x_7$ ). Квартира, купленная в первом полугодии ( $x_{11}$ ), дешевле на 13,0% – 22,0%, если квартира находится на одном из промежуточных этажей дома, то она дороже в среднем 11,0% - 12,0% ( $x_5$ ).

Для совокупности трехкомнатных квартир за анализируемый период увеличилось влияние фактора – площадь кухни. Эластичность цены по жилой площади ( $x_2$ ) изменилась с 0,43% до 0,31%. Квартира, проданная в первом полугодии, стоит на 12,0% дешевле аналогичной, но проданной во втором полугодии ( $x_{11}$ ), наличие балкона ( $x_6$ ) увеличивает цену квартиры в среднем на 7,0%-9,0%.

Сравнение полученных моделей показывает, что с ростом количества комнат увеличивается влияние коэффициента зонирования. Это является, на наш взгляд, отражением того факта, что чем более просторной (=престижной= дорогой) является покупаемая квартира, тем в меньшей степени покупатель согласен купить ее в непрестижном районе города и, следовательно, тем ниже та пороговая сумма, которую он готов за нее заплатить. В результате квартиры с большим количеством комнат в расчете на 1 кв. м общей площади в непрестижных и очень отдаленных районах оказываются наиболее дешевыми.

Этаж, на котором находится квартира, существенно влияет на цену многокомнатного жилья.

Фактор  $x_9$  – качество дома, который определяется через характеристику материала его стен – не вошел в модели. Однако это не означает, что квартиры в кирпичном и панельном домах имеют одинаковую стоимость. Можно говорить, что при заданном уровне значимости материал стен не оказывает существенного влияния на цену типового жилья г. Оренбурга.

Таким образом, наиболее существенными факторами, влияющими на стоимость квартиры, являются район города, характеристики размера квартиры, этаж, на котором находится квартира, тип дома, период продажи.

Построенные регрессионные модели оценки стоимости жилья по временным сечениям (по годам) свидетельствуют о том, что существует изменение влияния основных факторов на стоимость жилья не только в пространстве, но и во

времени, то есть происходит смещение параметров регрессионной модели. Поэтому целесообразным представляется применять различные модели регрессии для панельных данных.

Объектами наблюдения при формировании панели рассматривались одно-, двух-, трех- и четырехкомнатные квартиры. Тип квартиры характеризуется поквартальными данными за период с 2004 по I полугодие 2006 годы по следующим показателям:

$y_{it}$  – средняя цена квартиры для объекта  $i$  в момент времени  $t$ , тыс. руб;

$x_{2,it}$  – средняя площадь (жилая) для объекта  $i$  в момент времени  $t$ ,  $m^2$ ;

$x_{3,it}$  – средняя площадь кухни для объекта  $i$  в момент времени  $t$ ,  $m^2$ ;

$x_{4,it}$  – средняя площадь (дополнительная), которая представляет собой разницу между общей площадью, жилой площадью и площадью кухни, для объекта  $i$  в момент времени  $t$ ,  $m^2$ ,  $i = 1, 2, 3, 4$ ;  $t = 1, 2, \dots, 10$ . В данном случае, как и в предыдущем, выбрана степенная функция для модели оценки стоимости квартир.

С помощью статистического пакета Stata оценены модели с фиксированными и случайными эффектами. В таблице 4.7 представлены результаты исследования моделей.

По результатам проверки гипотез можно сделать вывод, что обе модели значимы, оправданным является использование в обоих случаях панельной структуры данных, что подтверждает предположение о неоднородности объектов продажи, различающихся по типу квартир. Была также построена двунаправленная модель с фиксированными эффектами, в рамках которой доказана значимость временного эффекта, однако индивидуальный эффект оказался незначимым.



Таблица 4.7 – Результаты оценки моделей с фиксированными и случайными эффектами

Модель с фиксированными эффектами		
$\hat{\ln}(y_{it}) = 2,194 + 0,886 \ln(x_{2,it}) + 0,607 \ln(x_{3,it}), \hat{R}^2 = 0,9719$ <small>(0,945)                      (0,013)                      (0,053)</small>		
Проверяемая гипотеза	Наблюдаемое значение статистики	Значимость нулевой гипотезы
$H_0 : \beta_1 = \beta_2 = 0$	F(2;34)=659,82	$p = 0,0000$
$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$	F(3,34) = 59,66	$p = 0,0000$
Модель со случайными эффектами		
$\hat{\ln}(y_{it}) = 2,938 + 0,676 \ln(x_{2,it}) + 0,632 \ln(x_{3,it}), \hat{R}^2 = 0,9980$ <small>(0,792)                      (0,275)                      (0,084)</small>		
Проверяемая гипотеза	Наблюдаемое значение статистики	Значимость нулевой гипотезы
$H_0 : \beta_1 = \beta_2 = 0$	<i>Wald</i> (2) = 63,17	$p = 0,0000$
$H_0 : \sigma_u^2 = 0$	<i>LM</i> (1) = 6,09	$p = 0,0000$

Все коэффициенты при объясняющих переменных в моделях значимы (таблица 4.7). Оценки коэффициентов детерминации для моделей с фиксированными и случайными эффектами составили соответственно 0,97 и 0,99. По результатам проверки гипотез можно отметить, что обе модели обладают достаточно хорошими статистическими свойствами, однако, тест Хаусмана ( $W(2) = 0,20$ ,  $p = 0,6514$ ) подтвердил целесообразность применения моделей со случайными эффектами, так как анализируемые объекты попали в выборку из большой совокупности, а эффекты, связанные с типом дома, временем продажи квартиры, вероятно, более существенны, нежели индивидуальные особенности объектов.

Таким образом, для моделирования стоимости жилья на вторичном рынке г. Оренбурга предлагается следующая регрессионная модель:

$$\ln(\hat{y}_{it}) = 2,938 + 0,676 \ln(x_{2,it}) + 0,632 \ln(x_{3,it}), \quad i = \overline{1, \dots, 4}, \quad t = \overline{1, \dots, 9}.$$

По полученной модели можно сделать вывод о том, что изменение стоимости жилья объясняется на 99,8% вариацией жилой площадью и площадью кухни.

Эластичность цены квартир по жилой площади составляет 0,676%, по площади кухни 0,632%.

Несомненно, что истинная цена сделки будет определяться, безусловно, в результате проведения торгов между покупателем и продавцом жилой недвижимости, тем не менее, полученная на основе эконометрических моделей информация окажет помощь продавцам и покупателям в определении цены предложения. Использование различных методов позволяет сделать наиболее точную, максимально приближенную к реальности оценку стоимости жилья.

При построении классической модели регрессии предполагалось, что исходные данные являются однородными по всей исследуемой области. В результате получается модель, описывающая изменения, характерные в среднем для всей исследуемой области. Данный тип моделей достаточно хорошо изучены как за рубежом, так и в нашей стране, и характеризуются наглядностью интерпретации получаемых результатов, а также доступностью исходных данных.

На практике зачастую исходные данные являются пространственно зависимыми и измеряются на достаточно большой территории неоднородной по составу, например, при моделировании ценообразования недвижимости для рынков различных городов. В любой модели цена жилья зависит от набора количественных и качественных переменных, таких как возраст дома, площадь и т.п., социальных признаков, а также от характеристик местоположения, например удаленность от метро, мест отдыха и т.п. Поэтому возникает проблема адекватного учета местоположения объекта на территории исследуемой области, а возможных корреляций цен близкорасположенных объектов.

Учет территориальной неоднородности рассмотрен выше в контексте общей проблемы построения регрессионных моделей по неоднородным данным. Для анализа территориальных данных, регрессионную неоднородность учитывают разделением исследуемой области на однородные группы. Для этого в модель вводят фиктивные переменные, характеризующие территориальную принадлежность объектов. Однако, зонирование исследуемой территории зависит от

выбора масштаба, оказывающего влияние на результаты эконометрического моделирования и зависящего от определения границ.

Различают районы (группы) с четкими и нечеткими границами. При разделении исследуемой территории на зоны с четкими границами предполагают, что они определяют различия в закономерностях социально-экономических явлений. При произвольном определении зон часто изменение границ приводит к совершенно противоположным результатам в оценках коэффициентов регрессионной модели и ошибочной интерпретации, поэтому в процессе зонирования используют нечеткие границы. Нечеткие границы означают, что каждый объект с той или иной вероятностью имеет возможность пасть в любую зону. Вероятность зависит от близости объекта к центру зоны и с увеличением расстояния уменьшается.

Кроме определения границ возникает проблема влияния масштаба выделяемых областей на значимость коэффициентов модели. Возможна ситуация, когда зоны, получают различные или даже совершенно противоположные результаты в оценках и значимости параметров модели, что затрудняет выбор модели и трактовку результатов.

Для избежания этих трудностей, применяют метод географически взвешенной регрессии, которую можно рассматривать как некоторое обобщение фиктивных переменных, позволяющей частично устранить перечисленные проблемы и получить модель с непрерывно меняющейся структурой. Исследование регрессионных моделей, использующих представление данных с учетом их координат, осуществлено в работах В.А. Балаша, О.С. Балаш, А.В. Харламова, А.Д. Лунькова и др. В представленных работах методом географически взвешенной регрессии проведено исследование изменений коэффициентов модели стоимости жилой недвижимости на территории г. Саратова и определены локальные особенности ценообразующих факторов.

При географическом подходе считается, что модель не является постоянной для всего исследуемого региона, а меняется в зависимости от местных условий, и ее коэффициенты являются функциями координат.

В работе Харламова А.В. модель ценообразования на рынке жилой недвижимости предлагается представить в виде [7, 8, 9]:

$$y_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{1i} + \beta_2(u_i, v_i)x_{2i} + \alpha_1(u_i, v_i)z_{1i} + \alpha_2(u_i, v_i)z_{2i} + \alpha_3(u_i, v_i)z_{3i} + \alpha_4(u_i, v_i)z_{4i} + \alpha_5(u_i, v_i)z_{5i} + \varepsilon_i,$$

где  $y_i$  – цена  $i$ -го объекта жилой недвижимости, тыс. руб.;

$x_1$  – жилая площадь, м<sup>2</sup>;

$x_2$  – вспомогательная площадь, м<sup>2</sup>;

$z_1$  – этаж, на котором расположен объект жилой недвижимости;

$z_2$  – наличие/отсутствие балкона;

$z_3$  – наличие/отсутствие телефона;

$z_4$  – тип дома;

$z_5$  – материал стен;

$(u_i, v_i)$  – координаты  $i$ -го объекта жилой недвижимости.

Для оценивания параметров предлагаемой модели ценообразования, необходимо к исходным данным, характеризующим объекты жилой недвижимости, добавить условные координаты объектов. После получения дополнительной информации можно приступить к расчету оценок параметров географически взвешенной регрессии. В результате оценивания получим матрицы оценок параметров для каждого регрессора. Так, например, в таблице 4.8 представлена матрица оценок коэффициента регрессии при переменной «жилая площадь» в зависимости от координат.

Как видно, из таблицы в центральной части города ( $u = 60, v = 32$ ) располагается квадрат с самыми дорогими квартирами (около 30 тыс. руб. за м<sup>2</sup>). Вокруг него стоимость квадратного метра жилой площади превышает 20 тыс. руб. Четко выделяются окраины города, где цена жилой площади порядка 10 тыс. руб. Прослеживается тенденция к снижению стоимости квартир от центра в направлении «левого верхнего» и «левого нижнего» углов таблицы.

Аналогично могут быть оценены и проанализированы матрицы оценок параметров различных характеристик объектов жилой недвижимости.

Таблица 4.8 – Матрица коэффициента регрессии при переменной «жилая площадь» в зависимости от координат<sup>1</sup>

		u											
		53	54	55	56	57	58	59	60	61	62	63	64
v	36		9,1	8,6		12,2	11,2	8,9	8,6				
	35		9,7	10,9	11,9	12,9	9,8			12,4			16,7
	34				11,9	11,5	11,8	14,7		16,6	15,2	16,6	16,8
	33				11,2		12	15,7	20,3	19,1		15,7	
	32					14,1	16,6	23,2	28,4	21,4	14,1	14,1	
	31				13	15,7	17,2	19,7	22,4	17,4	11,6	13,5	
	30	10,3	12,9	14,2	13,6	14,7	16,4	18,6					
	29	9,6	11	13,5	13,7	14	12,7						
	28	9,8	10	10,3									
	27		9,8										

Таким образом, географический подход позволяет выявить специфические особенности ценообразования жилой недвижимости, присущие отдельным районам города и аппроксимировать специфику ценообразования в данном местоположении.

#### 4.4 Построение интегрального индикатора, характеризующего уровень демографической безопасности с учетом неоднородности данных во времени

Существующее смещение рейтинговых позиций муниципальных образований по уровню демографической безопасности в динамике свидетельствует о временной неоднородности, что обусловило необходимость построения интегрального индикатора демографической безопасности, учитывающего одновременно фактор времени и неоднородность в пространстве. В этом случае исходные данные предлагается представлять в виде панели, содержащей сведения об

<sup>1</sup> Расчетные данные взяты из работы Харламова А.В. Исследование динамики цен на жилую недвижимость методом географически взвешенной регрессии //Известия Саратовского университета.-2011.- Том 11, Выпуск 2.- С.25-29

одном и том же множестве объектов за ряд последовательных периодов времени [1].

Ставится задача построения интегрального индикатора, характеризующего уровень демографической безопасности, по заданному апостериорному набору частных критериев этого интегрального свойства  $x_1, x_2, \dots, x_k$ , учитывающих неоднородность объектов во времени.

Будем рассматривать уровень демографической безопасности региона как сводную характеристику  $f(\bar{X})$  – определяемую поддающимися учету и измерению признаками  $\bar{X} = (x_1, x_2, \dots, x_k)$ . Интуитивное экспертное восприятие сводной характеристики будем рассматривать в виде динамической модели латентного показателя:

$$\tilde{y}(t) \equiv f_t(X, \Theta) = \Theta_1 x_{1,t} + \Theta_2 x_{2,t} + \dots + \Theta_k x_{k,t}. \quad (4.15)$$

Объектами наблюдения являются административно-территориальные единицы Оренбургской области, каждый из которых характеризуется данными за период 2002-2005 годы по апостериорному набору из 23 показателей.

$x_{1.1.1,t}$  - общий коэффициент рождаемости в момент времени  $t$ ;

$x_{1.1.2,t}$  - общий коэффициент смертности в момент времени  $t$ ;

$x_{1.1.3,t}$  - коэффициент младенческой смертности в момент времени  $t$ ;

$x_{1.1.4,t}$  - уровень брачности населения в момент времени  $t$ ;

$x_{1.1.5,t}$  - уровень разводимости населения в момент времени  $t$ ;

$x_{1.1.6,t}$  - коэффициент миграционного прироста в момент времени  $t$ ;

$x_{1.2.1,t}$  - удельный вес населения, в трудоспособном возрасте в момент времени  $t$ ;

$x_{1.2.3,t}$  - соотношение мужчин и женщин в момент времени  $t$ ;

$x_{1.3.1,t}$  - показатель общей заболеваемости в момент времени  $t$ ;

$x_{1.3.6,t}$  - болезни системы кровообращения в момент времени  $t$ ;

$x_{1.3.12,t}$  - обеспеченность населения больничными койками в момент времени  $t$ ;

$x_{2.2,t}$  - соотношение среднего размера назначенных месячных пенсий с величиной прожиточного минимума в момент времени  $t$ ;

$x_{2.3,t}$  - площадь жилищ, приходящейся в среднем на одного жителя в момент времени  $t$ ;

$x_{2.8,t}$  - удельный вес учащихся, занимающихся во вторую и третью смены в момент времени  $t$ ;

$x_{3.3,t}$  - оборот розничной торговли в момент времени  $t$ ;

$x_{3.5,t}$  - уровень официально зарегистрированной безработицы в момент времени  $t$ ;

$x_{3.8,t}$  - инвестиции в основной капитал на душу населения в момент времени  $t$ ;

$x_{4.1,t}$  - численность лиц, состоящих на диспансерном учете по поводу алкоголизма и алкогольных психозов в момент времени  $t$ ;

$x_{4.2,t}$  - численность лиц, состоящих на диспансерном учете по поводу наркомании в момент времени  $t$ ;

$x_{4.3,t}$  - число пострадавших с утратой трудоспособности на один рабочий день и более в момент времени  $t$ ;

$x_{4.4,t}$  - травмы и отравления в момент времени  $t$ ;

$x_{5.1,t}$  - объем использования свежей воды в момент времени  $t$ ;

$x_{5.2,t}$  - выбросы загрязняющих веществ в атмосферный воздух в момент времени  $t$ .

Экспертная часть исходных данных представляется в форме булевой матрицы парных сравнений объектов  $O_i$  и  $O_j$ :

$$\gamma_m = \{\gamma_{ijt}^{(m)}\}, \quad (4.16)$$

где  $i = \overline{1, n}$ ;  $j = \overline{1, n}$ ;  $t = \overline{1, \dots, T}$

$$\gamma_{ijt}^{(m)} = \begin{cases} 1, & \text{если } O_i \text{ и } O_j \text{ принадлежат одной группе объектов;} \\ 0, & \text{в противном случае;} \end{cases}$$

$n$  – число объектов;

$m$  – показывает, каким экспертом предложена матрица  $\gamma_m$ ,  $m = \overline{1, l}$ .

Располагая экспертной и статистической информацией, с помощью специально разработанного программного обеспечения были оценены параметры искомой целевой функции  $f(\bar{X}, \bar{\Theta})$  [14]. Ниже представлена динамическая модель влияния факторов на уровень демографической безопасности с учетом пространственной и временной неоднородности:

$$\begin{aligned} \tilde{y}(t) \equiv f_t(X, \Theta) = & 0,82x_{1.1.1,t} - 0,61x_{1.1.2,t} - 0,15x_{1.1.3,t} + 0,33x_{1.1.4,t} - 1,05x_{1.1.5,t} + 0,89x_{1.1.6,t} + 0,35x_{1.2.1,t} + \\ & + 0,77x_{1.2.3,t} - 0,73x_{1.3.1,t} - 0,43x_{1.3.6,t} + 0,19x_{1.3.12,t} + 0,50x_{2.2,t} + 0,69x_{2.3,t} - 0,34x_{2.8,t} + 0,52x_{3.3,t} - \\ & - 0,44x_{3.5,t} + 0,32x_{3.8,t} - 0,46x_{4.1,t} - 0,36x_{4.2,t} - 0,48x_{4.3,t} - 0,52x_{4.4,t} + 0,25x_{5.1,t} - 0,36x_{5.2,t} \end{aligned}$$

Анализируя модель латентного показателя можно сделать вывод, что уровень демографической безопасности формируется большей частью под положительным влиянием таких факторов как - общий коэффициент рождаемости, коэффициент миграционного прироста, соотношение среднего размера назначенных месячных пенсий с величиной прожиточного минимума, площадь жилищ, приходящаяся в среднем на одного жителя. Существенное отрицательное воздействие на уровень демографической безопасности оказывают: общий коэффициент смертности, общая заболеваемость, уровень безработицы, а также численность лиц, состоящих на диспансерном учете по поводу алкоголизма и алкогольных психозов.



Далее рассчитаем характеристики, которые позволят отслеживать улучшение или ухудшение демографической безопасности в динамике (таблица 4.9).

Таблица 4.9 – Ранжирование объектов Оренбургской области по уровню демографической безопасности за период 2002-2005 годы с учетом пространственной и временной неоднородности.

Города и районы	Рейтинг						
	2002г.	2003г.		2004г.		2005г.	
		факт	$\delta_{i,2003}$	факт	$\delta_{i,2004}$	факт	$\delta_{i,2005}$
1	2	3	4	5	6	7	8
Абдулинский	31	30	+1	30	0	29	+1
Адамовский	17	17	0	16	+1	16	0
Акбулакский	33	35	-2	34	+1	34	0
Александровский	37	37	0	38	-1	35	+3
Асекеевский	25	24	+1	24	0	22	+2
Беляевский	24	25	-1	25	0	25	0
Бугурусланский	30	31	-1	32	-1	32	0
Бузулукский	14	16	-2	17	-1	19	-2
Гайский	38	38	0	37	+1	37	0
Грачевский	36	34	+2	35	-1	38	-3
Домбаровский	26	29	-3	26	+3	26	0
Илекский	34	36	-2	36	0	36	0
Кваркенский	42	41	+1	39	+2	39	0
Красногвардейский	43	42	+1	45	-3	46	-1
Кувандыкский	46	46	0	47	-1	45	+2
Курманаевский	22	22	0	22	0	24	-2
Матвеевский	47	47	0	46	+1	47	-1
Новоорский	12	12	0	12	0	13	-1
Новосергиевский	29	26	+3	28	-2	28	0
Октябрьский	28	28	0	29	-1	30	-1
Оренбургский	2	2	0	3	-1	3	0
Первомайский	15	15	0	15	0	15	0
Переволоцкий	35	33	+2	33	0	33	0
Пономаревский	41	43	-2	43	0	44	-1
Сакмарский	32	32	0	31	+1	31	0
Саракташский	23	23	0	23	0	23	0
Светлинский	44	44	0	44	0	43	+1
Северный	27	27	0	27	0	27	0
Соль-Илецкий	20	19	+1	21	-2	21	0
Сорочинский	21	20	+1	19	+1	17	+2
Ташлинский	19	21	-2	20	+1	20	0
Тоцкий	18	18	0	18	0	18	0
Тюльганский	40	39	+1	40	-1	40	0
Шарлыкский	39	40	-1	41	-1	42	-1
Ясненский	45	45	0	42	+3	41	+1
Абдулино	13	11	+2	11	0	11	0
Бузулук	3	3	0	1	+2	1	0
Бугуруслан	6	6	0	6	0	5	+1

Продолжение таблицы 4.9

1	2	3	4	5	6	7	8
Гай	8	9	-1	9	0	10	-1
Кувандык	9	8	+1	7	+1	7	0
Медногорск	16	14	+2	14	0	14	0
Новотроицк	4	5	-1	5	0	4	+1
Оренбург	1	1	0	2	-1	2	0
Орск	5	4	+1	4	0	4	0
Сорочинск	11	13	-2	13	0	14	-1
Соль-Илецк	7	7	0	8	-1	8	0
Ясный	10	10	0	10	0	9	+1

Наибольший рост рейтинга за 2003 год наблюдался у Новосергиевского района, за 2004 год – у Домбаровского и Ясненского районов, за 2005 год – у Александровского района (+ 3 пункта).

Наибольшее падение рейтинга за 2003 год наблюдалось у Домбаровского района, за 2004 год – у Красногвардейского района, за 2005 год – у Грачевского района.

За исследуемый период количество муниципальных образований, в которых наблюдалось повышение уровня демографической безопасности, немного превышало количество административно-территориальных образований, в которых наблюдалось снижение уровня демографической безопасности. За 2003, 2004 гг. рейтинг не изменился у 21 муниципальной единицы, за 2005 год – у 26 муниципальных единиц.

По результатам рейтинговой оценки городов и районов Оренбургской области можно сделать следующие выводы. Лидерами по уровню демографической безопасности стабильно являются практически все города области, а также Новоуральский и Оренбургский районы.

Город Оренбург занимает ведущую позицию среди всех объектов области (1 место в 2002-2003 годы), поскольку это областной центр и данный объект во многом является экономическим и финансовым центром области. Города в среднем существенно опережают сельскую местность по уровню демографической безопасности: здесь сосредоточены предприятия, предоставляющие торговые, культурные, медицинские, образовательные услуги. Образовательный уровень

работающих в городах выше, чем в среднем по области. Областной центр и наиболее крупные города области концентрируют наиболее квалифицированные и образовательные кадры. Здесь находится большинство высших учебных заведений области, много средних специальных учебных заведений и профессионально-технических училищ, которые готовят специалистов для различных отраслей хозяйства. Лечебные учреждения городов области пополняются новейшим медицинским оборудованием, внедряются современные технологии диагностики и лечения. Кроме того, имеются санатории, дома отдыха, туристические базы, профилактории, учреждения физической культуры и спорта, где можно укрепить здоровье.

Легко объяснить также, почему два сельских района – Оренбургский и Новоорский – входят в группу с сравнительно высоким уровнем демографической безопасности: примыкая к крупнейшим городам области, они используют преимущества своего положения.

Ряд сельских районов, где сосредоточены, главным образом, месторождения нефти и относительно развита сельская инфраструктура (Первомайский, Ташлинский, Бузулукский, Адамовский, Тоцкий, Соль-Илецкий) характеризуются показателями, обеспечивающими средний уровень демографической безопасности. Это явление можно объяснить тем, что в условиях естественной убыли населения прирост числа жителей происходит только за счет миграционного притока. Переселенцами являются в основном либо сельские жители других регионов области, либо иммигранты из Казахстана и государств Центральной Азии, где уровень жизни в среднем ниже, чем в регионах России. Они, как правило, не обладают достаточными средствами для приобретения жилья и адаптации в Оренбурге. Рынок труда в городе не дает возможности получить соответствующие затратам доходы в ближайшем будущем. Поэтому мигранты вынуждены расселяться в сельской местности. Привлекательны для мигрантов, прежде всего, районы с выгодным географическим положением, а также с развивающимся производством, например, нефтедобычей. Поэтому при размещении они отдадут предпочтение Оренбургскому, Акбулакскому, Первомайскому, Ташлин-

скому и Красногвардейскому районам, а также крупным городам – Оренбургу, Орску, Новотроицку.

В тоже время население таких городов как Гай, Сорочинск, Соль-Илецк, а также Северного, Абдулинского, Пономаревского и Шарлыкского районов выезжают в другие регионы России (как правило, в Самарскую, Саратовскую и Челябинскую области, Татарстан и Башкортостан), сальдо миграции здесь отрицательное. Экономика этих регионов характеризуется достаточно высоким уровнем развития, многопрофильностью. Такое соседство благоприятно для Оренбуржья: оно создает возможности для производственного кооперирования, торговли и культурного обмена. Однако, административные центры этих областей и республик значительно крупнее Оренбурга: все они – города-«миллионеры». Эти центры располагают более разветвленной сетью учебных заведений, мест приложения труда, учреждений культуры и поэтому привлекательны для молодежи. При наличии хорошей железнодорожной связи западного Оренбуржья с соседними территориями это ведет к оттоку части населения за пределы области.

Демографическая безопасность районов, расположенных вдоль границы с Казахстаном (Ясненский, Кувандыкский, Матвеевский и Светлинский районы), остается на сравнительно низком уровне: высокие показатели заболеваемости и смертности, в том числе младенческой, объясняются низким уровнем медицинского обслуживания, а также отсутствием экстренной медицинской помощи в сельских районах. В то время как данные районы сохранили достаточно высокий уровень рождаемости. Это объясняется тем, что сельские жители более привержены традициям и ценностям, которых придерживались предыдущие поколения. Заметное воздействие на рождаемость оказывает и национальный состав этих районов. Некоторые народы (например, казахи) сохранили традиции многодетности, и там, где доля этих народов в населении выше, выше и показатели рождаемости [15].

Таким образом, предложенный подход позволил сформировать редуцированные (апостериорные) наборы из 23 показателей, наиболее полно отражающие среду и систему обеспечения демографической безопасности. Агрегирование

отдельных статистических показателей в сводный интегральный индикатор, характеризующий уровень ДБ, с учетом неоднородности данных позволило провести сравнительный анализ административно-территориальных единиц по рассматриваемой синтетической категории. В частности, были выделены группы неблагополучных районов области по уровню демографической безопасности, нуждающиеся в поддержке администрации, а также районы, относительно благополучные по основным параметрам, включая здоровье населения и состояние окружающей природной среды, куда на случай чрезвычайных ситуаций можно направлять для реабилитации миграционные потоки.

#### **4.5 Моделирование зависимостей между показателями, характеризующими инвестиционный потенциал, инвестиционную активность и инвестиционный риск с учетом пространственной и временной неоднородности данных**

Для выявления характера взаимосвязей факторов, оказывающих существенное влияние на составляющие инвестиционной привлекательности, с учетом пространственной и временной неоднородности данных, построим рекурсивную систему одновременных уравнений (СОУ) на основе панельных данных. Предварительно был осуществлен нетрадиционный корреляционный анализ, который позволил выявить силу связи во времени между показателями, характеризующими инвестиционную активность, инвестиционный потенциал и инвестиционный риск и определить, с каким лагом включать в модели переменные [16]. Информационной базой послужили показатели, характеризующие инвестиционную привлекательность региона за период с 1999 по 2006 гг., представленные в таблице 4.10. Объектами наблюдения являются муниципальные образования Оренбургской области.

Таблица 4.10 – Показатели, характеризующие инвестиционную привлекательность региона

Показатели, характеризующие инвестиционную активность и инвестиционный риск	Показатели, характеризующие инвестиционный потенциал
<p><i>Показатели, характеризующие инвестиционную активность</i></p> <p><math>y_{1,it}</math> – инвестиций в основной капитал на душу населения для объекта <math>i</math> в момент времени <math>t</math>, руб.</p> <p><math>y_{2,it}</math> – ввод в действие жилых домов на 1000 населения для объекта <math>i</math> в момент времени <math>t</math>, <math>m^2</math>;</p> <p><math>y_{3,it}</math> – объем промышленной продукции на душу населения для объекта <math>i</math> в момент времени <math>t</math>, руб.</p> <p><math>y_{4,it}</math> – производство (реализация) скота и птицы для объекта <math>i</math> в момент времени <math>t</math>, центнеров на душу населения.</p> <p><i>Показатели, характеризующие инвестиционный риск</i></p> <p><math>y_{5,it}</math> – задолженность организаций по заработной плате для объекта <math>i</math> в момент времени <math>t</math>, в процентах от общего фонда заработной платы;</p> <p><math>y_{6,it}</math> – удельный вес убыточных предприятий и организаций для объекта <math>i</math> в момент времени <math>t</math>, в процентах от общего числа предприятий;</p> <p><math>y_{7,it}</math> – просроченная кредиторская задолженность предприятий для объекта <math>i</math> в момент времени <math>t</math>, в процентах от общей задолженности.</p>	<p><math>x_{1.1,it}</math> – уровень зарегистрированной безработицы для объекта <math>i</math> в момент времени <math>t</math>, в процентах;</p> <p><math>x_{1.2,it}</math> – доля населения в трудоспособном возрасте в общей численности населения для объекта <math>i</math> в момент времени <math>t</math>, в процентах;</p> <p><math>x_{1.3,it}</math> – доля лиц моложе трудоспособного возраста для <math>i</math>-го объекта в момент времени <math>t</math>, в общей численности населения, в процентах;</p> <p><math>x_{1.4,it}</math> – среднегодовая численность работников, занятых в сельскохозяйственном производстве для объекта <math>i</math> в момент времени <math>t</math>, человек;</p> <p><math>x_{1.5,it}</math> – среднегодовая численность работников, занятых в промышленности для объекта <math>i</math> в момент времени <math>t</math>, человек;</p> <p><math>x_{1.6,it}</math> – число зарегистрированных иностранных рабочих на 1000 человек для объекта <math>i</math> в момент времени <math>t</math>, человек;</p> <p><math>x_{2.1,it}</math> – сальдированный финансовый результат (прибыль минус убыток) одного предприятия для объекта <math>i</math> в момент времени <math>t</math>, рублей;</p> <p><math>x_{2.2,it}</math> – уровень рентабельности реализованной продукции сельского хозяйства в сельскохозяйственных организациях для объекта <math>i</math> в момент времени <math>t</math>, в процентах;</p> <p><math>x_{3.1,it}</math> – оборот розничной торговли на душу населения для объекта <math>i</math> в момент времени <math>t</math>, рублей.</p>

На основе панельных данных, с помощью статистического пакета Stata, была построена модель с фиксированными эффектами для первой эндогенной переменной:

$$y_{1,it} = \alpha_i + \beta_1 x_{1.1,it} + \beta_2 x_{1.6,it} + \beta_3 x_{2.1,it} + \varepsilon_{it}, \quad i = \overline{1, \dots, 47}, \quad t = \overline{1, \dots, 8}. \quad (4.17)$$

Оценка коэффициентов модели с фиксированными эффектами:

$$\hat{y}_{1,it} = \hat{\alpha}_i - 2556,603 x_{1.1,it} + 112,746 x_{1.6,it} + 0,004 x_{2.1,it}, \quad i = \overline{1, \dots, 47}, \quad t = \overline{1, \dots, 8}. \quad (4.18)$$

(1055,257)                      (57,805)                      (0,001)

Для того, чтобы доказать, что введение в модель фиксированных эффектов оправдано, необходимо проверить гипотезу об их значимости. Если они равны между собой, то модели с фиксированными эффектами следует предпочесть обычную регрессию. Для проверки гипотезы  $H_0: \alpha_1 = \alpha_2 \dots = \alpha_n$  используется F-статистика (4.6).

В нашем случае, наблюдаемое значение  $F=7,88$  попадает в критическую область ( $F_{кр}(0,05;46;326)=1,405$ ), следовательно, нулевая гипотеза о равенстве всех индивидуальных эффектов отвергается.

С помощью F-критерия (4.4) проверим гипотезу  $H_0: \beta_1 = \beta_2 \dots \beta_k = 0$ .

В нашем случае, наблюдаемое значение  $F=16.19$  попадает в критическую область ( $F_{кр}(0,05;3;326)=2,632$ ), следовательно, коэффициенты значимо отличаются от нуля.

На основе панельных данных, была построена модель со случайными эффектами для первой эндогенной переменной:

$$y_{1,it} = \alpha + \beta_1 x_{1.1,it} + \beta_2 x_{1.6,it} + \beta_3 x_{2.1,it} + \varepsilon_{it}, \quad (4.19)$$

где  $\varepsilon_{it} = u_i + v_{it}$ ,  $i = \overline{1, \dots, 47}$ ,  $t = \overline{1, \dots, 8}$ .

Оценка коэффициентов модели со случайными эффектами:

$$\hat{y}_{1,it} = \underset{(1043,39)}{6476,28} - \underset{(918,93)}{2141,95} x_{1.1,it} + \underset{(56,78)}{133,14} x_{1.6,it} + \underset{(0,0006)}{0,004} x_{2.1,it}, \quad i = \overline{1, \dots, 47}, \quad t = \overline{1, \dots, 8}, \quad (4.20)$$

Для проверки гипотезы о значимости случайных эффектов:

$$H_0: \sigma_u^2 = 0$$

$$H_1: \sigma_u^2 \neq 0$$

Строится статистика: 
$$LM = \frac{nT}{2 \cdot (T-1)} \left[ \frac{\sum_{i=1}^n \left( \sum_{t=1}^T e_{it} \right)^2}{\sum_{i=1}^n \sum_{t=1}^T e_{it}^2} - 1 \right]^2.$$

При справедливости нулевой гипотезе LM подчиняется закону распределения Хи–квадрат с одной степенью свободы. Так как  $LM(1) = 221,78$  ( $p=0,0000$ ), то существуют основания предпочесть модель со случайными эффектами простой регрессии.

Согласно тесту Хаусмана ( $W(3)=8,52$ ,  $p=0,0368$ ) предпочтение следует отдать модели с фиксированными эффектами.

Таким образом, оценка уравнения регрессии будет иметь вид:

$$\hat{y}_{1,it} = \hat{\alpha}_i - 2556,603 x_{1.1,it} + 112,746 x_{1.6,it} + 0,004 x_{2.1,it}, i = \overline{1, \dots, 47}, t = \overline{1, \dots, 8}, \bar{R}^2 = 0,748$$

(1055,257)
(57,805)
(0,001)

Оценки индивидуальных эффектов модели:

$a_1 = 1985,43$	$a_{13} = 2342,81$	$a_{25} = 3422,07$	$a_{37} = 13323,96$
$a_2 = 4551,00$	$a_{14} = 3616,18$	$a_{26} = 3213,78$	$a_{38} = 17198,18$
$a_3 = 5039,34$	$a_{15} = 4162,58$	$a_{27} = 4588,64$	$a_{39} = 12306,61$
$a_4 = 3619,8$	$a_{16} = 4463,95$	$a_{28} = 3253,39$	$a_{40} = 3649,83$
$a_5 = 4164,94$	$a_{17} = 4097,43$	$a_{29} = 3385,8$	$a_{41} = 5006,00$
$a_6 = 5319,24$	$a_{18} = 6392,72$	$a_{30} = 10219,33$	$a_{42} = 15039,85$
$a_7 = 4473,5$	$a_{19} = 4093,67$	$a_{31} = 5647,14$	$a_{43} = 10633,21$
$a_8 = 8174,55$	$a_{20} = 4962,59$	$a_{32} = 4095,02$	$a_{44} = 4094,43$
$a_9 = 5783,25$	$a_{21} = 14959,61$	$a_{33} = 8384,81$	$a_{45} = 5412,73$
$a_{10} = 4932,3$	$a_{22} = 32503,58$	$a_{34} = 4816,98$	$a_{46} = 23832,39$
$a_{11} = 4306,81$	$a_{23} = 3999,49$	$a_{35} = 7532,37$	$a_{47} = 3314,49$
$a_{12} = 2851,07$	$a_{24} = 4282,7$	$a_{36} = 2769,37$	

Для выявления характера взаимосвязей факторов, оказывающих существенное влияние на составляющие инвестиционной привлекательности, с учетом пространственной и временной неоднородности данных предложены следующие модели в форме рекурсивной системы одновременных уравнений на основе панельных данных:



$$\left\{ \begin{array}{l} y_{1,it} = \alpha_i + C_1 x_{1.1,it} + C_2 x_{1.6,it} + C_3 x_{2.1,it} + \varepsilon_{it} - \text{модель с фиксированными эффектами} \\ y_{2,it} = \alpha_i + \beta_1 y_{1,it} + C_1 x_{1.2,it} + C_2 x_{1.5,it} + \varepsilon_{it} - \text{модель с фиксированными эффектами} \\ y_{3,it} = \alpha_i + \beta_1 y_{1,it} + C_1 x_{1.5,it} + C_2 x_{3.1,it} + \varepsilon_{it} - \text{модель с фиксированными эффектами} \\ y_{4,it} = \alpha_i + C_1 x_{1.2,it} + C_2 x_{1.4,it} + \varepsilon_{it} - \text{модель с фиксированными эффектами} \\ y_{5,it} = \alpha_i + \beta_1 y_{2,it} + C_1 x_{1.4,it} + C_2 x_{1.5,it} + C_3 x_{2.2,it} + \varepsilon_{it} - \text{модель с фиксированными эффектами} \\ y_{6,it} = \alpha + \beta_1 y_{1,it} + \beta_2 y_{3,it} + C_1 y_{5,it-1} + C_2 x_{2.2,it} + C_3 x_{3.1,it} + \varepsilon_{it} - \text{модель со случайными эффектами} \\ y_{7,it} = \alpha + \beta_1 y_{2,it} + \beta_2 y_{5,it} + C_1 y_{6,it-1} + C_2 x_{1.2,it} + C_2 x_{2.1,it} + \varepsilon_{it} - \text{модель со случайными эффектами} \end{array} \right.$$

где  $i = \overline{1, \dots, 47}$ ,  $t = \overline{1, \dots, 8}$

Оценка моделей осуществлялась с помощью статистического пакета Stata. Результаты исследования каждого уравнения приведены в приложении Р. Согласно тесту Хаусмана, в первых пяти случаях предпочтение отдано моделям с фиксированными эффектами, а для двух последних – моделям со случайными эффектами:

$$\left\{ \begin{array}{l} \hat{y}_{1,it} = \hat{\alpha}_i - 2556,603 x_{1.1,it} + 112,746 x_{1.6,it} + 0,004 x_{2.1,it}, \quad \hat{R}^2 = 0,748 \\ \quad \quad \quad (1055,257) \quad \quad \quad (57,805) \quad \quad \quad (0,001) \\ \hat{y}_{2,it} = \hat{\alpha}_i + 0,002 y_{1,it} + 10,675 x_{1.2,it} - 0,008 x_{1.5,it}, \quad \hat{R}^2 = 0,916 \\ \quad \quad \quad (0,0007) \quad \quad \quad (1,921) \quad \quad \quad (0,003) \\ \hat{y}_{3,it} = \hat{\alpha}_i + 1,516 y_{1,it} + 4,793 x_{1.5,it} + 1,951 x_{3.1,it}, \quad \hat{R}^2 = 0,779 \\ \quad \quad \quad (0,559) \quad \quad \quad (1,078) \quad \quad \quad (0,799) \\ \hat{y}_{4,it} = \hat{\alpha}_i + 0,022 x_{1.2,it} + 0,0001 x_{1.4,it}, \quad \hat{R}^2 = 0,802 \\ \quad \quad \quad (0,012) \quad \quad \quad (0,00005) \\ \hat{y}_{5,it} = \hat{\alpha}_i - 0,068 y_{2,it} + 0,004 x_{1.4,it} - 0,0004 x_{1.5,it} - 0,027 x_{2.2,it}, \quad \hat{R}^2 = 0,873 \\ \quad \quad \quad (0,013) \quad \quad \quad (0,001) \quad \quad \quad (0,0002) \quad \quad \quad (0,015) \\ \hat{y}_{6,it} = 49,249 - 0,0003 y_{1,it} - 0,0001 y_{3,it} + 0,214 y_{5,it-1} - 0,538 x_{2.2,it} - 0,0005 x_{3.1,it}, \quad \hat{R}^2 = 0,935 \\ \quad \quad \quad (2,091) \quad \quad \quad (0,0001) \quad \quad \quad (0,00004) \quad \quad \quad (0,083) \quad \quad \quad (0,047) \quad \quad \quad (0,0002) \\ \hat{y}_{7,it} = 187,809 - 0,048 y_{2,it} + 1,015 y_{5,it} + 0,120 y_{6,it-1} - 2,425 x_{1.2,it} - 0,000003 x_{2.1,it}, \quad \hat{R}^2 = 0,864 \\ \quad \quad \quad (36,408) \quad \quad \quad (0,021) \quad \quad \quad (0,402) \quad \quad \quad (0,042) \quad \quad \quad (0,601) \quad \quad \quad (0,00001) \end{array} \right.$$

где  $i = \overline{1, \dots, 47}$ ,  $t = \overline{1, \dots, 8}$ .

Согласно полученным результатам, инвестиции в основной капитал складываются под влиянием уровня официально зарегистрированной безработицы, числа зарегистрированных иностранных мигрантов и сальдированного финансового результата предприятий. Увеличение уровня безработицы на 1% приводит к

снижению инвестиций в основной капитал в среднем на 2556,6 руб. Увеличение числа зарегистрированных иностранных мигрантов на 1% влечет рост инвестиций в среднем на 112,7 руб. Рост сальдированного финансового результата на 1000 руб. способствует росту инвестиций в среднем на 4 руб.

На ввод в действие жилых домов положительное влияние оказывают такие показатели, как инвестиции в основной капитал и доля населения в трудоспособном возрасте, а отрицательное – среднегодовая численность работников, занятых в промышленности. Увеличение инвестиций в основной капитал на 1000 руб. приводит к росту ввода в действие жилых домов в среднем на  $2 \text{ м}^2$ . Увеличение населения в трудоспособном возрасте на 1% приводит к росту ввода в действие жилых домов в среднем на  $10,7 \text{ м}^2$ .

Увеличение в среднем объема промышленной продукции в Оренбургской области происходит за счет увеличения инвестиций в основной капитал, среднегодовой численности работников, занятых в промышленности и оборота розничной торговли. Увеличение инвестиций в основной капитал на 1 руб. приводит к росту объема промышленной продукции в среднем на 1,5 руб. Рост среднегодовой численности работников, занятых в промышленности на 1 человека, способствует росту объема промышленной продукции в среднем на 4,8 руб. При увеличении оборота розничной торговли на 1 руб. происходит рост объема промышленной продукции в среднем на 2 руб.

На производство скота и птицы оказывают положительное влияние такие показатели, как доля населения в трудоспособном возрасте и среднегодовая численность работников, занятых в сельскохозяйственном производстве. Увеличение среднегодовой численности работников, занятых в сельскохозяйственном производстве на 1 человека приводит к росту производства скота и птицы в среднем на 0,0001 центнер.

Увеличение ввода в действие жилых домов, среднегодовой численности работников, занятых в промышленности и уровня рентабельности реализованной продукции сельского хозяйства приводит к снижению задолженности организаций по заработной плате. Рост ввода в действие жилых домов на  $1 \text{ м}^2$  приводит к

снижению задолженности в среднем на 0,07%; Увеличение среднегодовой численности работников, занятых в промышленности на 1 человека способствует снижению задолженности в среднем на 0,0004%.

На удельный вес убыточных предприятий отрицательное влияние оказывают инвестиции в основной капитал, объем промышленной продукции, уровень рентабельности реализованной продукции сельского хозяйства и оборот розничной торговли, а положительное – задолженность организаций по заработной плате. С ростом инвестиций в основной капитал на 1 руб. происходит снижение удельного веса убыточных предприятий в среднем на 0,0003%; увеличение объема промышленной продукции на 1 руб. приводит к уменьшению убыточных предприятий в среднем на 0,0001%. Увеличение задолженности по заработной плате в предыдущий момент времени на 1% способствует росту убыточных предприятий в среднем на 0,2%.

Просроченная кредиторская задолженность складывается под влиянием таких показателей, как ввод в действие жилых домов, задолженность по заработной плате, удельный вес убыточных предприятий, доля населения в трудоспособном возрасте, сальдированный финансовый результат. Увеличение задолженности по заработной плате на 1% приводит к росту просроченной кредиторской задолженности в среднем на 1,02%; увеличение сальдированного финансового результата на 1руб. способствует снижению просроченной кредиторской задолженности в среднем на 0,000003%; рост убыточных предприятий в предыдущий момент времени на 1% влечет увеличение просроченной задолженности в среднем на 0,12%;

#### **4.6 Модели эколого-экономического риска с учетом пространственной и временной неоднородности данных**

Для моделирования зависимости уровня экологического риска от показателей социально-экономического состояния муниципалитета с учетом пространственной и временной неоднородности данных рассмотрим бинарную модель вида

$$y_{it} = \begin{cases} 1, & y_{it}^* > 0 \\ 0, & y_{it}^* \leq 0 \end{cases}, \quad i = \overline{1, \dots, n}, \quad t = \overline{1, \dots, T}. \quad (4.21)$$

где  $y_{it}^* = x_{it}^T \beta + \alpha_i + \varepsilon_{it}$ ,

$x_{it}$  – вектор значений объясняющих переменных для  $i$ -го объекта в момент времени  $t$ ;

$\beta = (\beta_1, \dots, \beta_p)$  – вектор коэффициентов модели;

$\alpha_1, \dots, \alpha_N$  – индивидуальные эффекты;

$\varepsilon_{it}$  – случайные величины, одинаково распределенные для всех  $i$  и  $t$ , независимые для  $\forall i, j: i \neq j$  и для  $\forall t, \tau: t \neq \tau$ , а также независимые с  $x_{it}$ .

В нашем случае объектами выступают муниципальные образования Оренбургской области, а показателями, их характеризующими, введенные выше доля убыточных предприятий (DUB), розничный товарооборот (ROZN) и др. (CITY, ZP, TRAVMPR, NOV, BER1, OSf2, T02).

В зависимости от того, какая функция рассматривается в качестве функции  $F$ , разделяют, например, пробит и логит-модели. Отметим, что при оценке моделей дискретного или ограниченного выбора по панельным данным возникают дополнительные трудности, связанные с более сложным видом функции правдоподобия и алгоритмами ее максимизации из-за естественного наличия корреляций между различными компонентами случайной ошибки, поскольку при использовании панельных данных наблюдения, относящиеся к одному объекту в разные моменты времени, трудно считать независимыми.

Аналогично линейным моделям панельных данных различные предположения относительно  $\alpha_i$  приводят к двум типам моделей: с фиксированными эффектами и со случайными эффектами.

Если  $\alpha_i$  трактуются как неизвестные детерминированные параметры, то имеем модель с фиксированными эффектами, которая используется, когда исследуемая совокупность состоит из уникальных объектов и нас интересует получение прогноза для каждого объекта (говорят, что выборка исчерпывает совокупность). Такую модель нельзя распространить на объекты, не вошедшие в пер-

воначальную выборку, так как для них неизвестны индивидуальные эффекты. Кроме того, в модели с фиксированными эффектами, которые предполагаются постоянными во времени, невозможно использование в качестве регрессоров переменных, не изменяющихся во времени, из-за их линейной зависимости с индивидуальными эффектами. Таким образом, указанная трактовка  $\alpha_i$  соответствует включению в модель  $N$  фиктивных переменных. В этом случае логарифм функции правдоподобия имеет вид:

$$\ln L(\beta, \alpha_1, \dots, \alpha_N) = \sum_{i,t} y_{it} \ln F_{\alpha_i}(x_{it}^T \beta + \alpha_i) + \sum_{i,t} (1 - y_{it}) \ln [1 - F_{\alpha_i}(x_{it}^T \beta + \alpha_i)]. \quad (4.22)$$

В общем случае максимизация по  $\beta, \alpha_1, \dots, \alpha_N$  дает состоятельные оценки только при  $T \rightarrow \infty$ . Для получения состоятельных оценок  $\beta$  при конечном  $T$  используется метод условного максимального правдоподобия, для чего необходимо иметь некоторое множество статистик, достаточных для  $\alpha_1, \dots, \alpha_N$ , при заданных значениях которых вклад  $i$ -го объекта в правдоподобие уже не зависит от  $\alpha_i$ . Для пробит-модели таких статистик не существует [17, 18]. Проверка гипотезы об отсутствии индивидуальных эффектов для логит-модели осуществляется с помощью критерия типа Хаусмана, использующего разность между оценкой коэффициентов, полученной методом условного правдоподобия, и обычной оценкой максимального правдоподобия, полученной без учета панельной структуры данных.

Если  $\alpha_i$  рассматриваются как случайные величины,  $\alpha_i \sim NID(0, \sigma_{\alpha_i}^2)$ , то имеем модель со случайными эффектами, которая используется, когда нужно описать поведение совокупности в целом и построить прогнозы для типичного представителя совокупности, то есть в тех случаях, когда данные являются случайной выборкой из большой популяции. Предполагается, что индивидуальные эффекты и ошибки некоррелированы. Требованием этой модели является также некоррелированность эффектов и объясняющих переменных. Это требование необходимо проверять. Если оно нарушается, то полученные оценки будут несо-

стоятельными. С другой стороны, использование модели с фиксированными эффектами при отсутствии корреляции дает менее качественные оценки. Логарифм функции правдоподобия для модели со случайными эффектами имеет вид:

$$\ln L(\beta, \alpha_1, \dots, \alpha_N) = \sum_{i=1}^N y_{it} L_i, \quad (4.23)$$

$$\text{где } L_i = \int_{-\infty}^{+\infty} f(\alpha_i) \prod_{t=1}^T \left\{ \left[ \Phi \left( \frac{x_{it}^T \beta + \alpha_i}{\sqrt{1 - \sigma_{\alpha_i}^2}} \right) \right]^{y_{it}} \cdot \left[ 1 - \Phi \left( \frac{x_{it}^T \beta + \alpha_i}{\sqrt{1 - \sigma_{\alpha_i}^2}} \right) \right]^{1 - y_{it}} \right\} d\alpha_i,$$

$$\varphi(\alpha_i) = \frac{1}{\sqrt{2\pi\sigma_{\alpha_i}^2}} \exp\left(-\frac{\alpha_i^2}{2\sigma_{\alpha_i}^2}\right).$$

Обозначим  $u_{it} = \alpha_i + \varepsilon_{it}$ .

Тогда

$$M(u_{it}) = 0,$$

$$D(u_{it}) = \sigma_{\alpha_i}^2 + \sigma_{\varepsilon}^2,$$

$$\text{cov}(u_{it}, u_{is}) = \sigma_{\alpha_i}^2, \quad t \neq s$$

$$\rho(u_{it}, u_{is}) = \frac{\sigma_{\alpha_i}^2}{\sigma_{\alpha_i}^2 + \sigma_{\varepsilon}^2}, \quad t \neq s$$

В стандартной спецификации  $D(u_{it})=1$ , и, следовательно,  $\rho(u_{it}, u_{is}) = \sigma_{\alpha_i}^2$ .

Поэтому проверка гипотезы об отсутствии индивидуальных эффектов  $H_0: \sigma_{\alpha_i}^2 = 0$  сводится к проверке на основе статистики Вальда гипотезы  $H_0: \rho = 0$ .

В результате оценивания в пакете Stata бинарной пробит-модели со случайными эффектами нами было получено следующее выражение для вероятности того, что для  $i$ -го города или района в момент времени  $t$  будет характерен высокий или средний эколого-экономический риск:

$$\hat{P}\{y_{it} = 1|x_{it}\} = \Phi(1,01 - 1,01 \cdot CITY_i - 0,18 \cdot ROZN_i - 3,36 \cdot LOGZP_i - 0,003 \cdot DUB_i - 0,09 \cdot TRAMPR_i + \frac{0,48}{NOV_i} - 0,62 \cdot BER1 + 0,59 \cdot OSf2 - 0,44 \cdot T02 + \alpha_i)$$

Величины большинства коэффициентов модели, а также направление влияния отобранных показателей несущественно отличаются от модели, построенной без учета панельной структуры. Стоит, однако, отметить меньшую значимость коэффициента при переменной DUB (доли убыточных предприятий), что сигнализирует о снижении влияния данного фактора на уровень эколого-экономического риска [19].

#### **Вопросы и задания по теоретическому материалу раздела 4**

1. Запишите линейную регрессионную модель панельных данных с фиксированными эффектами. Какие условия накладываются на эту модель? Что характеризуют индивидуальные эффекты в линейной модели с фиксированными эффектами?
2. Каким методом можно оценить коэффициенты линейной регрессионной модели с фиксированными эффектами? Опишите алгоритм исследования линейной модели с фиксированными эффектами.
3. Запишите двунаправленную модель с фиксированными эффектами. Опишите алгоритм исследования двунаправленной модели с фиксированными эффектами.
4. Запишите линейную регрессионную модель панельных данных со случайными эффектами. Как можно оценить коэффициенты линейной регрессионной модели со случайными эффектами? Как осуществляется проверка гипотезы о значимости случайных эффектов?
5. С помощью какого теста можно подтвердить или опровергнуть априорные предположения относительно спецификации модели? В чем он заключается?

6. Опишите, каким образом можно сформировать панели данных по категориям безработных, различающихся по значимым социально-демографическим и профессионально-квалификационным характеристикам.

7. В результате оценки коэффициентов двунаправленной модели с фиксированными эффектами для категорий безработных, различающихся по полу, возрасту и уровню профессионального образования, получены следующие результаты:

$$\hat{y}_{it} = a_i + c_t + \underset{(0,04)}{0,32} x_{1,it} + \underset{(0,04)}{0,41} x_{2,it} - \underset{(0,03)}{0,69} x_{3,it} + \underset{(0,05)}{0,50} x_{4,it}, \quad i = \overline{1, \dots, 24}, \quad t = \overline{1, \dots, 24}.$$

Дайте интерпретацию оценкам коэффициентов при объясняющих переменных уравнения регрессии, поясните смысл и возможности практического использования значений индивидуальных и временных эффектов.

8. В результате моделирования задолженности организаций по заработной плате, с учетом пространственной и временной неоднородности данных, были получены следующие результаты (n=47, T=8):

### Результаты оценивания моделей с фиксированными и случайными эффектами

Проверяемая гипотеза	Наблюдаемое значение статистики	Значимость нулевой гипотезы
Модель с фиксированными эффектами		
$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_5 = 0$	F(4,325)=59,61	p=0,0000
$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_{47} = 0$	F(46,325)=8,93	p=0,0000
$\hat{y}_{1it} = \hat{\alpha}_i - \underset{(0,013)}{0,068} x_{1,it} + \underset{(0,001)}{0,004} x_{2,it} - \underset{(0,0002)}{0,0004} x_{3,it} - \underset{(0,015)}{0,027} x_{4,it}$		
Модель со случайными эффектами		
$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_5 = 0$	Wald(4)=206,00	p=0,0000
$H_0 : \sigma_u^2 = 0$	LM(1) = 160,39	p=0,0000
$\hat{y}_{1,it} = 4,946 - \underset{(0,004)}{0,035} x_{1,it} + \underset{(0,0003)}{0,004} x_{2,it} - \underset{(0,00001)}{0,0001} x_{3,it} - \underset{(0,015)}{0,025} x_{4,it}$		

где  $y_{1,it}$  – задолженность организаций по заработной плате для объекта  $i$  в момент времени  $t$ , в процентах от общего фонда заработной платы;



$x_{1,it}$  – ввод в действие жилых домов на 1000 населения для объекта  $i$  в момент времени  $t$ ,  $m^2$  .;

$x_{2,it}$  – среднегодовая численность работников, занятых в сельскохозяйственном производстве для объекта  $i$  в момент времени  $t$ , человек;

$x_{3,it}$  – среднегодовая численность работников, занятых в промышленности для объекта  $i$  в момент времени  $t$ , человек;

$x_{4,it}$  – уровень рентабельности реализованной продукции сельского хозяйства в сельскохозяйственных организациях для объекта  $i$  в момент времени  $t$ , в процентах;

### Тест Хаусмана

Наблюдаемое значение статистики	Значимость нулевой гипотезы
$W(4)=69,41$	$p=0,0000$

Проведите математико-статистическое исследование моделей. Какую модель следует выбрать: с фиксированными или случайными эффектами?

9. Укажите особенности построения моделей бинарного выбора на основе панельных данных: формирование информационной базы, методы оценивания, свойства оценок максимального правдоподобия и др.

10. Опишите известные Вам подходы к построению моделей бинарного выбора с фиксированными эффектами.

11. Опишите известные Вам подходы к построению моделей бинарного выбора со случайными эффектами.

### Задания к самостоятельной и исследовательской работе

1. Провести моделирование числа трудоустроенных безработных на основе панели данных, представленной в файле panel.xls и содержащей поквартальную информацию по категориям безработных, характеризующихся показателями:

$y_t$  – число безработных, трудоустроенных при содействии Центра занятости населения в квартале  $t$ , чел.;

$x_{1,t}$  – число безработных, зарегистрированных в Центре занятости населения в целях поиска работы в квартале  $t$ , чел.;

$x_{2,t}$  – число безработных, состоявших на учете в Центре занятости населения в целях поиска работы на начало квартала  $t$ , чел.;

$x_{3,t}$  – число безработных, получавших пособие по безработице в квартале  $t$ , чел.;

$x_{4,t}$  – число безработных, состоявших на учете в Центре занятости населения, которым предлагались вакансии в квартале  $t$ , чел.

2. Осуществите моделирование зависимостей на основе панельных данных между показателями, характеризующими демографическую безопасность.

3. Осуществите моделирование зависимостей на основе панельных данных между показателями, характеризующими экономическую безопасность.

4. Осуществите моделирование зависимостей на основе панельных данных между показателями, характеризующими инвестиционную привлекательность.

5. Осуществите моделирование зависимостей на основе панельных данных между показателями, характеризующими антропогенную нагрузку и эколого-экономический риск.

Замечания к заданиям 2-5:

1) объект и предмет исследования выберите самостоятельно;  
2) самостоятельно сформируйте информационную базу, указав источник статистических данных.

6. Сформируйте информационную базу по интересующей Вас социально-экономической проблеме, требующей наблюдения за одними и теми же объектами за 2 периода времени, и проведите исследование на основе логит-модель бинарного выбора с фиксированными эффектами. Подробно опишите структуру информационной базы и выполненные Вами преобразования. Примеры: взаимо-

связь безработицы и материального благосостояние домохозяйства, взаимосвязь безработицы и удовлетворенностью жизнью, взаимосвязь эколого-экономического риска и социально-экономических характеристик региона.

7. Сформируйте информационную базу по интересующей Вас социально-экономической проблеме, требующей наблюдения за одними и теми же объектами за несколько периодов времени, и проведите исследование на основе модели бинарного выбора со случайными эффектами. Подробно опишите структуру информационной базы и выполненные Вами преобразования.

#### **Список использованных источников к разделу 4**

1 Балаш, В.А. Модели линейной регрессии для панельных данных: учебное пособие для ВУЗов / В.А. Балаш, О.С. Балаш. – М., 2002. – 65 с.

2 Эконометрика: учебник / И.И. Елисеева, С.В. Курышева, Т.В. Костеева и др.; Под ред. И.И. Елисеевой. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2005. – 576 с.

3 Магнус, Я.Р. Эконометрика. Начальный курс: учеб / Я.Р. Магнус, П.К. Катышев, А.А. Пересецкий. – 6-е изд., перераб. и доп. – М.: Дело, 2004. – 576 с.

4 Green, W. H. *Econometric Analysis* / W. H. Green. Third Edition. New York University, Pritice Hall, 1997. – 1070 p.

5 Matyas L. and Sevestre P. *The Econometrics of Panel Data. A Handbook of Theory and Application*. 2nd revised edition, Kluwer Academic Publisher, Dordrecht, 1995.

6 Ратникова, Т. А. Введение в эконометрический анализ панельных данных / Т.А. Ратникова // Экон. журн. Высш. шк. экономики. - 2006. - Т. 10, 2. - С. 267-316; Т. 10, 3. - С. 492-519; Т.10, 4. – С.638-669.

7 Харламов, А.В.. Оценка коэффициентов линейной регрессии с дополнительными ограничениями / В.Н. Михайлов, А.В. Харламов// Заводская лаборатория. Т. 66.- № 11.- 2000. - С. 57-60.

8 Харламов, А.В. Особенности построения географически взвешенной регрессии для моделирования рынка недвижимости / В.А. Балаш, О.С. Балаш, А.В. Харламов // Вестник Саратовского государственного социально экономического университета; 2008, №5 (24) - С. 125-127.

9 Харламов, А.В. Адаптивные методы учета территориальной неоднородности при моделировании цен на недвижимость / О.С. Балаш, А.В. Харламов // Вестник Самарского государственного экономического университета; 2009, №1 (51). - С. 147-153.

10 Харламов, А.В. Техника геокодирования в построении географически взвешенных регрессионных моделей при массовой оценке в условиях неопределенности и неоднородности исходных данных / А.В. Захаров, А.В. Харламов // Имущественные отношения в Российской Федерации, 2010, №7, с.76-85.

11 Харламов, А.В. Исследование динамики цен на жилую недвижимость методом географически взвешенной регрессии /А.В. Харламов //Известия Саратовского университета.-2011.- Том 11, Выпуск 2.-С.25-29

12 Бравичева, О.С. Методы и инструментальные средства исследования влияния факторов на продолжительность поиска работы / О.С. Бравичева // Вестник ОГУ. – 2006. – №2. – С. 140–144.

13 Реннер, А.Г. Моделирование стоимости жилья на вторичном рынке жилья / А.Г. Реннер, О.С. Стебунова // Вестник ОГУ. – 2005. – №10. Приложение. – С. 179–182.

14 Реннер, А.Г. Моделирование уровня демографической безопасности Оренбургской области с учетом неоднородности данных / А.Г. Реннер, О.И. Бантикова // Взаимодействие реального и финансового секторов в трансформационной экономике: материалы Международной конференции. – Оренбург, 2008.

15 Бантикова, О.И. Построение интегрального показателя, характеризующего уровень демографической безопасности / О.И. Бантикова // Вестник ОГУ. – 2007. – № 2. – С. 98 – 104.

16 Жемчужникова, Ю.А. Моделирование зависимостей между показателями, характеризующими инвестиционную привлекательность / Ю.А. Жемчужникова // Вестник ОГУ. – 2008. – №9. – С. 118 – 121.

17 17. Носко, В. П. Эконометрика. Кн. 1. Ч. 1, 2: учебник. – М.: Издательский дом «Дело» РАНХиГС.-2011. – 672 с.

18 18. Носко, В. П. Эконометрика. Кн. 2. Ч. 3, 4: учебник. – М.: Издательский дом «Дело» РАНХиГС.- 2011. – 576 с.

19 19. Седова, Е.Н. Моделирование взаимосвязи экологического риска и социально-эколого-экономического профиля региона / Е.Н. Седова // Материалы VIII Международной научно-методической конференции «Информатика: проблемы, методология, технологии» (7-8 февраля 2008 г). – Воронеж, Издательско-полиграф. Центр Воронежского государственного университета. - 2008. – Ч.2, - С.216-220.

## **5 Моделирование взаимосвязей социально-экономических показателей на основе многомерных временных рядов**

При изучении взаимосвязей между социально-экономическими показателями, измеренными во времени применение обычного корреляционного и регрессионного анализа является некорректным. Применение специальных (нетрадиционных коэффициентов корреляции) позволяет выявить структуру связей между переменными, представленными временными рядами, а также формировать набор предопределенных переменных.

В зависимости от свойств временных рядов выбирается модель для прогнозирования. Следует отметить, что если все участвующие в модели ряды динамики стационарны, то специальных предварительных исследований не требуется, то есть можно строить уравнение регрессии по исходным (не преобразованным) рядам динамики показателей. Для нестационарных временных рядов необходимы предварительные исследования на характер тренда, так как подходы к построению многофакторных моделей для них отличаются.

Раздел организован следующим образом. В подразделе 5.1 рассматривается метод нетрадиционного корреляционного анализа: обосновывается необходимость применения специальных коэффициентов корреляции для данных временного характера, приводятся формулы расчета модифицированного, адаптивного, вероятностного и условного коэффициентов корреляции. В подразделе 5.2 на основе нетрадиционного корреляционного анализа произведен отбор факторов оказывающих влияние на показатели объемов производства сельского хозяйства и промышленности, демографические, миграционные процессы, инвестиционную привлекательность и потенциал региона. В подразделе 5.3 рассматриваются подходы к моделированию и прогнозированию показателей, характеризующих социально-экономические процессы в регионе, на основе многомерных временных рядов. В начале дается краткое описание основных подходов к построению многофакторных моделей в зависимости от свойств временных рядов. Далее приводятся примеры по-

строения и исследования многофакторных моделей, как в форме одного уравнения, так и группы независимых и одновременных уравнений. Так изучена динамика структуры показателей, характеризующих социально-экономические процессы, осуществлено прогнозирование объемов производства промышленности и сельского хозяйства региона, основных индикаторов демографической безопасности и показателей рынка труда региона.

### **5.1 Особенности изучения взаимосвязей показателей, измеренных во времени. Методы нетрадиционного корреляционного анализа факторов**

Классический корреляционный анализ, применяемый для анализа связи переменных, измеренных в пространстве, имеет ограниченное применение в случае временных рядов. Для применения стандартных коэффициентов корреляции необходимо, чтобы ряды динамики социально-экономических показателей были эргодическими случайными процессами. Эргодическими случайными процессами называется подкласс стационарных случайных процессов, у которых числовые характеристики оцениваются не по множеству наблюдаемых значений в любом временном сечении, а по одной траектории [24, 47]. Применение корреляционного анализа для заведомо нестационарных временных рядов приведет к эффекту «ложной» корреляции – связь будет обусловлена наличием тренда, а не реальной причинной зависимостью. К тому же временные ряды представлены единственной реализацией, так как большое число реализаций в экономике при одних и тех же условиях принципиально невозможно [47].

Если временной ряд нестационарен, то он известными способами приводится к приблизительно стационарному виду [44]. Но, из временного ряда удаляется весьма существенная информация, то есть имеется возможность исследования только краткосрочной связи, кроме этого нет возможности проследить изменение этих связей во времени.

В работе [33] для анализа взаимосвязей между показателями, измеренными во времени (для нестационарных временных рядов) предлагается пользоваться методами нетрадиционного корреляционного анализа.

Рассматриваются коэффициенты корреляции, которые позволяют измерять изменение силы связи двух переменных во времени, а также в зависимости от амплитуды их отклонения от некоторой критической величины.

О наличии положительной или отрицательной корреляции в момент  $t$  предлагается судить по совпадению или несовпадению знаков прироста интересующих переменных. Для получения общего представления об усредненных корреляционных свойствах двух нестационарных рядов используют коэффициент, основывающийся на использовании приростов переменных. Модифицированный коэффициент корреляции рассчитывается по формуле:

$$r_{\text{mod}} = \frac{\sum_{t=2}^T \Delta x_{1t} \Delta x_{2t}}{\sum_{t=2}^T |\Delta x_{1t} \Delta x_{2t}|}, \quad (5.1)$$

где  $\Delta x_{it} = x_{it} - x_{it-1}$ ,  $i = 1, 2$ .

Знаменатель в этой формуле играет роль нормирующего коэффициента, благодаря которому  $r_{\text{mod}}$  находится в пределах  $[-1; 1]$ . При таком измерителе связи не происходит деформации данных [30].

При этом модифицированный коэффициент корреляции любых двух монотонно возрастающих или монотонно убывающих функций всегда будет равен единице. В этом случае целесообразно перейти к анализу разностей второго (или даже более высокого) порядка. Получаемый коэффициент можно назвать модифицированным коэффициентом корреляции второго (или более высокого) порядка.

Корреляционная связь между рядами динамики исследуемых показателей может изменяться во времени (становиться слабее или сильнее). Одним



из способов выявления динамики этой связи, является использование рекуррентного (адаптивного) коэффициента корреляции:

$$r_t(\alpha) = \frac{s_t}{d_t}, \quad (5.2)$$

где  $s_t = (1 - \alpha)s_{t-1} + \alpha(\Delta x_{1t}\Delta x_{2t})$ ;

$d_t = (1 - \alpha)d_{t-1} + \alpha|\Delta x_{1t}\Delta x_{2t}|$ ;

$t = 1, 2, \dots, T$ ;

$0 < \alpha < 1$ .

Здесь  $s_t$  и  $d_t$  являются экспоненциально-взвешенными скользящими средними произведений приростов и абсолютных произведений приростов двух рядов. Параметр  $\alpha$  – постоянная сглаживания или параметр адаптации. Очевидно, что и здесь  $-1 < r_t(\alpha) < 1$ . График  $r_t(\alpha)$  относительно времени показывает, как изменяется сила корреляционной связи на выборочном периоде [33].

Используют и так называемый адаптивный коэффициент знаковой корреляции приростов, позволяющий проследивать корреляцию знаков приращений переменных, который вычисляется по формуле:

$$r_t = (1 - \alpha)r_{t-1} + \alpha \frac{\Delta x_{1t}\Delta x_{2t}}{|\Delta x_{1t}\Delta x_{2t}|}, \quad 0 < \alpha < 1 \quad (5.3)$$

где в качестве  $r_0$  можно взять отношение

$$r_0 = \frac{\sum_{j=1}^{T_0} \Delta x_{1j}\Delta x_{2j}}{\sum_{j=1}^{T_0} |\Delta x_{1j}\Delta x_{2j}|}. \quad (5.4)$$

Для определения наилучшего значения  $\alpha$  можно рассматривать  $r_{t-1}(\alpha)$  как прогноз совпадения колебаний  $x_{1t}$  и  $x_{2t}$  в момент  $t$  и ввести ошибку этого прогноза в момент  $t$  в виде:

$$e_t = \frac{\Delta x_{1j} \Delta x_{2j}}{|\Delta x_{1j} \Delta x_{2j}|} - r_{t-1}(\alpha). \quad (5.5)$$

Тогда оптимальное значение  $\alpha$  находится путем минимизации функции потерь  $Q(\alpha)$ :

$$Q(\alpha) = \sum_{t=2}^T e_t^2, \quad (5.6)$$

где  $e_t$  – ошибка прогноза в момент  $t$ .

При необходимости таким же образом устанавливается наилучшее значение  $T_0$ . В этом случае функция потерь рассматривается как функция двух переменных  $Q(T_0, \alpha)$  [33].

Рассмотренные коэффициенты корреляции отражают силу взаимосвязи колебаний двух рядов в целом. Для измерения сила связи, например, положительных отклонений  $x_t$  с отклонениями  $z_t$  используют условный коэффициент корреляции:

$$r_c(x_t, z_t | A < x_t - E_x \leq B) = \frac{\sum (x_t - E_x)(z_t - E_z)}{\sum |(x_t - E_x)(z_t - E_z)|} | A < x_t - E_x \leq B, \quad (5.7)$$

где  $A, B$  – границы области амплитуд интересующих отклонений;

$E_x, E_z$  – уровни, относительно которых рассчитываются отклонения рядов  $x_t$  и  $z_t$  соответственно.

Точно также вычисляют коэффициент взаимосвязи отрицательных отклонений ряда  $x_t$  с отклонением  $z_t$ . В обоих случаях коэффициенты корреляции могут принимать неодинаковые значения.

Таким образом, особенности взаимосвязи двух показателей, измеренных во времени теперь не скрываются за одним значением коэффициента корреляции, как это предполагается в классической теории, а представляются двумя коэффициентами.

Преимущество такого подхода в тех случаях, когда одна переменная является контролируемой, а другая – непосредственно недоступной, но коррелированной с первой.

В пользу применения условного коэффициента корреляции можно привести следующее - при низких значениях амплитуды отклонений корреляционная связь может не просматриваться из-за влияния случайных помех, то есть быть незначительной. При больших отклонениях определенность этой связи может возрасти до детерминированности. Тогда при большой амплитуде отклонений одной переменной резко возрастает степень предопределенности области нахождения второй переменной. Легче определить и лаг одной переменной относительно другой [33].

Существует еще один метод исследования силы сопряженности двух переменных на основе оценки условной вероятности их связи. Вероятностный коэффициент корреляции рассчитывается по формуле:

$$r_p(x_t, z_t | A \leq x_t - E_x \leq B, C \leq z_t - E_z \leq D) = P(C \leq z_t - E_z \leq D | A \leq x_t - E_x \leq B), \quad (5.8)$$

где  $P$  – вероятность того, что  $C \leq z_t - E_z \leq D$  при условии  $A \leq x_t - E_x \leq B$ .

Областью существования вероятностного коэффициента корреляции является отрезок  $[0, 1]$ .

Оценку вероятностного коэффициента на основе статистических данных получают как отношение по следующей формуле:

$$\hat{r}_p(x_t, z_t | A \leq x_t - E_x \leq B, C \leq z_t - E_z \leq D) = \frac{n_{z|x}}{N_x}, \quad (5.9)$$

где  $n_{z|x}$  – число наблюдений, когда  $z_t - E_z$  оказывалось в пределах  $C \leq z_t - E_z \leq D$  при  $A \leq x_t - E_x \leq B$ ;

$N_x$  – общее число наблюдений в выборке, удовлетворяющих условию  $A \leq x_t - E_x \leq B$ .

Обратное направление связи характеризуется величиной:

$$\hat{r}_P(z_t, x_t | C \leq z_t - E_z \leq D, A \leq x_t - E_x \leq B) = \frac{n_{x|z}}{N_z}. \quad (5.10)$$

Вероятностный коэффициент корреляции позволяет исследовать взаимосвязь переменных с вероятностной точки зрения [33].

Условный и вероятностный коэффициенты корреляции называют также парциальными коэффициентами корреляции, так как они способны отражать частичные свойства двух исследуемых временных рядов, наблюдаемые при определенных условиях.

## 5.2 Моделирование и прогнозирование показателей на основе многомерных временных рядов

Прежде чем переходить к построению многофакторных моделей прогнозирования необходимо исследовать многомерные временные ряды: на стационарность; компонентный состав и характер тренда. В зависимости от того, являются ли изучаемые ряды стационарными или нет, рядами TS (детерминированный тренд) и/или DS (стохастический тренд) в совокупности или в отдельности, делим ли мы изучаемые переменные на эндогенные и экзогенные или нет, будем использовать различные подходы к их моделированию.

Вспомним, что к TS-рядам относятся стационарные временные ряды и ряды с детерминированным трендом. В случае, когда ряды динамики изучаемых показателей являются стационарными можно строить регрессию непосредствен-

но по исходным рядам динамики, так как ложная регрессия возникает только для нестационарных временных рядов вследствие наличия стохастического и/или детерминированного тренда.

Для стационарных временных рядов можно помимо построения модели множественной регрессии, можно строить модели, где в качестве регрессоров будут выступать как лаговые значения результативной переменной, так и лаговые значения объясняющей переменной. Речь идет об авторегрессионных моделях с распределенными лагами. В литературе можно встретить аббревиатуры ADL или ARMAX [26,37,44].

Общий случай: модель с лаговыми значениями результативной и факторной переменных ADL (p,q):

$$\begin{aligned}
 y_t &= \theta + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \beta_0 x_t + \beta_1 x_{t-1} + \dots + \beta_q x_{t-q} + \varepsilon_t, \\
 t &= k+1, k+2, \dots, T \\
 k &= \max(p, q).
 \end{aligned}
 \tag{5.11}$$

Можно заменить значения процессов  $y_t$  и  $x$  в различные моменты времени на их значения в момент  $t-1$  и отклонения и после некоторых преобразований [37] получают следующее выражение:

$$\Delta y_t = \beta_0 \Delta x_t + \sum_{i=1}^{p-1} \delta_i \Delta y_{t-i} + \sum_{i=1}^{q-1} \gamma_i \Delta x_{t-i} - \alpha_p(1) \left[ y_{t-1} - \frac{\theta}{\alpha_p(1)} - \frac{\beta_q(1)}{\alpha_p(1)} \right] + \varepsilon_t. \tag{5.12}$$

Это представление модели с распределенными лагами называется моделью коррекции ошибками ЕСМ (или ее называют моделью с исправлением ошибки, модель корректировки ошибок). Название связано с тем, что выражение в скобках может трактоваться как отклонение от долгосрочного равновесия в момент  $t-1$ . Выражение в скобках – корректирующий член, подправляющий отклонениями от долгосрочного равновесия на предыдущем шаге. Таким образом, в одной модели объединяют долгосрочное и краткосрочное равновесие

Параметры модели оценивают методом максимального правдоподобия или МНК (если регрессоры и остатки некоррелированы и все переменные стационарные) [37].

Если ряды динамики исследуемых показателей являются рядами TS, то в этой ситуации можно вычистить детерминированные составляющие и строить регрессию по отклонениям от детерминированных трендов. В этом случае речь идет о моделировании краткосрочного равновесия или можно строить регрессию  $Y$  на  $X$  с включением в правую часть фактора времени (метод Фриша–Фоу):

$$y_t = a_0 + a_1 t + \theta x_t + \varepsilon_t, \quad (5.13)$$

где остатки являются белым шумом. Кстати, если между исследуемыми временными рядами есть «паразитная» (ложная) связь, то при включении в модель регрессора  $t$  приведет к тому, что параметр  $\theta$  перестанет быть значимым [9].

При построении регрессионной модели на основе нестационарных временных рядов может возникнуть эффект «ложной, паразитной регрессии». В литературе часто приводят пример связи между потреблением электроэнергии и рекордами прыжков в высоту [4, 27, 35]. Применение в этом случае регрессионного анализа к непретобразованным данным приводит к получению значимого уравнения регрессии. Связано это с наличием возрастающих/убывающих трендов в рядах динамики обоих показателей, но не реальной причинной зависимостью, хотя формальное применение регрессионного анализа зачастую дает статистически значимые оценки коэффициентов детерминации близким к единице. «Паразитная» связь между переменными может возникнуть, если переменные имеют в своем составе детерминированный и/или стохастический тренд [36, 49].

Подход, позволяющий напрямую работать с нестационарными рядами, основан на концепции коинтеграции.

Коинтеграция – долгосрочное равновесие между трендами временных рядов. При этом коинтеграция возможна только между DS-рядами, то есть рядами со стохастическим трендом [4, 33].

В частном случае признаками коинтеграции переменных являются:

- 1) интегрированность первого порядка индивидуальных временных рядов;
- 2) интегрированность нулевого порядка линейной комбинации временных рядов.

Если первые разности ряда стационарны, то говорят, что этот ряд является интегрированным первого порядка или  $I(1)$ . Если стационарны вторые разности ряда, то это интегрированный ряд второго порядка, или  $I(2)$ . Если же ряд является стационарным, то он называется интегрированным рядом нулевого порядка, или  $I(0)$  [33, 35].

Таким образом, проверке коинтегрированности предшествует этап проверки характера тренда. Как правило, для этих целей используют критерии Дики-Фуллера (простой и расширенный), которые проверяют гипотезу о том, что ряд динамики типа DS, альтернативная гипотеза – ряд типа TS. Указанные критерии «хорошо работают» только в случае отсутствия структурных сдвигов, сезонности, а также в условиях гомоскедастичности и не автокоррелированности остатков [2, 3, 35].

Пусть мы имеем  $N$  временных рядов  $x_{1t}, \dots, x_{Nt}$ , каждый из которых является интегрированным порядка 1. Если существует такой вектор  $\beta = (\beta_1, \dots, \beta_N)^T$ , отличный от нулевого, для которого  $\beta_1 x_{1t} + \dots + \beta_N x_{Nt} \sim I(0)$  - стационарный ряд, то говорят, что эти ряды коинтегрированы, такой вектор  $\beta$  называется коинтегрирующим вектором [27, 28].

У коинтегрированной системы  $I(1)$  рядов может быть несколько линейно независимых коинтегрирующих векторов.

Ранг коинтеграции  $r$  - максимальное количество линейно независимых коинтегрирующих векторов.

Для коинтегрированной системы из  $N$  рядов  $r = 1, \dots, N - 1$ . Формально, если ряды некоинтегрированы, то  $r = 0$ .

Для определения ранга коинтеграции можно воспользоваться процедурой Йохансена [28].

Проверить коинтегрированность рядов можно по ряду остатков. Выдвигается гипотеза о некоинтегрированности (о наличии единичного корня в остатках). Проверка осуществляется либо на основе расчета статистики Дарбина-Уотсона, критические значения здесь будут отличаться от критических значений обычной статистики DW, либо применения расширенного критерия Дики-Фуллера (ADF-тест), но с уточненными для этого случая критическими значениями (поскольку остатки вычисляются уже после оценивания коэффициентов модели, МНК «делает» остатки «похожими» на стационарные и слишком часто ошибочно принимается гипотеза о наличии коинтеграции) [33].

Если нулевая гипотеза отвергается, то вектор берется в качестве оцененного коинтегрирующего вектора и модель отражает действительную связь и применима для прогнозирования.

При изучении динамики структуры показателей, характеризующих социально-экономические процессы, зачастую применяют показатели структурных сдвигов, однако существует подход, позволяющий, во-первых, проследить динамику структуры и более чем за два года, во-вторых, ранжировать структурные составляющие по их величине (доле), в-третьих, включать непрерывно вновь поступающую информацию [41]. Данный подход основан на аппроксимации зависимости  $\tilde{y}_t$  от  $\tilde{x}_{t1}, \dots, \tilde{x}_{tk}$  моделью регрессии:

$$\tilde{y}_t = \beta_1 \tilde{x}_{t1} + \beta_2 \tilde{x}_{t2} + \dots + \beta_k \tilde{x}_{tk} + \varepsilon_t, \quad t = \overline{1; T}, \quad (5.14)$$

где  $\tilde{y}_t = \frac{y_t - \bar{y}}{S_{y_i}}$ ;  $\tilde{x}_{ti} = \frac{x_{ti} - \bar{x}_i}{S_{x_{ii}}}$  - стандартизованные переменные ( $i = \overline{1, k}$ );

$\varepsilon_t$  - регрессионные остатки в момент времени  $t$ ;

$\beta_i$  - стандартизованные коэффициенты ( $i = \overline{1, k}$ ), которые оцениваются рекуррентным методом наименьших квадратов, позволяющим подключать к оценке «вновь» поступающие данные в каждый последующий момент времени  $t$ , и таким образом, отслеживать динамику коэффициентов  $\beta(t)$  [41].



Основное достоинство стандартизованных коэффициентов регрессии заключается в том, что они сравнимы между собой. Сравнивая их, друг с другом, можно ранжировать факторы по силе их воздействия на результат [43].

### 5.3 Метод Зельнера при оценивании коэффициентов группы регрессионных уравнений

Зачастую при моделировании сложных социально-экономических процессов и систем осуществляют построение группы регрессионных уравнений, среди множества объясняющих переменных которых в разных уравнениях обязательно есть и различные. В ситуации, когда регрессионные остатки различных уравнений группы коррелированы между собой, целесообразно воспользоваться следующими соображениями, высказанными Зельнером: при условии коррелированности между собой регрессионных остатков различных уравнений, гомоскедастичности и некоррелированности остатков каждого из уравнений применение обобщенного метода наименьших квадратов (ОМНК), приводит к асимптотически более эффективным оценкам, чем в случае отдельного их оценивания с помощью МНК [18].

Рассматривается модель регрессии в виде группы уравнений:

$$\begin{aligned} y_i &= X_i b_i + \varepsilon_i, \quad i = 1, \dots, m \\ \Sigma_{\varepsilon_i} &= \sigma_i^2 E = \sigma_{ii} E, \end{aligned} \tag{5.15}$$

где  $y_i$  - вектор размерности  $T \times 1$ ;

$X_i$  - матрица порядка  $T \times k$ ;

$b_i$  - вектор размерности  $k_i \times 1$ ;

$\varepsilon_i$  - вектор размерности  $T \times 1$ ,  $M(\varepsilon_i) = 0$ .

По определению матрица ковариаций для возмущающего члена  $\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_m \end{pmatrix}$

$$\text{равна: } \Sigma = M(\varepsilon\varepsilon^T) = \begin{pmatrix} M(\varepsilon_1\varepsilon_1^T) & M(\varepsilon_1\varepsilon_2^T) & \dots & M(\varepsilon_1\varepsilon_m^T) \\ M(\varepsilon_2\varepsilon_1^T) & M(\varepsilon_2\varepsilon_2^T) & \dots & M(\varepsilon_2\varepsilon_m^T) \\ \vdots & \vdots & & \vdots \\ M(\varepsilon_m\varepsilon_1^T) & M(\varepsilon_m\varepsilon_2^T) & \dots & M(\varepsilon_m\varepsilon_m^T) \end{pmatrix}.$$

$$M(\varepsilon_i\varepsilon_j) = \sigma_{ij}E, \quad i, j = 1, \dots, m, \quad (5.16)$$

$$\text{где } M(\varepsilon_{it}\varepsilon_{j\tau}) = \begin{cases} \sigma_{ij}, & t = \tau \\ 0, & t \neq \tau \end{cases}$$

Если  $i \neq j$ , то существует ненулевая корреляция между одновременными возмущениями и нулевая корреляция между возмущениями  $\varepsilon_i$  и  $\varepsilon_j$ . Подставим (5.16) в матрицу  $\Sigma$  и получим:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2m} \\ \vdots & \vdots & & \vdots \\ \sigma_{m1} & \sigma_{m2} & \dots & \sigma_{mm} \end{pmatrix} \otimes E = \Sigma_c \otimes E, \quad (5.17)$$

где  $E$  - единичная матрица порядка  $T \times T$ .

Применение обобщенного метода наименьших квадратов к модели (5.15) даст нам оценку

$$b = (X^T \hat{\Sigma}^{-1} X)^{-1} X^T \hat{\Sigma}^{-1} Y, \quad (5.18)$$

$$\hat{\Sigma}^{-1} = \hat{\Sigma}_c^{-1} \otimes E,$$

где  $\hat{\Sigma}$  - оценка матрицы  $\Sigma$ ,

$\hat{\Sigma}_c$  - оценка матрицы  $\Sigma_c$ .

Основная трудность практического использования оценивания (5.18) состоит в том, что матрица  $\Sigma$  неизвестна. Зельнер предложил применить к каждому уравнению обыкновенный метод наименьших квадратов и с помощью полученных таким образом остатков оценить элементы матрицы  $\Sigma_c$ . Для  $i$ -го уравнения остатки, соответствующие обыкновенному методу наименьших квадратов, будут находиться как

$$\varepsilon_i = (y_i - X_i b_i), \quad (5.19)$$

где  $b_i = (X_i^T X_i)^{-1} X_i^T y_i$ .

Поэтому  $\sigma_{ii}$  оценивается величиной  $s_{ii} = \frac{(y_i - X_i b_i)^T (y_i - X_i b_i)}{T - k_i}$ , а  $\sigma_{ij}$  оценивается как  $s_{ij} = \frac{(y_i - X_i b_i)^T (y_j - X_j b_j)}{\sqrt{T - k_i} \sqrt{T - k_j}}$ .

Эти оценки подставляют в матрицу  $\hat{\Sigma}_c$ , обращают эту матрицу и получают оценку матрицы [18].

Выигрыш в эффективности оценок, полученный в результате применения оценивания Зельнера, по сравнению с применением обыкновенного метода наименьших квадратов возрастает с увеличением корреляции между возмущениями, соответствующими различным уравнениям, и убывает при увеличении корреляции между различными множествами объясняющих переменных. Даже в том случае, когда истинное значение коэффициента корреляции между возмущениями в уравнениях равно нулю, выборочные значения остатков, соответствующих методу наименьших квадратов, могут обладать не столь малыми ковариациями, и применение обобщенного метода наименьших квадратов будет нецелесообразным. Полученные таким образом оценки будут обладать несколько большими стандартными ошибками по сравнению с оценками, найденными обыкновенным методом наименьших квадратов. Это справедливо даже для очень малых корреляций между возмущениями, если же эти корреляции растут, то эффективность

оценок при применении обобщенного метода наименьших квадратов существенно увеличивается [39].

## **5.4 Анализ взаимосвязей между показателями, характеризующими социально-экономические процессы, измеренными во времени**

### **5.4.1 Отбор факторов, оказывающих влияние на показатели объемов производства сельского хозяйства и промышленности региона**

При моделировании объемов производства базовых отраслей экономики региона существенным является выявление факторов оказывающих на них влияние. Актуальность такого рода исследований увеличивается в связи со вступлением России в ВТО, ведь существенных последствий от вступления ВТО следует ожидать в отраслях, на которых специализируется регион. Исследования показали, что отраслями специализации, например, Оренбургской области, являются промышленность и сельское хозяйство [39, 45]. Особый интерес представляло бы изучение влияния на объемы производства в промышленности и сельском хозяйстве показателей наиболее подверженные влиянию вступления России в ВТО, такие как объем импорта, экспорта и цены на бензин. Так условием ВТО является уравнивание внутренних и мировых цен на топливо и электроэнергию, а также транспортных тарифов [16, 17]. Ликвидация таможенных пошлин и квот на вывоз сырья повлечет за собой такой скачок цен на энергоносители, что может парализовать не только промышленность и сельское хозяйство, но и системы жизнеобеспечения населения (например, теплоснабжение). Кроме того обязательной нормой ВТО, которая не подвергается обсуждению, является отказ государства от избирательной поддержки отдельных отраслей, таких как сельское хозяйство, или предприятий посредством субсидий, дотаций, льготных цен, налогов и т.д. Снижение и последующая их отмена таможенных пошлин приведет к росту им-

порта, что особенно важно, учитывая неконкурентоспособность продукции ряда отраслей отечественной экономики [23, 31].

Выявление степени и характера влияния перечисленных факторов позволит получить количественные оценки последствий вступления России в ВТО для экономики региона, что позволит в свою очередь заранее предпринять меры по снижению негативных последствий от вступления России в ВТО.

Проведем оценивание нетрадиционных коэффициентов корреляции по следующим показателям:

- объем выпуска продукции сельского хозяйства млн.руб. (Prod\_cx);
- объем промышленного производства, млн.руб.(Prod);
- следующих показателей: оборот розничной торговли, млн.руб.(Obor);
- экспорт товаров, млн. дол. США (Ex);
- импорт товаров, млн. дол. США (Im);
- инвестиции в основной капитал, млн. руб. (Inv);
- стоимость бензина, руб. за тонну (Ben);
- индекс цен производителей на реализованную сельхозпродукцию (Ind\_cx).

Информационной базой служат ряды месячной динамики перечисленных показателей за период с 1998 по 2005 гг. по Оренбургской области (Приложение С, таблица С.1).

Для выявления факторов оказывающих влияние на показатели выпуска базовых отраслей экономики региона используем нетрадиционный корреляционный анализ, поскольку показатели представлены временными рядами. Применение нетрадиционного корреляционного анализа позволит измерить изменение силы связи переменных во времени и определить, с каким лагом включать в модели переменные.

Рассчитаем модифицированные коэффициенты корреляции с помощью электронной таблицы Excel, результаты представим в таблице 5.1.

Таблица 5.1 – Модифицированные коэффициенты корреляции

Переменная	Оборот розничной торговли (Obor)	Экспорт товаров (Ex)	Импорт товаров (Im)	Цена на бензин (Ben)	Индекс цен производителей сельхозпродукции (Ind_cx)
Объем промышленного производства (Prod)	0,730	0,322	0,182	0,149	-0,120
Объем производства сельхозпродукции (Prod_cx)	0,272	-0,121	-0,466	-0,548	0,187

Как видно из таблицы 5.1 между объемом промышленной продукции и оборотом розничной торговли существует сильная положительная корреляционная связь, а между объемом промышленной продукции и экспортом – умеренная положительная. Между объемом сельскохозяйственной продукции и ценой на бензин существует достаточно сильная отрицательная корреляционная связь, а между объемом сельскохозяйственной продукции и оборотом розничной торговли, а также индексом цен сельхозпроизводителей – умеренная положительная.

Расчет модифицированных коэффициентов корреляции можно проводить также для рядов с запаздыванием - коэффициенты корреляции между взятым с различными лагами исследуемыми переменными и остальными признаками. Результаты расчетов показали, что объем промышленной продукции сильно коррелирован с оборотом розничной торговли не только в текущий момент времени. Такой же вывод можно сделать о коррелированности объема продукции сельского хозяйства и цены на бензин.

Корреляционная связь может изменяться во времени. Для выявления динамики этой связи могут быть рассчитаны адаптивные коэффициенты корреляции. Было выявлено, что наиболее устойчивая положительная корреляционная связь наблюдается между объемом промышленной продукции и оборотом розничной торговли; объемом промышленной продукции и экспортом товаров, а между объемом сельскохозяйственной продукции и ценой на бензин наблюдается наиболее устойчивая отрицательная корреляционная связь.

Рассмотрим более подробно связь между исследуемыми рядами. Рассчитаем условный коэффициент корреляции для первых разностей рядов и возьмем в

качестве критического нулевой уровень (что эквивалентно исследованию исходных рядов, но взятию в качестве критического уровня предыдущего значения ряда). На рисунке 5.1 представлены условные коэффициенты с положительными приростами объемов продукции промышленности и сельского хозяйства.

Как видно из рисунка 5.1 условные коэффициенты корреляции между положительными приростами объема промышленной продукции и оборота розничной торговли, экспорта товаров, цены на бензин близки к единице. Это означает, что увеличение экспорта товаров и оборота розничной торговли тесно связано с увеличением объема производства промышленной продукции. Условные коэффициенты корреляции между положительными приростами объема сельхозпродукции и оборота розничной торговли, экспорта товаров, цены на бензин также высоки.

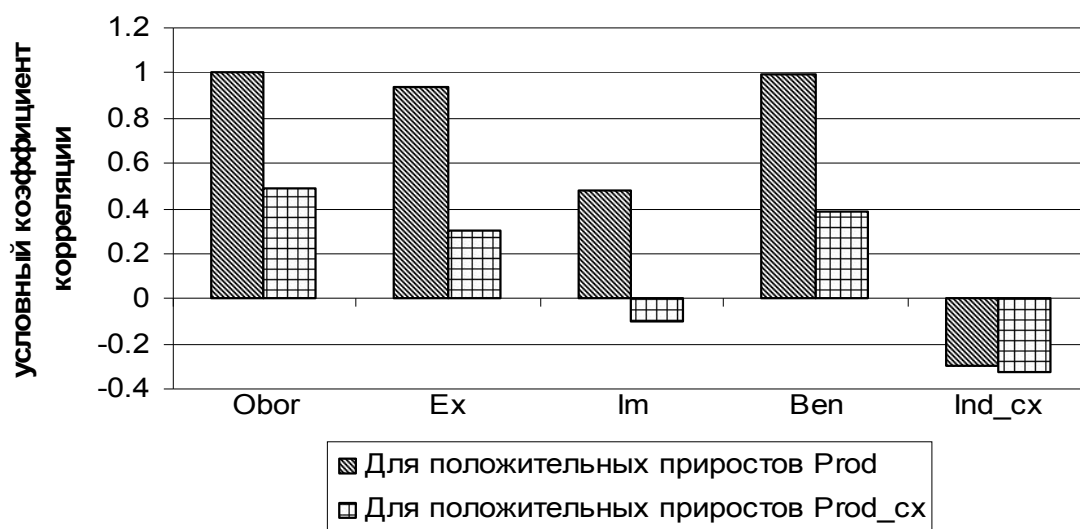


Рисунок 5.1 – Условные коэффициенты корреляции с положительными приростами объемов продукции промышленности и сельского хозяйства

На рисунке 5.2 представлены условные коэффициенты с отрицательными приростами объемов продукции промышленности и сельского хозяйства.

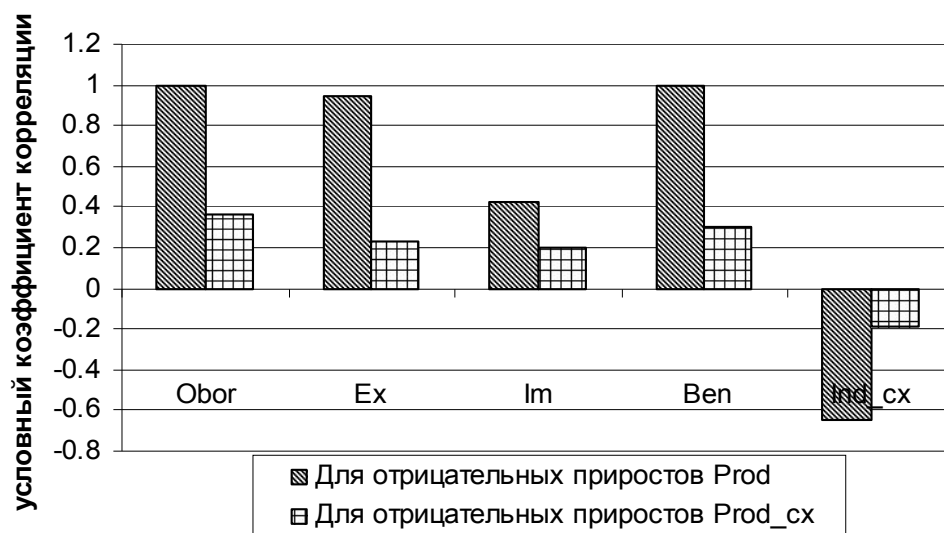


Рисунок 5.2 – Условные коэффициенты корреляции с отрицательными приростами объемов продукции промышленности и сельского хозяйства

Как видно из рисунка 5.2 условные коэффициенты корреляции между отрицательными приростами объема промышленной продукции и оборота розничной торговли, экспорта товаров, цены на бензин близки к единице. Это означает, что уменьшение экспорта товаров и оборота розничной торговли тесно связано с уменьшением объема производства промышленной продукции. Условные коэффициенты корреляции между отрицательными приростами объема сельхозпродукции и оборота розничной торговли, экспорта товаров, цены на бензин также высоки. Кроме того, условный коэффициент корреляции между отрицательными приростами объема сельхозпродукции и индекса цен сельхозпроизводителей близок к минус единице, что говорит о том, что снижение индекса цен тесно связано с увеличением объема производства сельхозпродукции.

Рассчитаем, с помощью электронной таблицы Excel, силу сопряженности исследуемых переменных на основе оценки условной вероятности их связи, то есть, вычислим вероятностные коэффициенты корреляции. На рисунке 5.3 представлены величины рассчитанных вероятностных коэффициентов.



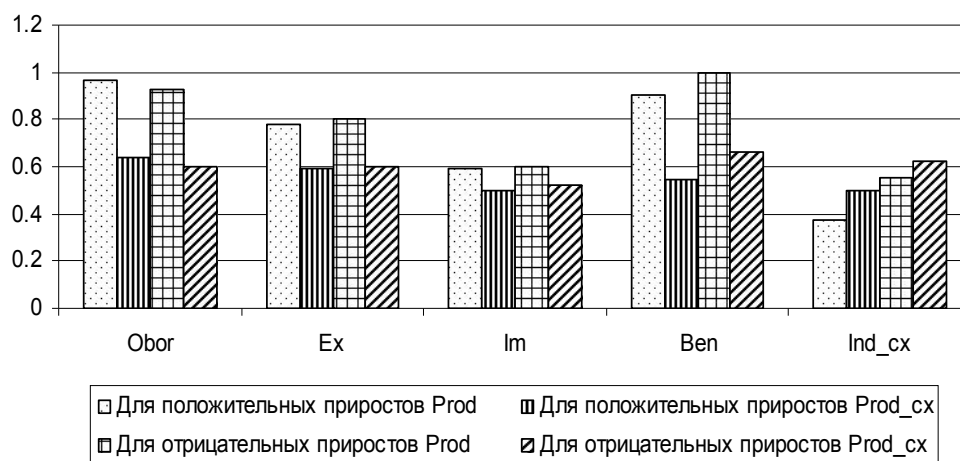


Рисунок 5.3 – Вероятностные коэффициенты корреляции между приростами объемов продукции промышленности и сельского хозяйства и остальных показателей

Согласно рисунку 5.3, вероятность увеличения объема промышленной продукции при условии увеличения оборота розничной торговли равна 0,97, при увеличении экспорта товаров – 0,78, импорта – 0,59. Вероятность уменьшения объема промышленной продукции при условии сокращения экспорта товаров немного выше и составляет 0,8. Что касается объема сельскохозяйственной продукции, то вероятность увеличения выпуска при увеличении оборота розничной торговли составляет 0,64, при увеличении экспорта – 0,59, при повышении индекса цен сельхозпроизводителей – 0,5. Вероятность снижения объема сельскохозяйственной продукции при повышении цены на бензин составляет 0,66.

Таким образом, был осуществлен отбор факторов оказывающих влияние на исследуемые показатели, при этом получена возможность определить, с каким лагом включать факторы в последующее регрессионное моделирование.

#### 5.4.2 Отбор факторов, оказывающих влияние на демографические и миграционные процессы в регионе

Говоря о демографических и миграционных процессах, имеющих сложный и многогранный характер, следует отметить многообразие связей между показателями

телями, характеризующими состояние этих процессов и тот факт, что они протекают под влиянием множества факторов, которое к тому же может усиливаться или ослабевать с течением времени [38].

Информационной базой служат ряды годовой динамики за период с 2003 по 2007 годы следующих показателей:

$y_{1,t}$  - показатель общей заболеваемости населения в момент времени  $t$  (‰);

$y_{2,t}$  - общий коэффициент смертности в момент времени  $t$  (‰);

$y_{3,t}$  - уровень внешней миграции в момент времени  $t$  (‰);

$y_{4,t}$  - уровень брачности населения в момент времени  $t$  (‰);

$y_{5,t}$  - общий коэффициент рождаемости в момент времени  $t$  (‰);

$x_{1,t}$  - уровень разводимости населения в момент времени  $t$  (‰);

$x_{2,t}$  - площадь жилищ, приходящаяся в среднем на одного жителя в момент времени  $t$  (кв.м);

$x_{3,t}$  - удельный вес населения в трудоспособном возрасте в момент времени  $t$  (‰);

$x_{4,t}$  - удельный вес населения старше трудоспособного возраста в момент времени  $t$  (‰);

$x_{5,t}$  - удельный вес учащихся, занимающихся во вторую и третью смены в момент времени  $t$  (‰);

$x_{6,t}$  - уровень официально зарегистрированной безработицы в момент времени  $t$  (‰);

$x_{7,t}$  - среднемесячная номинальная начисленная заработная плата работников в момент времени  $t$  (руб.);

$x_{8,t}$  - число пострадавших с утратой трудоспособности на 1 рабочий день и более в момент времени  $t$  (на 1000 работающих);

$x_{9,t}$  - выбросы загрязняющих веществ в атмосферный воздух от стационарных источников в момент времени  $t$  (тысяч тонн/км<sup>2</sup>).

Результаты расчета модифицированного коэффициента корреляции представлены в таблице 5.2.

Таблица 5.2 – Модифицированные коэффициенты корреляции

Переменная	$y_{1,t}$	$y_{2,t}$	$y_{3,t}$	$y_{4,t}$	$y_{5,t}$
$y_{1,t}$ - показатель общей заболеваемости	1	<b>0,660</b>	-0,413	<b>-0,601</b>	<b>-0,690</b>
$y_{2,t}$ - общий коэффициент смертности	<b>0,660</b>	1	<b>-0,622</b>	-0,325	-0,246
$y_{3,t}$ - уровень внешней миграции	-0,413	<b>-0,622</b>	1	<b>0,714</b>	<b>0,654</b>
$y_{4,t}$ - уровень брачности населения	<b>-0,601</b>	-0,325	<b>0,714</b>	1	<b>0,598</b>
$y_{5,t}$ - общий коэффициент рождаемости	<b>-0,690</b>	-0,246	<b>0,654</b>	<b>0,598</b>	1
$x_{1,t}$ - уровень разводимости населения	0,321	<b>0,527</b>	0,226	0,089	<b>-0,640</b>
$x_{2,t}$ - площадь жилищ, приходящаяся в среднем на одного жителя	-0,264	0,123	<b>0,721</b>	<b>0,614</b>	0,456
$x_{3,t}$ - удельный вес населения в трудоспособном возрасте	-0,258	-0,107	<b>0,557</b>	<b>0,591</b>	0,415
$x_{4,t}$ - удельный вес населения старше трудоспособного возраста	<b>0,612</b>	0,347	-0,379	0,412	<b>-0,566</b>
$x_{5,t}$ - удельный вес учащихся, занимающихся во вторую и третью смены	0,403	0,086	0,215	-0,246	0,115
$x_{6,t}$ - уровень официально зарегистрированной безработицы	0,155	0,407	<b>-0,601</b>	0,141	<b>-0,598</b>
$x_{7,t}$ - среднемесячная номинальная начисленная заработная плата работников	-0,212	-0,089	<b>0,624</b>	<b>0,663</b>	0,398
$x_{8,t}$ - число пострадавших с утратой трудоспособности на 1 рабочий день и более	<b>0,543</b>	0,245	0,112	-0,113	-0,012
$x_{9,t}$ - выбросы загрязняющих веществ в атмосферный воздух от стационарных источников	0,497	0,154	0,245	0,211	0,254

Из приведенных результатов следует: между общей заболеваемостью и удельным весом населения старше трудоспособного возраста, а также между общей заболеваемостью и числом пострадавших с утратой трудоспособности на 1 рабочий день и более наблюдается тесная положительная корреляционная связь ( $r_{\text{mod}} = 0,612$  и  $r_{\text{mod}} = 0,543$  соответственно). Можно отметить, что уровень внешней миграции имеет тесную положительную корреляционную связь со среднемесячной заработной платой работников ( $r_{\text{mod}} = 0,624$ ), а тесную отрицательную – с уровнем официально зарегистрированной безработицы ( $r_{\text{mod}} = -0,601$ ). Тесная

положительная корреляционная связь наблюдается между уровнем брачности населения и площадью жилищ, приходящаяся в среднем на одного жителя ( $r_{\text{mod}} = 0,614$ ), а также между уровнем брачности населения и среднемесячной заработной платой работников ( $r_{\text{mod}} = 0,663$ ). Общий коэффициент рождаемости имеет тесную отрицательную корреляционную связь с уровнем разводимости населения ( $r_{\text{mod}} = -0,640$ ).

Выявим динамику связи между рассматриваемыми показателями, рассчитав модифицированный коэффициент корреляции с лагом с помощью электронной таблицы Excel. Результаты расчетов представим графически. На рисунке 5.4 представлен график изменения значения модифицированного коэффициента корреляции для общей заболеваемости и удельного веса учащихся, занимающихся во вторую и третью смены

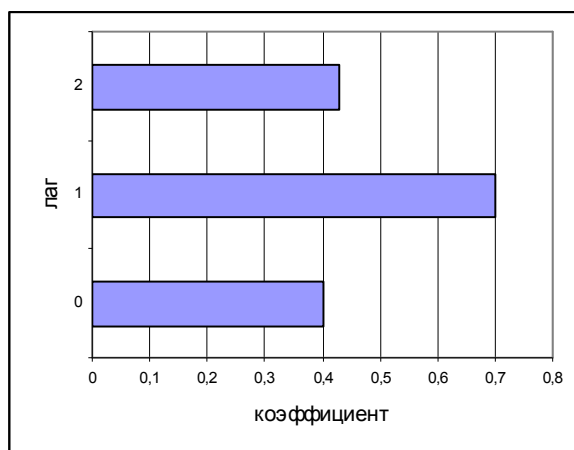


Рисунок 5.4 – Динамика значения модифицированного коэффициента корреляции для общей заболеваемости и удельного веса учащихся, занимающихся во вторую и третью смены

По рисунку 5.4 видно, что общая заболеваемость населения наиболее тесно коррелирует с удельным весом учащихся, занимающихся во вторую и третью смены, взятым с единичным лагом, так как модифицированный коэффициент корреляции принимает в таком случае свое наибольшее абсолютное значение.

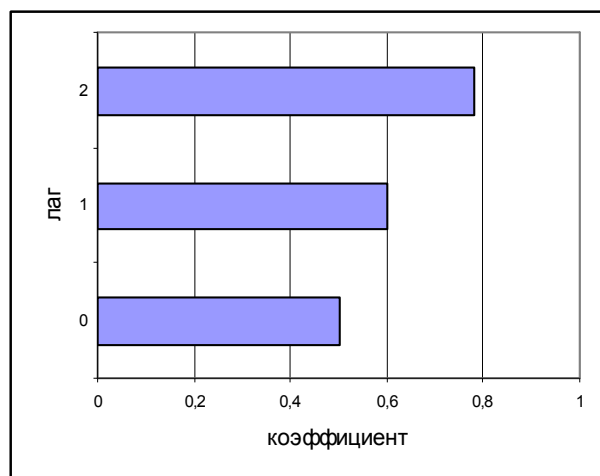


Рисунок 5.5 – Динамика значения модифицированного коэффициента корреляции для общей заболеваемости и выбросами загрязняющих веществ в атмосферный воздух от стационарных источников

По рисунку 5.5 видно, что между общей заболеваемости и выбросами загрязняющих веществ в атмосферный воздух от стационарных источников, существует тесная положительная связь. Причем модифицированный коэффициент корреляции принимает свое наибольшее абсолютное значение лаге, равным двум.

На рисунке 5.6 представлен график изменения значения модифицированного коэффициента корреляции для общего коэффициента смертности и уровнем официально зарегистрированной безработицы.

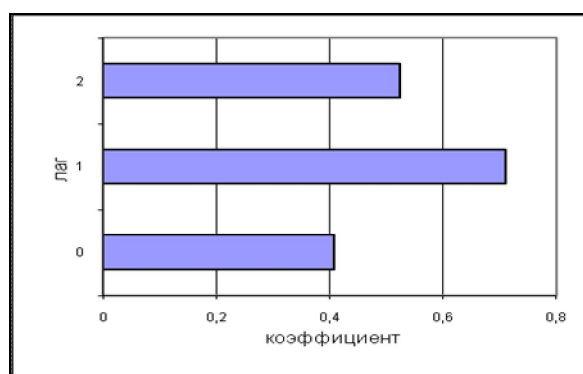


Рисунок 5.6 – Динамика значения модифицированного коэффициента корреляции для общего коэффициента смертности и уровнем официально зарегистрированной безработицы

Общий коэффициент смертности наиболее тесно коррелирует с уровнем официально зарегистрированной безработицы, взятым с единичным лагом (мо-

дифицированный коэффициент корреляции принимает в таком случае свое наибольшее абсолютное значение).

Аналогичная ситуация для показателей общего коэффициента рождаемости и площадью жилищ, приходящаяся в среднем на одного жителя (рисунок 5.7).

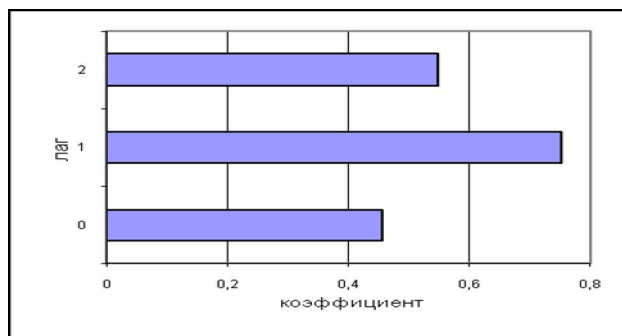


Рисунок 5.7 – Динамика значения модифицированного коэффициента корреляции для показателя общего коэффициента рождаемости и площадью жилищ, приходящаяся в среднем на одного жителя

В остальных случаях наибольшее значение модифицированный коэффициент корреляции принимает при нулевом лаге.

Используя результаты нетрадиционного корреляционного анализа, для выявления взаимосвязей факторов, характеризующих демографические и миграционные процессы можно строить многофакторные модели, например в форме систем одновременных регрессионных уравнений, в том числе на панельных данных.

#### **5.4.3 Анализ взаимосвязей между показателями, характеризующими инвестиционную активность, инвестиционный потенциал и инвестиционный риск**

Для выявления силы связи во времени между показателями, характеризующими инвестиционную активность, инвестиционный потенциал и инвестиционный риск применим нетрадиционный корреляционный анализ, поскольку

показатели, характеризующие данные категории, представлены временными рядами, которые к тому же в большинстве случаев являются нестационарными. Применение нетрадиционного корреляционного анализа позволит измерить изменение силы связи переменных во времени и определить, с каким лагом включать в модели переменные.

В нашем случае информационной базой послужили ряды годовой динамики за период с 1995 по 2006 гг. Результаты расчетов модифицированных коэффициентов корреляции представлены в таблице 5.3.

Таблица 5.3 – Модифицированные коэффициенты корреляции

Переменная	$y_{1,t}$	$y_{2,t}$	$y_{3,t}$	$y_{4,t}$	$y_{5,t}$	$y_{6,t}$	$y_{7,t}$
1	2	3	4	5	6	7	8
$y_{1,t}$ – объем инвестиций в основной капитал на душу населения, рублей	1	<b>0,811</b>	<b>0,989</b>	0,387	<b>-0,513</b>	<b>-0,598</b>	<b>-0,555</b>
$y_{2,t}$ – ввод в действие жилых домов на 1000 человек населения, кв.метров общей площади	<b>0,811</b>	1	0,487	0,459	<b>-0,588</b>	<b>-0,555</b>	<b>-0,753</b>
$y_{3,t}$ – объем промышленной продукции на душу населения, рублей	<b>0,989</b>	0,487	1	-0,402	-0,406	<b>-0,625</b>	<b>-0,505</b>
$y_{4,t}$ – производство (реализация) скота и птицы, центнеров на душу населения	0,387	0,459	-0,402	1	-0,111	-0,356	-0,196
$y_{5,t}$ – задолженность организаций по заработной плате, в процентах от общего фонда заработной платы	<b>-0,513</b>	<b>-0,588</b>	-0,406	-0,071	1	<b>0,531</b>	<b>0,699</b>
$y_{6,t}$ – удельный вес убыточных предприятий и организаций, в процентах от общего числа предприятий	<b>-0,598</b>	<b>-0,555</b>	<b>-0,625</b>	-0,293	<b>0,531</b>	1	<b>0,799</b>
$y_{7,t}$ – просроченная кредиторская задолженность предприятий, в процентах от общей задолженности	<b>-0,555</b>	<b>-0,753</b>	<b>-0,505</b>	-0,317	<b>0,699</b>	<b>0,799</b>	1
$x_{1,1,t}$ – уровень официально зарегистрированной безработицы, в процентах	<b>-0,922</b>	-0,277	-0,359	-0,233	0,074	0,205	0,120

Продолжение таблицы 5.3

1	2	3	4	5	6	7	8
$x_{1,2,t}$ – доля населения в трудоспособном возрасте в общей численности населения, в процентах	0,356	<b>0,598</b>	0,356	<b>0,599</b>	-0,269	-0,156	<b>-0,599</b>
$x_{1,3,t}$ – доля лиц моложе трудоспособного возраста, в общей численности населения, в процентах	0,211	0,111	0,245	0,378	-0,136	-0,178	-0,136
$x_{1,4,t}$ – среднегодовая численность работников, занятых в сельскохозяйственном производстве, человек	0,373	-0,072	-0,498	<b>0,738</b>	<b>0,581</b>	0,379	0,204
$x_{1,5,t}$ – среднегодовая численность работников, занятых в промышленности, человек	0,126	<b>-0,571</b>	<b>0,698</b>	<b>-0,564</b>	<b>-0,546</b>	-0,499	0,354

Из приведенных результатов следует: ввод в действие жилых домов, объем промышленной продукции, уровень официально зарегистрированной безработицы, сальдированный финансовый результат одного предприятия тесно коррелируют с инвестициями в основной капитал; между вводом в действие жилых домов и просроченной кредиторской задолженностью предприятий существует тесная отрицательная связь ( $r_{\text{mod}} = -0,753$ ); объем промышленной продукции коррелирует с удельным весом убыточных предприятий ( $r_{\text{mod}} = -0,625$ ) и оборотом розничной торговли ( $r_{\text{mod}} = -0,789$ ); между просроченной кредиторской задолженностью и удельным весом убыточных предприятий существует тесная положительная связь ( $r_{\text{mod}} = 0,799$ ).

Выявим динамику связи между рассматриваемыми показателями. Результаты расчетов представим графически. На рисунке 5.8 представлен график изменения значения модифицированного коэффициента корреляции для показателя число убыточных предприятий и задолженность организаций по заработной плате в Оренбургской области.



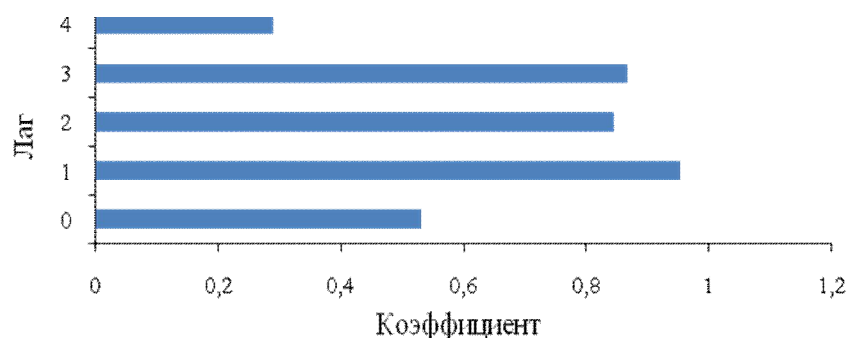


Рисунок 5.8 – Динамика значения модифицированного коэффициента корреляции для показателя число убыточных предприятий и задолженность организаций по заработной плате в Оренбургской области

Согласно рисунку 5.8, число убыточных предприятий наиболее тесно коррелирует с показателем задолженность организаций по заработной плате, взятым с единичным лагом (модифицированный коэффициент корреляции принимает в таком случае свое наибольшее абсолютное значение).

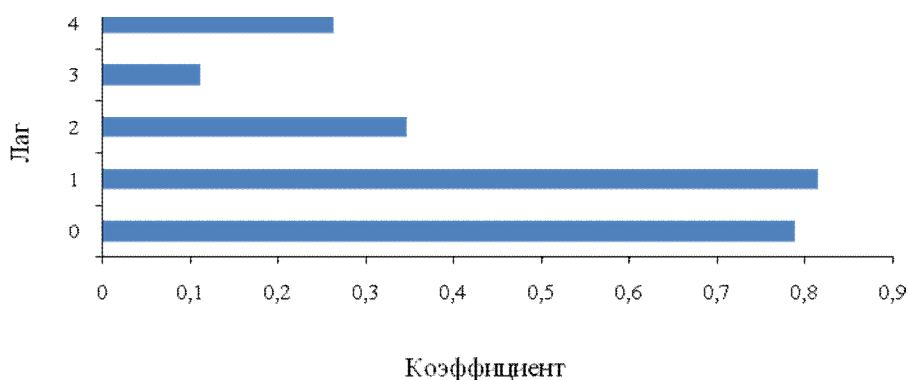


Рисунок 5.9 – Динамика значения модифицированного коэффициента корреляции для показателя просроченная кредиторская задолженность предприятий и число убыточных предприятий и организаций в Оренбургской области

Показатель просроченная кредиторская задолженность наиболее тесно коррелирует с показателем число убыточных предприятий и организаций с единичным лагом (рисунок 5.9).

В остальных случаях наибольшее значение модифицированный коэффициент корреляции принимает при нулевом лаге.

## 5.5 Моделирование и прогнозирование динамики структуры показателей, характеризующих социально-экономические процессы

### 5.5.1 Моделирование динамики структуры общего прироста населения

Общий прирост населения складывается из естественного прироста и механического. Естественный прирост населения характеризует разницу между родившимися и умершими, механический прирост – разницу между прибывшими и ушедшими. Структурно-динамический анализ общего прироста населения (рисунок 5.10) позволяет констатировать тот факт, что структура общего прироста населения на протяжении 37 лет претерпевала существенные изменения. В связи с чем, представляет интерес исследование структуры общего прироста населения на основе многомерных временных рядов.

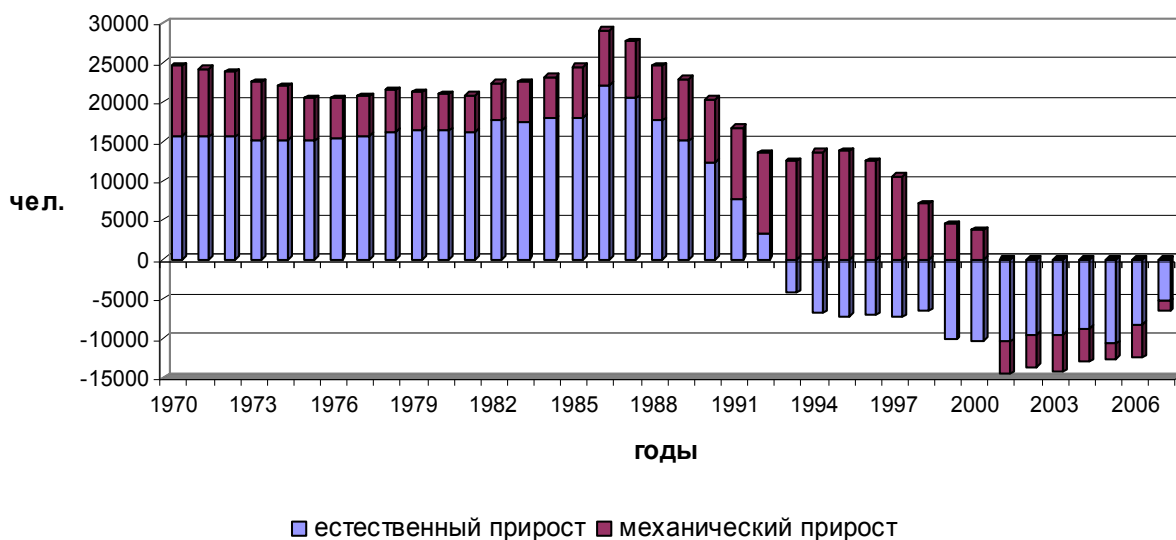


Рисунок 5.10 - Динамика естественного и механического прироста населения Оренбургской области за период с 1970 по 2007 гг.

Если до 1988 г. общий прирост населения складывался большей частью из естественного прироста, то в период с 1993 по 2000 г. прирост населения происходил исключительно за счет миграции, что связано, в том числе, с приграничным расположением области. Большой приток мигрантов, с одной стороны, в какой-то мере компенсирует естественную убыль населения, с другой стороны,

обостряет ситуации на рынке труда, что влечет за собой рост безработицы среди местного населения. За 1990–2004 гг. 81,1% составили прибывшие из Казахстана и республик Средней Азии. Приток беженцев и вынужденных переселенцев способствует распространению опасных заболеваний, росту преступности, незаконному обороту наркотиков, что создает угрозу демографической и национальной безопасности. Реальной угрозой демографической безопасности являются также нерациональные миграционные потоки между сельской и городской местностью, на долю которых в Оренбургской области приходится 76%. Необходимо отметить, что 90% в объеме чистой миграции из села в город составляет молодежь. В результате усиливается деформация возрастной структуры сельского населения. Это создает трудности в обеспечении сельского хозяйства рабочей силой.

На основе годовых данных о приросте населения за период с 1970 по 2007 гг.:

$y_t$  - общий прирост населения Оренбургской области в момент времени  $t$ ;

$x_{t1}$  - естественный прирост населения Оренбургской области в момент времени  $t$ ;

$x_{t2}$  - механический прирост населения Оренбургской области в момент времени  $t$ ;

причем  $y_t = x_{t1} + x_{t2}$ .

Перейдя к стандартизованным переменным, получим:  $\tilde{y}_t = \frac{\sigma_{x_{t1}}}{\sigma_{y_t}} \cdot \tilde{x}_{t1} + \frac{\sigma_{x_{t2}}}{\sigma_{y_t}} \cdot \tilde{x}_{t2}$ .

Аппроксимируя зависимость  $\tilde{y}_t$  от  $\tilde{x}_{t1}$  и  $\tilde{x}_{t2}$  моделью регрессии:

$$\tilde{y}_t = \beta_1 \tilde{x}_{t1} + \beta_2 \tilde{x}_{t2} + \varepsilon_t, \quad t = \overline{1; T}, \quad (5.20)$$

и оценивая коэффициенты  $\beta_1, \beta_2$  рекуррентным методом наименьших квадратов, получим их зависимость от  $t$  и изучим динамику структуры прироста населения Оренбургской области [41].

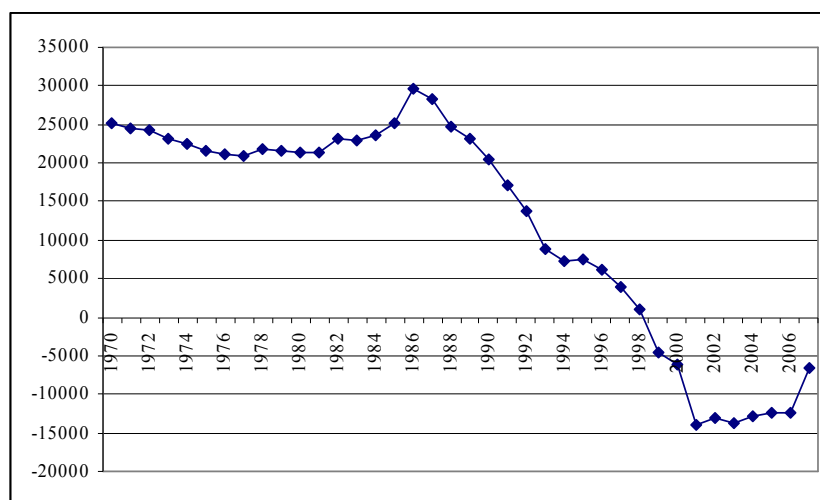


Рисунок 5.11 – Динамика общего прироста населения Оренбургской области за период с 1970 по 2007гг.

Согласно рисунку 5.11 на протяжении рассматриваемого периода времени наблюдалось снижение общего прироста населения, поэтому можно предположить наличие тренда в данном временном ряду.

Представим графически ряд первых разностей общего прироста населения (рисунок 5.12).

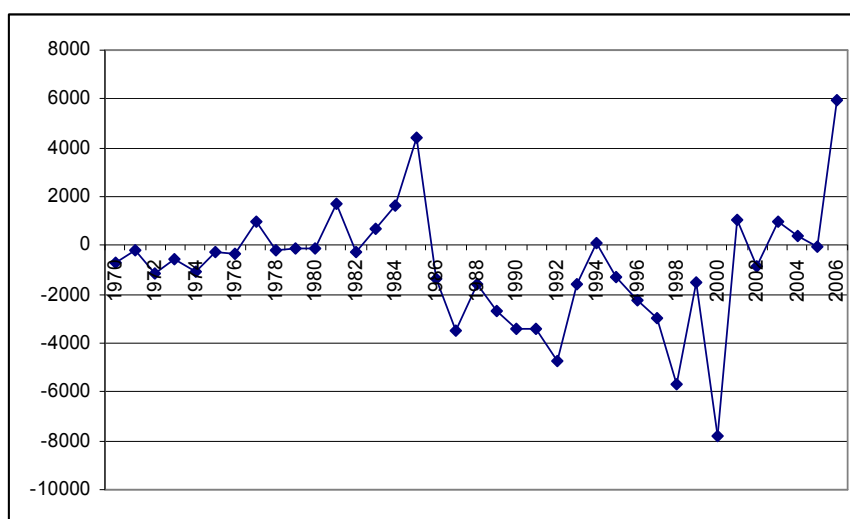


Рисунок 5.12 – Результат взятия первых разностей ряда динамики общего прироста населения Оренбургской области за период с 1970 по 2007гг.

В соответствии с рисунком 5.12 можно предположить отсутствие тенденции в ряду динамики после взятия первых разностей.

Визуальный анализ остальных рядов динамики и рядов первых разностей (приложение У) также позволил сделать предположение о присутствии тренда в

исходных временных рядах, которое было проверено использованием критерия Дикки-Фуллера (таблица 5.4) (расчеты проводились в ППП Eviews).

Таблица 5.4 – Результаты проверки гипотезы об отсутствии тренда в рядах динамики исследуемых показателей

Ряд	Значение статистики	Вывод
$y_t$	-1,35 (критич=-2,27)	нестационарный
$y_t - y_{t-1}$	-9,84 (критич=-2,27)	стационарный
$x_{1t}$	-1,15 (критич=-2,27)	нестационарный
$x_{1t} - x_{1,t-1}$	-10,21 (критич=-2,27)	стационарный
$x_{2t}$	-0,22 (критич=-2,27)	нестационарный
$x_{2t} - x_{2,t-1}$	-7,44 (критич=-2,27)	стационарный

Согласно таблице исследуемые временные ряды являются интегрированными первого порядка. Если после оценивания регрессионные остатки окажутся стационарными, то ряды будут коинтегрированными. С помощью расширенного критерия Дикки–Фуллера (ADF) доказана стационарность регрессионных остатков (таблица 5.5) (расчеты проводились в ППП Eviews).

Таблица 5.5 - Проверка регрессионных остатков на стационарность

Значение ADF	Уровень значимости	Критическое значение ADF
ADF = -4,02	0,01	-3,64
ADF= -4,02	0,05	-2,95
ADF= -4,02	0,1	-2,61

Найденные рекуррентным методом наименьших квадратов значения стандартизованных коэффициентов модели (5.20), отражающие динамику структуры прироста населения Оренбургской области за период 1985-2007 гг., так как в этот период наблюдались существенные изменения в характере развития данных, представлены в таблице 5.6.

Таблица 5.6 – Доли естественного и механического прироста в общем приросте населения Оренбургской области

Год	Доля естественного прироста	Доля механического прироста
1985	0,7241	0,2754
1986	0,7284	0,2711
1987	0,7342	0,2652
1988	0,7333	0,2661
1989	0,7291	0,2701
1990	0,7230	0,2768
1991	0,7117	0,2879
1992	0,6917	0,3077
1993	0,6414	0,3582
1994	0,5765	0,4233
1995	0,5123	0,4874
1996	0,4441	0,5557
1997	0,3473	0,6522
1998	0,0126	0,9870
1999	0,0569	0,9426
2000	0,1055	0,8939
2001	0,1247	0,8747
2002	0,1416	0,8581
2003	0,1574	0,8422
2004	0,1724	0,8274
2005	0,1909	0,8085
2006	0,1924	0,8070
2007	0,1977	0,8021

Представим графически изменения каждого из коэффициентов.

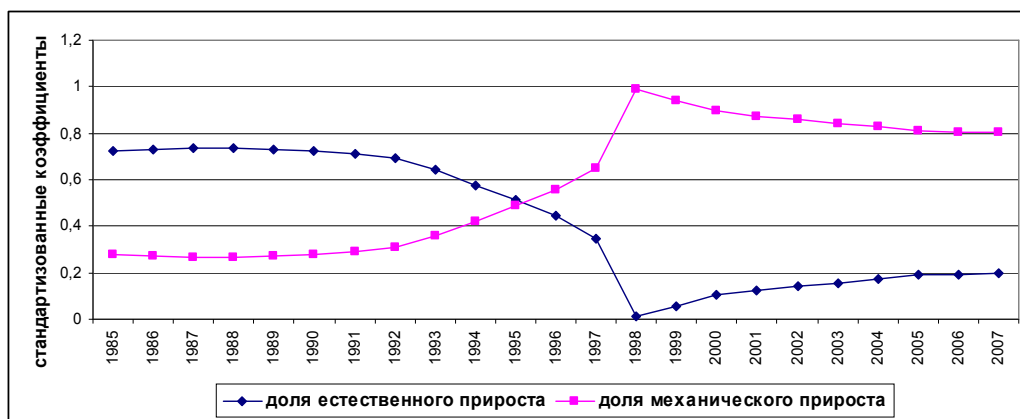


Рисунок 5.13 - Динамика доли естественного и механического прироста в общем приросте населения Оренбургской области

Согласно полученным результатам, можно сделать вывод, что за период с 1970 по 1995 год общий прирост населения Оренбургской области происходил за счет естественного прироста. Если за период с 1970 по 1985 гг. на долю естественного прироста приходилось в среднем 72,41%, на долю механического прироста лишь 27,54%, то за период с 1970 по 1995 гг. их доли практически сравнялись, составив в среднем 51,23% - доля естественного прироста, 48,74% - доля механического прироста в общем приросте населения. После чего общий прирост населения происходил исключительно за счет механического прироста, на его долю в среднем за период с 1970 по 1998 гг. приходилось 98,7%. На сегодняшний день на долю естественного прироста населения в среднем приходится лишь около 20% общего прироста населения, основная же часть (около 80%) приходится на механический прирост населения.

Негативные тенденции демографического развития и впредь будут обострять проблемы, связанные со старением населения, демографической нагрузкой и т.д. Поэтому, принимая во внимание такую безрадостную демографическую перспективу, необходимо проводить четкую иммиграционную политику, направленную на прием и интеграцию мигрантов в социальную сферу и рынок труда. Поощрение мигрантов с помощью грамотно выстроенной социальной политики совместными усилиями органов государственной власти, местного самоуправления и общественных объединений способно отчасти исправить сложившуюся ситуацию [13]. Однако не следует забывать, что миграционный приток,

хотя и в какой-то мере компенсирует естественную убыль населения, но вместе с тем, приток беженцев и вынужденных переселенцев способствует оттоку капитала, распространению опасных заболеваний, росту преступности, незаконному обороту наркотиков, что создает угрозу национальной безопасности. Кроме того, неразумно уповать лишь на поощрение притока мигрантов в область как на средство преодоления кризисной демографической ситуации. Безусловно, международная миграция может смягчить остроту последствий демографического кризиса, способствовать решению отдельных демографических проблем, но не более того. Преодоление демографического кризиса может быть только результатом общего оздоровления общества, восстановления ценности и роли семьи, стабилизации уровня рождаемости, сокращения смертности и уже потом привлечения мигрантов. На это должны быть нацелены меры социально-демографической и миграционной политики.

### **5.5.2 Моделирование динамики структуры внешней миграции по основным отраслям экономики**

В отраслевой структуре занятости в Оренбургской области за 2003-2007 гг. произошли некоторые изменения: сократилась занятость в промышленности на 18 % при параллельном наращивании занятости населения в сфере строительства – на 30 %, в торговле - на 35 % и в сельском и лесном хозяйстве – на 7% [14]. Специалисты рабочих профессий остаются самыми востребованными на рынке труда Оренбурга. По информации городского Центра занятости населения, из всех вакансий, вошедших в базу данных в за 2007 г., 78 % составили предложения по рабочим профессиям, таким как токарь, сварщик, плотник, штукатур-маляр, каменщик и другие [14,15]. По той же информации, две трети из числа обратившихся в Центр занятости, имели рабочие специальности. В случае, если бы предлагаемые работодателями условия труда устраивали ищущих работу, они были бы уже трудоустроены. Основной причиной отказа потенциальных работ-



ников заключить трудовое соглашение остается сравнительно низкая заработная плата специалистов рабочих профессий на предприятиях города. В связи с этим создаются объективные предпосылки для миграционного прироста иностранных работников, которые готовы работать за низкую заработную плату и в «не престижных сферах экономики», где не желает работать местное население.

Численность иностранных работников, осуществляющих трудовую деятельность на законных основаниях в Оренбургской области, в 2007 г. по сравнению с 2003 г. увеличилась в 5,7 раз и составила 10293 человек. Средний темп прироста иностранных трудящихся в области за 2003 – 2007 гг. составил 53,15 %. Коэффициент иммиграции рабочей силы в 2007 году составил 4,843 ‰, в то время как в 2003 году он составлял 0,858 ‰.

Заметим, что Оренбургская область является привлекательным для трудящихся мигрантов еще и тем, что в регионе в 2007 г. наблюдался самый низкий уровень безработицы (0,8%) по сравнению с другими регионами Приволжского федерального округа. Максимальный уровень безработицы зарегистрирован в Ульяновской области – 4,2%. В Самарской и Саратовской областях уровень безработицы составил 1,6%, в Татарстане – 1,3%, в Башкортостане - 1,4% [13]. Тем не менее, в Оренбуржье, мигрантам очень сложно устроиться по специальностям. Предлагаемые службами занятости должностные вакансии не соответствуют профилю мигрантов, как правило, низкооплачиваемые и, зачастую там, где они имеются, отсутствует жилье. Территориальный орган Минфедерации России в Оренбургской области регулярно проводит выборку среди тех вакансий, которые могут привлечь мигрантов (например, с предоставлением общежития, служебной квартиры или другого жилья). Однако такие места существуют в отдаленных районах области и также малопопулярны.

Между регионами, принимающими иностранных работников, имеются существенные различия в структуре и занятости. Так, в Приморском крае наибольший удельный вес занятых иностранных рабочих приходится на торговлю, коммерцию и общественное питание, что в значительной мере объясняется высокой активностью здесь китайских предпринимателей, в Свердловской области

– на коммерческую деятельность по обеспечению рынка, Волгоградской области – на сельское хозяйство, в Иркутской области – на геологию и горнодобывающую промышленность. В Москве и Московской области, где концентрируются финансовые, коммерческие, культурные и прочие функции, около половины иностранцев работают в строительстве, примерно каждый пятый – в торговле.

В 2003 году иностранная рабочая сила в Оренбургской области направлялась в основном в сферу строительства, и ее удельный вес составлял 37,67 % от всех отраслей экономики (рисунок 5.14). Второе место в отраслевой структуре занимала промышленность (24,61 %), третье - сельское хозяйство (19,65%), а четвертое - коммерческая деятельность, в том числе торговля (15,74 %).

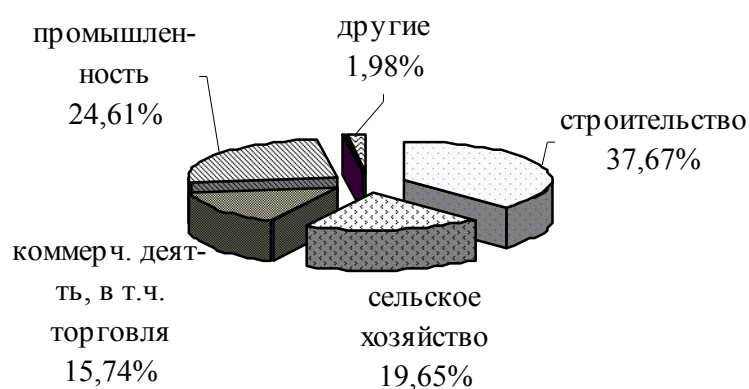


Рисунок 5.14 – Отраслевая структура зарегистрированной иностранной рабочей силы в Оренбургской области в 2003 г.

Анализ структуры занятости зарегистрированных иностранных работников в 2007 г. показывает, что прибывающие в область трудовые зарегистрированные мигранты направлялись, прежде всего, в сферу строительства (74,95 %), где достаточно высокая оплата труда и большой фронт работы (рисунок 5.15).

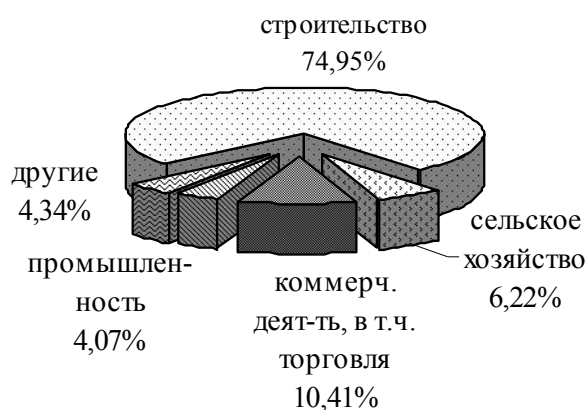


Рисунок 5.15 – Отраслевая структура зарегистрированной иностранной рабочей силы в Оренбургской области в 2007 г.

Торгово-коммерческой деятельностью в 2007 г. занимались 10,41 % всех зарегистрированных иностранных работников. Работа иностранных граждан в сельском хозяйстве (6,22%) носит сезонный характер, они привлекаются по выращиванию овощей в весеннее – летний период. Гораздо меньшее применение иностранные рабочие находят в социальной сфере. Удельный вес таких отраслей в совокупности составил в 2007 году 0,23 % от общей численности зарегистрированных иностранных трудящихся.

Отследим изменение отраслевой структуры занятости иностранных работников в 2007 г. по сравнению с 2003 г. с помощью коэффициентов структурных сдвигов (таблица 5.7)

Таблица 5.7 - Коэффициенты структурных сдвигов зарегистрированных иностранных работников в 2007 г. по сравнению с 2003 г.

Основные отрасли экономики	2003 год, %	2007 год, %	Относительный коэффициент структурных сдвигов	Абсолютный коэффициент структурных сдвигов, %
Строительство	37,67	74,95	1,9897	37,28
Сельское хозяйство	24,61	6,22	0,3164	-13,43
Коммерческая деятельность	15,74	10,41	0,6617	-5,33
Промышленность	24,61	4,07	0,1654	-20,54

С 2003 по 2007 гг. наблюдается тенденция снижения доли иностранной рабочей силы, занятой в промышленности (в 6 раз), в сельском хозяйстве (в 3

раза) и в коммерческой деятельности (в 1,5 раз) при параллельном наращивании темпов занятости в строительстве (в 2 раза). Подобные сдвиги в структуре занятости в значительной мере являются результатом падения объемов производства, следствием усиления кризисных явлений в экономическом развитии региона и страны.

Если рассмотреть различия рода деятельности по отраслям экономики трудовых мигрантов из стран ближнего и дальнего зарубежья, то можно сказать, что к 2007 году иностранные рабочие из стран бывшего СНГ охватили практически все отрасли экономики, вытолкнув трудовых мигрантов из стран дальнего зарубежья. Так, в сфере промышленности, доля зарегистрированных трудовых мигрантов из ближнего зарубежья, составляет 95,94 %, соответственно на долю мигрантов из дальнего зарубежья приходится 4,05 %. В сфере строительства на законных основаниях трудятся 96,01 %, мигрантов из стран СНГ, а в отрасли сельского хозяйства их доля составляет 100 %. В торговле и коммерческой деятельности удельный вес иностранных работников из стран ближнего зарубежья составляет 84,89 %, следовательно, 15,11 % составляют мигранты из дальнего зарубежья.

В 2007 г. в сфере строительства в основном были заняты иностранные работники из Узбекистана (60,89 %), в сфере промышленности – иностранные работники также из Узбекистана (47,25 %) и из Таджикистана (28,88%). Коммерческой деятельностью занимались 33,78 % трудящихся мигрантов из Таджикистана и 35,17 % - из Узбекистана. В сфере сельского хозяйства в основном занимались иностранные работники из Узбекистана (87,97%).

На основе месячных данных о числе иностранных работников Оренбургской области, занятых в основных отраслях экономики, за период с января 2003 г. по декабрь 2007 г.:

$q_t$  – общее количество работающих в Оренбургской области иностранных граждан на момент времени  $t$ ;

$z_{t1}$  – количество иностранных граждан, работающих в сфере строительства на момент времени  $t$ ;

$z_{t2}$  – количество иностранных граждан, занятых коммерческой деятельностью, в том числе торговлей, на момент времени  $t$ ;

$z_{t3}$  – количество иностранных граждан, работающих в сфере промышленности на момент времени  $t$ ;

$z_{t4}$  - количество иностранных граждан занятых в сельском хозяйстве на момент времени  $t$ ;

Аппроксимировав зависимость  $\tilde{q}_t$  от  $\tilde{z}_{t1}, \dots, \tilde{z}_{t4}$  ( $\tilde{q}_t, \tilde{z}_{t1}, \dots, \tilde{z}_{t4}$  - стандартизованные переменные) моделью регрессии, изучим динамику структуры зарегистрированных иностранных работников, занятых в основных отраслях экономики [41]:

$$\tilde{q}_t = \beta_1 \tilde{z}_{t1} + \beta_2 \tilde{z}_{t2} + \beta_3 \tilde{z}_{t3} + \beta_4 \tilde{z}_{t4} + \varepsilon_t, \quad t = \overline{1; T}, \quad (5.21)$$

Визуальный анализ рядов динамики и рядов первых разностей позволил сделать предположение о присутствии тренда в исходных временных рядах, которое было проверено использованием критерия Дикки-Фуллера (таблица 5.8) (расчеты проводились в ППП Eviews).

Таблица 5.8 – Результаты проверки временных рядов на интегрированность первого порядка с помощью критерия Дикки-Фуллера

Ряд	Значение статистики	Вывод
$q_t$	-1,26 (критич=-2,54)	ряд нестационарный
$q_t - q_{t-1}$	-7,075 (критич=-2,54)	ряд стационарный
$z_{t1}$	-1,92 (критич=-2,54)	ряд нестационарный
$z_{t1} - z_{t-1,1}$	-9,350 (критич=-2,54)	ряд стационарный
$z_{t2}$	-1,92 (критич=-2,54)	ряд нестационарный
$z_{t2} - z_{t-1,2}$	-7,533 (критич=-2,54)	ряд стационарный
$z_{t3}$	-2,15 (критич=-2,54)	ряд нестационарный
$z_{t3} - z_{t-1,3}$	-6,247 (критич=-2,54)	ряд стационарный
$z_{t4}$	-1,89 (критич=-2,54)	ряд нестационарный
$z_{t4} - z_{t-1,4}$	-10,113 (критич=-2,54)	ряд стационарный

С помощью расширенного критерия Дикки-Фуллера (ADF) доказана стационарность регрессионных остатков (таблица 5.9).

Таблица 5.9 - Проверка регрессионных остатков на стационарность

Значение ADF	Уровень значимости	Критическое значение ADF
ADF = -4,54	0,01	-2,59
ADF = -4,54	0,05	-1,94
ADF = -4,54	0,1	-1,62

Рекуррентным методом наименьших квадратов найдены значения стандартизованных коэффициентов модели (5.21), отражающие динамику в отраслевой структуре занятости иностранных работников в 2007 г. (таблица 5.10).

Таблица 5.10 – Стандартизованные коэффициенты

Период времени в 2007 г.	Отрасли экономики			
	строительство	коммерч. деятельность	промышленность	сельское хозяйство
Январь	0,583	0,226	0,085	0,055
Февраль	0,570	0,232	0,088	0,054
Март	0,554	0,239	0,087	0,051
Апрель	0,550	0,243	0,085	0,068
Май	0,553	0,244	0,084	0,072
Июнь	0,583	0,226	0,080	0,065
Июль	0,637	0,189	0,067	0,072
Август	0,674	0,152	0,058	0,074
Сентябрь	0,699	0,143	0,052	0,055
Октябрь	0,719	0,130	0,047	0,052
Ноябрь	0,730	0,120	0,047	0,051
Декабрь	0,742	0,111	0,042	0,048

Еще раз получили подтверждение, что первое место в отраслевой структуре иностранных трудящихся занимает сфера строительства. В 2007 г. удельный вес иностранных работников в сфере строительства имел тенденцию к увеличению и на декабрь 2007 г. составил 74,2%. Доля трудящихся мигрантов в сфере промышленности, в течение 2007 г., имела тенденцию к уменьшению, а доля иностранных работников в сфере сельского хозяйства существенно не изменялась. С января по декабрь 2007 г. удельный вес иностранных работников, занятых в коммерческой деятельности, снизился в 2 раза. Это объясняется тем, что вступило в силу Постановление Правительства РФ от 15.11.2006 г. "Об установлении на 2007 год допустимой доли иностранных работников, используемых хозяйствующими субъектами, осуществляющими деятельность в

сфере розничной торговли на территории Российской Федерации". Этот документ регламентирует трудовую деятельность иностранных граждан в сфере розничной торговли. Согласно постановлению, на первом этапе, с 15 января по 1 апреля, количество иностранных граждан на рынках и предприятиях розничной торговли должно было быть сокращено до 40% от общей численности работников, а с 1 апреля там вообще не должно было остаться ни одного работающего иностранца.

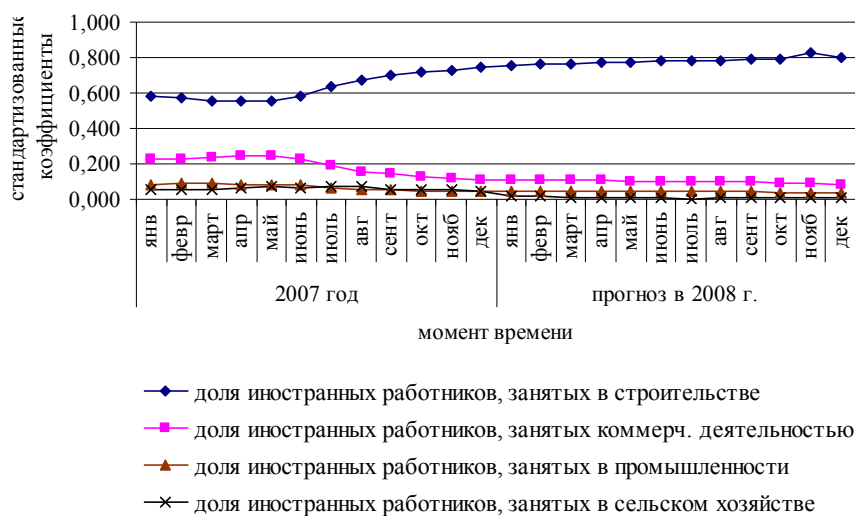


Рисунок 5.16 - Динамика стандартизованных коэффициентов, характеризующих структуру зарегистрированных иностранных работников по основным отраслям экономики в 2007 г. и их прогноз на 2008 г.

На рисунке 5.16 отражены зависимости стандартизованных коэффициентов от времени, характеризующих динамику структуры зарегистрированных иностранных работников, занятых в основных отраслях экономики в 2007 г. и прогноз стандартизованных коэффициентов на 12 месяцев вперед.

Как видно из рисунка 5.16, удельный вес трудящихся мигрантов, занятых в сфере строительства, в 2008 г. будет продолжать расти и концу прогнозного периода достигнет 80%. Вклад иностранных работников в общий объем работ, выполненных по виду деятельности «строительство» за 2003-2007 гг. увеличился в 25 раз и составил 10% от общего объема работ, что подтверждает прирост рабочей силы в реальный сектор экономики и, по-видимому, на сегодняшний день и в

краткосрочной перспективе внешняя миграция оказывает позитивное влияние на экономическую безопасность региона. Прогнозная сила, полученных результатов ограничена двенадцатимесечным промежутком времени. В связи с финансово - экономическим кризисом в стране и в мире, утверждать, что тенденция сохранится, безусловно, нельзя, в том числе по причине возможных административных ограничений на размеры миграционных потоков.

### **5.5.3 Моделирование динамики отраслевой структуры инвестиций**

Эффективность использования инвестиций в значительной степени зависит от их структуры. Структурная политика в современных условиях должна оказывать позитивное воздействие на динамику как экономики России, так и субъектов Федерации.

Анализ отраслевой структуры инвестиций в Оренбургской области показал, что с 1995 по 2006 гг. наибольшая доля капитальных вложений направлялась на поддержание и развитие промышленности (нефтедобывающей и газовой отраслей) и жилищно-коммунального хозяйства региона. Доля капитальных вложений в промышленность имела тенденцию к повышению с 38,5% в 1995 г. до 54,5% в 2006 г. Доля инвестиций в жилищно-коммунальное хозяйство в 2006 г. по сравнению с 1995 г. понизилась на 7,5 п.п. и составила 16,9% и 24,4% соответственно. Увеличились инвестиции в сельское и лесное хозяйства и составили 3,5% в 1995 г. и 6,5% в 2006 г. Малая доля инвестиций направлялась в образование, здравоохранение, физическую культуру и социальное обеспечение (Таблица 5.11).



Таблица 5.11 – Инвестиции в основной капитал в отрасли экономики Оренбургской области с 1995 – 2006гг. (в процентах к итогу)

Показатель	Год								
	1995	1999	2000	2001	2002	2003	2004	2005	2006
Инвестиции в основной капитал – всего	100	100	100	100	100	100	100	100	100
в том числе в отрасли:									
промышленность	38,5	46,8	47,3	51,4	42,8	49,5	55,5	57,2	54,5
сельское и лесное хозяйства	3,5	3,8	4,4	7,6	5,6	5,6	5,1	7,4	6,5
строительство	7,6	1,3	1,6	0,8	3,2	1,4	1,7	1,2	1,0
транспорт и связь	8,2	11,5	12,8	12,8	12,2	12,9	12,1	9,7	6,1
торговля и общественное питание, оптовая торговля продукцией производственно-технического назначения	5,0	1,6	4,7	3,0	5,6	3,2	2,3	2,2	8,7
жилищно-коммунальное хозяйство	24,4	22,3	16,9	14,9	16,9	18,3	16,3	12,8	16,9
здравоохранение, физическая культура и социальное обеспечение	2,0	2,5	1,3	1,4	1,1	1,2	0,9	1,1	3,2
образование	1,4	1,3	1,0	1,1	1,0	0,8	0,6	0,5	0,7
прочие	9,4	8,9	10	7	11,6	7,1	5,5	7,9	2,2

Обобщающие критерии для оценки меры существенности различий двух структур:

– коэффициента К. Гатева:  $K_{Г,1995-2006гг.} = 0,275$

– индекса Салаи:  $J_{с,1995-2006гг.} = 0,388$

– индекса Рябцева:  $J_{R,1995-2006гг.} = 0,198$

позволяют сделать вывод, что структура инвестиций в основной капитал по отраслям экономики в Оренбургской области за период 1995 – 2006 гг. имела существенный уровень различий.

Для изучения динамики отраслевой структуры инвестиций за период с 1995 по 2006 гг. на основе годовых данных об инвестициях:

$y_t$  – инвестиции в основной капитал Оренбургской области в момент времени  $t$  (всего по отраслям экономики), млн. руб.;

$x_{it}$  – инвестиции в промышленность Оренбургской области в момент времени  $t$ , млн. руб.;

$x_{t2}$  – инвестиции в сельское и лесное хозяйства Оренбургской области в момент времени  $t$ , млн. руб.;

$x_{t3}$  – инвестиции в строительство Оренбургской области в момент времени  $t$ , млн. руб.;

$x_{t4}$  – инвестиции в транспорт и связь Оренбургской области в момент времени  $t$ , млн. руб.;

$x_{t5}$  – инвестиции в жилищно–коммунальное хозяйство Оренбургской области в момент времени  $t$ , млн. руб.

построим модель регрессии:

$$i_{y_t} = \beta_1 i_{x_{t1}} + \beta_2 i_{x_{t2}} + \beta_3 i_{x_{t3}} + \beta_4 i_{x_{t4}} + \beta_5 i_{x_{t5}} + \varepsilon_t, \quad (5.22)$$

где  $i_{y_t}, i_{x_{t1}}, \dots, i_{x_{t5}}$  – стандартизованные переменные:  $i_{y_t} = \frac{y_t - \bar{y}}{S_{y_t}}$ ,  $i_{x_{ti}} = \frac{x_{ti} - \bar{x}_i}{S_{x_{ti}}}$ ;

$\varepsilon_t$  – регрессионные остатки в момент времени  $t$ .

В связи с тем, что исследуемый период времени с 1995 по 2006 гг. мал, то нам удастся «грубо» оценить динамику отраслевой структуры на основе модели регрессии, не претендуя на высокую достоверность результатов. По мере поступления информации модель будет усовершенствована.

Рекуррентным методом наименьших квадратов найдены значения стандартизованных коэффициентов модели (5.22), отражающие динамику в отраслевой структуре инвестиций в Оренбургской области за период с 1999 по 2006гг., так как в этот период наблюдалась активизация инвестиционной деятельности (таблица 5.12).

Согласно таблице 5.12, наибольшую долю в отраслевой структуре инвестиций составляют инвестиции в промышленность. При этом доля инвестиций в данную отрасль на протяжении исследуемого периода времени имела тенденцию к увеличению и на 2006 г. составила 49,6%. Противоречие данного результата реальности можно объяснить тем, что промышленное производство области в

значительной степени формируют электроэнергетика, черная и цветная металлургия, топливная промышленность. Топливная промышленность включает в себя нефтедобывающую и газовую отрасли, которые в настоящее время являются привлекательными для инвестирования, что и объясняет высокую долю инвестиций в промышленность.

Таблица 5.12 – Доли инвестиций в отрасли экономики Оренбургской области за период с 1999 по 2006гг.

Год	Доля инвестиций в промышленность	Доля инвестиций в сельское и лесное хозяйство	Доля инвестиций в строительство	Доля инвестиций в транспорт и связь	Доля инвестиций в жилищно-коммунальное хозяйство
1999	32,06	9,55	6,52	17,63	33,74
2000	33,75	9,06	6,06	18,77	31,86
2001	32,16	8,27	5,91	20,34	30,82
2002	31,76	8,19	5,94	21,96	28,65
2003	30,56	7,83	5,79	22,99	28,01
2004	29,86	8,01	5,29	21,46	27,87
2005	40,30	7,46	4,95	18,70	27,01
2006	49,57	5,52	4,71	11,49	28,22

Доля инвестиций в жилищно-коммунальное хозяйство на протяжении рассматриваемого периода имела тенденцию к уменьшению и на 2006г. составила 28,2%. Доля инвестиций в сельское и лесное хозяйства снизилась и составила 5,5% в 2006 г., против 9,6% в 1999 г. Доля инвестиций в транспорт и связь в 2006 г. снизилась по сравнению с 1999 г. на 6,14 процентных пункта. Малую долю в отраслевой структуре занимают инвестиции в строительство, которые направляются на развитие строительной базы. Доля инвестиций в данную отрасль на протяжении исследуемого периода времени имела тенденцию к уменьшению (6,5% в 1999 г. и 4,7% в 2006 г.) [43].

Методом экспоненциального сглаживания был осуществлен прогноз перечисленных выше показателей и на основе полученных результатов, применив рекуррентный метод наименьших квадратов, рассчитаны стандартизованные коэффициенты (таблица 5.13).

Таблица 5.13 – Доли инвестиций в отрасли экономики Оренбургской области за период с 2007 по 2009гг.

Год	Доля инвестиций в промышленность	Доля инвестиций в сельское и лесное хозяйство	Доля инвестиций в строительство	Доля инвестиций в транспорт и связь	Доля инвестиций в жилищно-коммунальное хозяйство
2007	51,20	5,43	4,32	11,04	27,51
2008	52,50	5,38	3,93	10,51	27,17
2009	53,42	5,38	3,55	9,92	27,24

На рисунке 5.17 представлен прогноз динамики отраслевой структуры инвестиций. Для оценки точности прогнозов были использованы данные за 2007 и 2008гг.

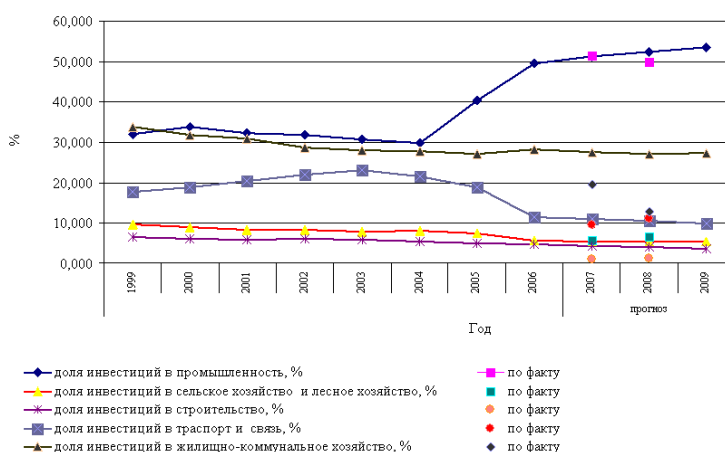


Рисунок 5.17 – Динамика стандартизованных коэффициентов, отражающих отраслевую структуру инвестиций в Оренбургской области

Согласно рисунку 5.17, доля инвестиций в промышленность будет увеличиваться и в 2009 г. составит 53,4%. Инвестиции в ЖКХ существенно не изменятся, их доля в 2009 г. составит 27,2%. Произойдет снижение инвестиций в таких отраслях, как транспорт и связь, сельское и лесное хозяйство, строительство.

В целом, можно отметить, что, несмотря на наметившиеся положительные тенденции, говорить об улучшении инвестиционной ситуации в области рано. Уход государства из сферы инвестиционной деятельности не обеспечил заполнение образовавшегося вакуума частной инициативой. Бюджетных средств совершенно недостаточно для того, чтобы оставить сколько-нибудь заметный след в

проведении экономической политики. В связи с этим необходимо более ошутимое участие государства в инвестировании [22].

## **5.6 Моделирование и прогнозирование социально-экономических процессов и систем на основе систем регрессионных уравнений**

### **5.6.1 Построение модели объемов производства промышленности и сельского хозяйства в виде группы регрессионных уравнений**

Исследуем возможные сценарии производства продукции сельского хозяйства как отрасли специализации региона при вступлении России в ВТО.

Выявлено влияние на объем выпуска продукции сельского хозяйства млн.руб. (Prod<sub>сх</sub>), объем промышленного производства, млн.руб. (Prod) следующих показателей: оборот розничной торговли, млн.руб.(Obor), экспорт товаров, млн. дол. США (Ex), импорт товаров, млн. дол. США (Im), инвестиции в основной капитал, млн. руб. (Inv), стоимость бензина автомобильного, руб. за тонну (Ben), индекс цен производителей на реализованную сельхозпродукцию (Ind<sub>сх</sub>). При выборе факторов ориентировались, в первую очередь на показатели наиболее подверженные влиянию вступления России в ВТО, такие как объем импорта, экспорта и цены на бензин.

Информационной базой служат ряды месячной динамики перечисленных показателей за период с 1998 по 2004 гг., данные за 2005 г. используются для оценки точности моделей.

Необходимым условием коинтегрированности является интегрированность первого порядка, то есть стационарность первых разностей. Анализ начнем с построения графиков исходных рядов и графиков первых разностей. Графическое изображение исходных рядов динамики и рядов их первых разностей представлено в Приложении Г. (рисунки Г.1-Г.15).

Для проверки наличия и характера тренда воспользовались критериями Фостера-Стюарта и Дики-Фуллера [38].

Выводы по результатам расчетов сведены в таблицу 5.14.

Таблица 5.14 – Проверка временных рядов на стационарность

Ряд	Вывод на основе критерия	
	Фостера-Стюарта	Дики-Фуллера
$Prod_t$	нестационарный	нестационарный
$Prod_t - Prod_{t-1}$	стационарный	стационарный
$Prod\_cx_t$	нестационарный	нестационарный
$Prod\_cx_t - Prod\_cx_{t-1}$	стационарный	стационарный
$Obor_t$	нестационарный	нестационарный
$Obor_t - Obor_{t-1}$	стационарный	стационарный
$Ex_t$	нестационарный	нестационарный
$Ex_t - Ex_{t-1}$	стационарный	стационарный
$Im_t$	нестационарный	нестационарный
$Im_t - Im_{t-1}$	стационарный	стационарный
$Ben_t$	нестационарный	нестационарный
$Ben_t - Ben_{t-1}$	стационарный	стационарный
$Ind\_cx_t$	нестационарный	нестационарный
$Ind\_cx_t - Ind\_cx_{t-1}$	стационарный	стационарный

Согласно таблице 5.14 исследуемые временные ряды являются интегрированными первого порядка. Если после оценивания регрессионные остатки окажутся стационарными, то ряды будут коинтегрированными.

Предварительно на основе нетрадиционного корреляционного анализа было установлено, что на множество факторов для изучаемых показателей не пересекаются, поэтому модель представим в виде группы регрессионных уравнений:

$$\begin{aligned} \tilde{Prod}_t &= \alpha_{11} Ex_{t-1} + \alpha_{12} Im_{t-1} + \alpha_{13} Inv_{t-1} \\ \tilde{Prod\_cx} &= \alpha_{21} Obor_{t-2} + \alpha_{22} Ben_t + \gamma_{21} D\_aug_t + \gamma_{22} D\_sep_t, \quad (5.23) \\ \tilde{Im}_t &= \alpha_{31} Prod_{t-1} + \alpha_{32} Ind\_cx_{t-1} \end{aligned}$$

где  $t = 3, \dots, T$ .

При построении модели объема выпуска сельскохозяйственной продукции мы учли регрессионную неоднородность данных, введя фиктивные переменные  $D\_aug_t$ ,  $D\_sep_t$ , которые принимают значения 1 в августе и сентябре соответственно. В силу того, что уравнения модели содержат разный набор predetermined переменных, а эндогенные переменные одного уравнения не входят в правую часть других уравнений, то оценку параметров производят методом наименьших квадратов.

МНК- оценки каждого из регрессионных уравнений (5.23):

$$\hat{Prod}_t = 50,28 Ex_{t-1} + 35,28 Im_{t-1} + 0,83 Inv_{t-1}, \hat{R}^2=0,81, F_H=123,1;$$

(19,2)                      (12,6)                      (0,27)

$$\hat{Prod\_cx}_t = 1,33 Obor_{t-2} - 0,22 Ben_t + 3060,85 D\_aug_t + 4203,79 D\_sep_t,$$

(0,6)                      (0,089)                      (1020)                      (1930)

$$R^2 = 0,79, F_H = 106,5;$$

$$\hat{Im}_t = 0,0025 Pr od_{t-1} + 0,21 Ind\_cx_{t-1} \quad R^2 = 0,68, \quad F_H = 97,5,$$

(0,001)                      (0,09)

где  $t = 3, \dots, 72$ .

Построены адекватные модели, результаты исследования приведены в приложении (Расчеты проведены в ППП Eviews, Statistica).

Согласно первому уравнению с ростом объемов экспорта, импорта и инвестиций в основной капитал в предыдущий период происходит увеличение объема промышленного производства. По модели объема выпуска сельскохозяйственной продукции следует, что при увеличении цен на бензин на 1 руб. за тонну объем выпуска продукции сельского хозяйства снизится в среднем на 0,22 млн.руб. На импорт положительное влияние оказывает индекс цен на сельскохозяйственную продукцию – с его ростом на 1% в предыдущем периоде - импорт в среднем увеличивается на 0,21 млн.долл.

При расчете коэффициентов корреляции между регрессионными остатками различных уравнений наблюдается наличие значимой связи.

$$r(e\_prod, e\_im) = -0,3349, \quad p(H_0) = 0,005;$$

$$r(e\_prod\_cx, e\_im) = 0,2684, \quad p(H_0) = 0,026.$$

Кроме того множества объясняющих переменных неодинаковы для всех уравнений. Таким образом, условия применения метода Зельнера выполнены, и можно приступить к его реализации.

Оценим элементы матрицы  $\Sigma_c$ .

$$\hat{\Sigma}_c = \begin{pmatrix} 2,522 \times 10^6 & -3,043 \times 10^5 & -1,061 \times 10^3 \\ -3,043 \times 10^5 & 1,137 \times 10^6 & 3,754 \times 10^3 \\ -1,061 \times 10^3 & 3,754 \times 10^3 & 284,058 \end{pmatrix}.$$

Используя  $\hat{\Sigma}_c$ , вычислим с помощью формулы (5.18) матрицу  $\hat{\Sigma}^{-1}$ , которую используем для оценки вектора  $b$  по формуле (5.17). Расчеты производились в MathCad, приведены в приложении Ф.

Оценка системы:

$$\begin{cases} \hat{Prod}_t = 47,08 Ex_{t-1} + 39,27 Im_{t-1} + 0,88 Inv_{t-1} \\ \quad \quad \quad (4,05) \quad \quad \quad (9,43) \quad \quad \quad (0,17) \\ \hat{Prod\_cx}_t = 1,43 Obor_{t-2} - 0,25 Ben_t + 3009,4 D\_aug_t + 4317,7 D\_sept \\ \quad \quad \quad (0,35) \quad \quad \quad (0,11) \quad \quad \quad (463,3) \quad \quad \quad (463,4) \\ \hat{Im}_t = 0,00167 Prod_{t-1} + 0,26 Ind\_cx_{t-1} \\ \quad \quad \quad (0,0007) \quad \quad \quad (0,05) \end{cases}$$

где  $t = 3, \dots, 72$ .

Построенные модели адекватны, результаты исследования приведены в приложении.

Анализ полученных моделей показал, что:

– увеличение стоимости бензина на 1 рубль за тонну приведет к снижению объема сельскохозяйственной продукции на 0,25 млн. руб., что свидетельствует о высокой чувствительности объема производимой сельхозпродукции по отношению к ценам на бензин;



- положительный знак при объеме импорта можно объяснить тем, что увеличение объема промышленной продукции происходит, в том числе, и за счет импортируемых машин и оборудования;
- увеличение экспорта на 1 млн. долл. ведет к росту объема промышленной продукции на 47,08 млн.руб.;
- увеличение индекса цен на продукцию сельского хозяйства на один процент приводит к увеличению импорта в среднем на 0,26 млн. долл.

На основе построенных моделей произведем сценарную оценку последствий присоединения России к ВТО для базовых отраслей экономики Оренбургской области.

### **5.6.2 Сценарное прогнозирование объемов производства промышленности и сельского хозяйства на основе группы регрессионных уравнений**

На основе построенных моделей можно произвести сценарное прогнозирование показателей объемов выпуска промышленности и сельского хозяйства, расчеты проводятся при условиях, что РФ вступила бы в ВТО в 2005 году.

Изменение цен на топливо это основной фактор, который повлияет на конкурентоспособность сельского хозяйства. Рост цен на топливо должно привести к падению объемов производства этой отрасли.

Прогнозирование последствий от вступления в ВТО производилось исходя из двух вариантов изменения цены на топливо: плавного и резкого возрастания цены на бензин до мирового уровня – 45385 руб. за тонну. Плавное возрастание цены предполагает увеличение цены на бензин до мировых уровней в течение 3-х лет. Резкое увеличение цены предполагает рост цены до мирового уровня в течение года. Будем исходить также из того, что в результате присоединения к ВТО снизятся импортные пошлины, что приведет к увеличению импорта. Со-

гласно мнению специалистов РАН ожидаемый рост импорта в течение первого года после вступления будет составлять 5% в целом по РФ [39]. В меньшей степени присоединение к ВТО скажется на экспорте региона. Во-первых, его топливно-сырьевая часть, составляющая 2/3 вывоза, не подвержена на мировых рынках серьезным ограничениям. Во-вторых, для остальной трети, на которую распространяются льготы ВТО, торгово-политическая ситуация не изменится, так как Россия, как правило, уже имеет доступ к тем конвенционным ставкам тарифов, которые страны - члены ВТО установили во взаимной торговле. При этом тенденции изменения остальных показателей предположим неизменными, а изменение цены на бензин также рассмотрим в рамках двух сценариев, описанных выше.

На рисунке 5.18 видно, что при росте цен на бензин до мирового уровня в течение трех лет, в первом году после вступления России в ВТО цена достигнет отметки 25000 руб. за тонну, реальная динамика цен на бензин также приведена на графике.

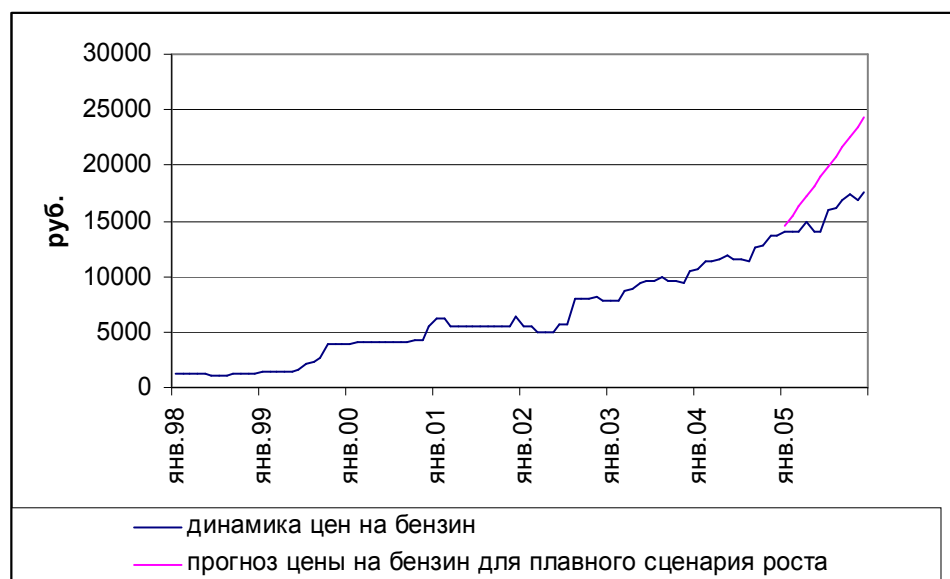


Рисунок 5.18 – Прогноз цены на бензин для плавного сценария роста цен

На рисунке 5.19 представлен резкий рост цен на бензин до мирового уровня в первый же год после вступления в ВТО.

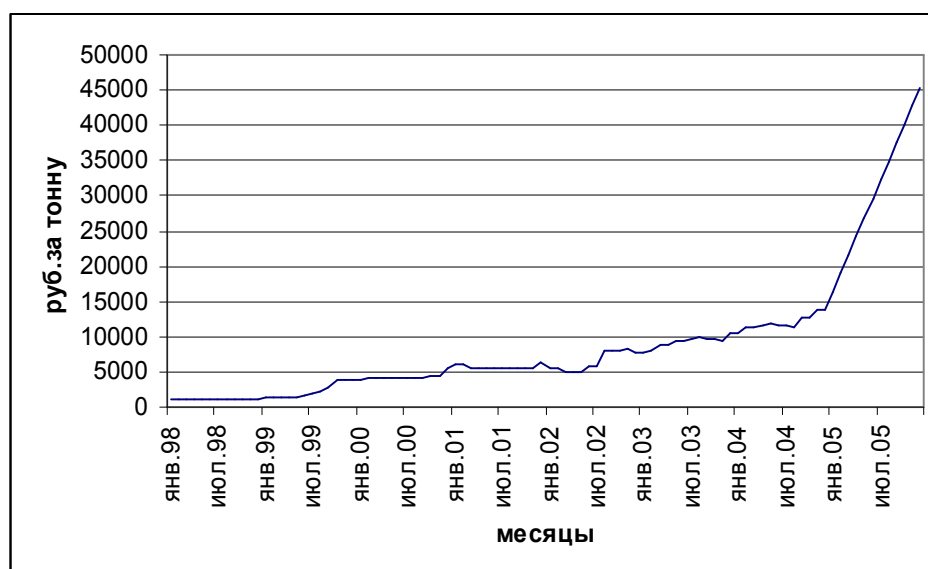


Рисунок 5.19 – Прогноз цены на бензин для резкого сценария роста цен

Также были построены прогнозные значения показателей, оказывающих влияние на объем сельхозпродукции, продукции промышленности и импорта товаров, а также рассчитанные на их основе объемы продукции сельского хозяйства при условии вступления и невступления России в ВТО при различных сценариях изменения цены на бензин.

На рисунке 5.20 представлены наблюдаемые и прогнозные значения объема сельхозпродукции при вступлении России в ВТО и плавном увеличении цены на бензин. На данном рисунке прослеживается снижение объема производства сельскохозяйственной продукции. Действительно, среднемесячное значение выпуска при условии невступления России в ВТО составит 2682,6 млн. р., а среднемесячное значение в год вступления – 1714,4 млн. р., то есть среднемесячный выпуск продукции сельского хозяйства Оренбургской области сократится на 36 процентов.

В случае резкого роста цен на топливо результаты прогнозирования еще более выпукло отражают негативные последствия для производства сельхозпродукции.

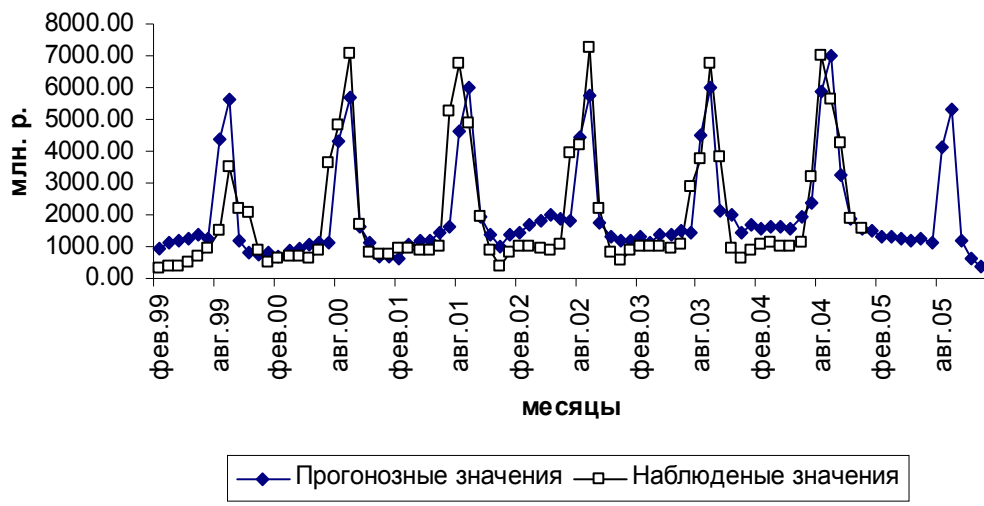


Рисунок 5.20 – Динамика и прогноз производства продукции сельского хозяйства Оренбургской области при условии вступления России в ВТО при плавном увеличении цены на бензин до мирового уровня

На рисунке 5.21 представлены наблюдаемые и прогнозные значения объема промышленной продукции при вступлении России в ВТО и плавном увеличении цены на бензин.

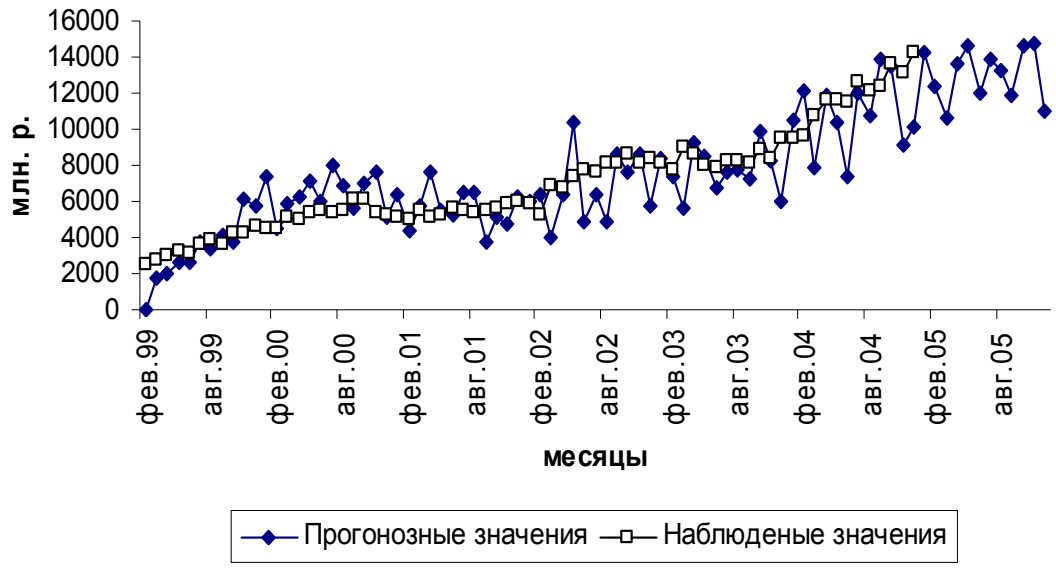


Рисунок 5.21 – Динамика и прогноз производства продукции промышленности Оренбургской области при условии вступления России в ВТО при плавном увеличении цены на бензин до мирового уровня

На рисунке 5.21 просматривается снижение среднего объема производства промышленной продукции. Действительно, среднемесячное значение выпуска при условии невступления России в ВТО 16356,7 млн. р., а среднемесячное зна-

чение при условии вступления – 16029,5 млн. р., то есть среднемесячный выпуск продукции промышленности Оренбургской области сократится на 2 процента.

### **5.6.3 Моделирование показателей рынка труда на основе систем одновременных регрессионных уравнений**

Влияние вступления России в ВТО на занятость в регионе можно попытаться оценить в зависимости от возможного увеличения доходов от экспорта, увеличения цен на топливо, роста импорта вследствие снижения импортных тарифов и увеличения объемов внешней трудовой миграции вследствие снятия экономических барьеров. Оценку будем производить на основании показателей базовых отраслей региона. Нами было определено, что отраслями специализации региона являются сельское хозяйство и промышленность. В 2004 г. в промышленности и сельском хозяйстве было задействовано соответственно 18,5 и 18,1% от всех занятых в Оренбургской области.

Исходная информационная база представлена многомерными временными рядами месячной динамики за период с 1999 по 2004 гг. следующих показателей:

- объем выпуска продукции сельского хозяйства млн.руб. (Prod\_cx);
- объем промышленного производства, млн.руб. (Prod);
- численность занятых в промышленности тыс.чел. (Zan\_prom);
- численность занятых в сельском хозяйстве тыс.чел. (Zan\_cx);
- численность занятых в торговле (Zan\_torg);
- экспорт товаров, млн. дол. США (Ex);
- импорт товаров, млн. дол. США (Im);
- инвестиции в основной капитал, млн. руб. (Inv);
- цена на энергоносители (стоимость бензина автомобильного), руб. за тонну (Ben);
- индекс цен промышленности (Ind);

- численность безработных, тыс.чел. (*Bezr*);
- численность прибывших. чел (*Prib*).

Получим количественные оценки вариации числа занятых в отраслях специализации и числа безработных Оренбургской области на основе моделей множественной регрессии.

Так как все рассматриваемые временные ряды являются интегрированными первого порядка, то выполняется необходимое условие коинтегрированности. Проверка на коинтегрированность осуществлялась с помощью расширенного критерия Дикки-Фуллера, в результате было выяснено, что временные ряды перечисленных показателей являются коинтегрированными. Факторы включались в модель на основе результатов нетрадиционного корреляционного анализа.

Модель представим в виде группы взаимозависимых регрессионных уравнений:

$$\begin{aligned} \tilde{P}rod\_cx &= \alpha_{21} Zarp_l_t + \alpha_{22} Zan\_cx_t + \gamma_{21} D_t \\ \tilde{Z}an\_prom_t &= \alpha_{31} Zarp_t + \alpha_{32} Prod_t + \alpha_{33} Inv_t + \alpha_{34} Bezr_t, \\ \tilde{Z}an\_cx_t &= \alpha_{41} Zarp_t + \alpha_{42} Prod\_cx_t + \alpha_{43} Benz_{t-1} \end{aligned} \quad (5.24)$$

где  $t = 2, \dots, T$ .

При построении модели объема выпуска сельскохозяйственной продукции мы учли регрессионную неоднородность данных, введя фиктивную переменную  $D_t$ , которая принимает значение 1 в августе, сентябре и октябре соответственно и принимает значение 0 иначе.

Система одновременных уравнений (5.24) является сверхидентифицируемой, поэтому для оценивания структурной модели будем использовать двухшаговый метод наименьших квадратов [37].

На основе приведенной формы модели (5.25) получим модельные значения эндогенных переменных, содержащихся в правой части уравнения.



Подтверждена нормальность распределения остатков каждого из этих уравнений и проверены гипотезы об адекватности построенных моделей реальным данным. Все уравнения значимы ( $F_{крит} = 2,49$ ), также значимы все коэффициенты ( $t_{крит} = 1,99$ ). Регрессионные остатки каждого из уравнений гомоскедастичны, Расчеты проводились в ППП Statistica, приведены в Приложении X.

Однако регрессионные остатки уравнений оказались автокоррелированными, поэтому возникла необходимость их отдельного моделирования [33].

Модели стационарных временных рядов позволяют использовать, как исходного временного ряда, так и моделирование остатков (ошибок). Как правило, ряд ошибок – это стационарный ряд. Наиболее распространенные модели стационарных временных рядов – это модели авторегрессии и модели скользящего среднего:

$$CC(q) \equiv MA(q): \quad y_t = \delta_t - \beta_1 \delta_{t-1} - \beta_2 \delta_{t-2} - \dots - \beta_q \delta_{t-q}, \quad (5.27)$$

где  $t = q + 1, q + 2, \dots, T$ .

$$AP(p) \equiv AR(p): \quad y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \delta_t, \quad (5.28)$$

где  $t = p + 1, p + 2, \dots, T$ .

$$APCC(p, q) \equiv ARIMA(p, q): \quad y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \delta_t - \beta_1 \delta_{t-1} - \beta_2 \delta_{t-2} - \dots - \beta_q \delta_{t-q}, \quad (5.29)$$

где  $t = k + 1, k + 2, \dots, T$ ,

$$k = \max(p, q).$$



Выбор модели основывается на анализе автокорреляционной, частной автокорреляционной функции, а также на основе расчета информационных критериев.

Ниже представлены АРСС-модели соответствующие уравнениям (5.26)

$$\begin{aligned}
 1) \text{ АРСС}(0,3): \hat{e}_t &= -0,407\delta_{t-1} - 0,225\delta_{t-2} - 0,243\delta_{t-3}, \\
 2) \text{ сезонная АРСС}(0,0)(1,0): \hat{e}_t &= -0,4085e_{t-s-1}, \\
 3) \text{ сезонная АРСС}(1,0)(1,0): \hat{e}_t &= 0,758e_{t-1} + 0,285e_{t-s-1}, \\
 4) \text{ АРСС}(1,0): \hat{e}_t &= 0,529e_{t-1},
 \end{aligned} \tag{5.30}$$

где  $t = s + 1, s + 2, \dots, 72$ ;  $s = 3$ .

Регрессионные остатки в построенных уравнениях являются белым шумом. В окончательной форме модель (5.26) будет иметь вид:

$$\begin{aligned}
 \hat{\text{Prod}}_t &= -3732,81 + 7,51Ex_{t-1} + 10,31Im_1 + 1,78Torg_t + 0,82Zarpl_t + \\
 &+ 33,69Ind_t - 0,407\delta_{t-1} - 0,225\delta_{t-2} - 0,243\delta_{t-3}, \quad \hat{R}^2 = 0,97, \quad F = 480, \\
 \hat{\text{Pr od}}_cx &= -45200,4 + 3,8Zarpl_t + 283,5Zan\_cx_t + 1774,4D_t - \\
 &- 0,408e_{t-4}, \quad \hat{R}^2 = 0,79, \quad F = 120, \\
 \hat{Zan\_prom}_t &= 237,4 - 0,0202Zarp_t + 0,0037Pr od_t + 0,0023Inv_t - \\
 &- 0,3716Bezr_t + 0,75e_{t-1} + 0,285e_{t-4}, \quad \hat{R}^2 = 0,91, \quad F = 290, \\
 \hat{Zan\_cx}_t &= 157,66 - 0,0081Zarp_t + 0,0021Pr od\_cx_t - 0,0019Benz_{t-1} + \\
 &+ 0,529e_{t-1}, \quad \hat{R}^2 = 0,98, \quad F = 1520,
 \end{aligned} \tag{5.31}$$

где  $t = 5, 6, \dots, 72$ .

Подтверждена нормальность распределения остатков каждого из этих уравнений и проверены гипотезы об адекватности построенных моделей реальным данным. Все уравнения значимы ( $F_{крит} = 2,49$ ), также значимы все коэф-

фициенты ( $t_{крит} = 1,99$ ). Регрессионные остатки каждого из уравнений некоррелированы и гомоскедастичны (Приложение X).

Согласно полученным результатам увеличение стоимости бензина на 1 рубль за тонну в предыдущий момент приведет к снижению численности занятых в сельском хозяйстве в среднем на 1,9 человек или на 1900 человек при увеличении цены на 1 рубль за литр. В первом уравнении положительный знак при объеме импорта можно объяснить тем, что увеличение объема промышленной продукции происходит, в том числе и за счет импортируемых машин и оборудования. Увеличение экспорта на 1 млн. долл. ведет к росту объема промышленной продукции в среднем на 7,51 млн.руб. Внимание обращает на себя чувствительность объема выпуска сельского хозяйства от численности занятых в отрасли, это указывает на неконкурентоспособность продукции сельского хозяйства вследствие изношенности основных средств и недостаточности внедрения передовых технологий (на эту отрасль приходится лишь 5% от общего объема инвестиций).

На основе построенных моделей была произведена оценка последствий присоединения России к ВТО для численности занятых базовых отраслей экономики Оренбургской области. Прогнозирование последствий от вступления в ВТО производилось исходя из двух вариантов изменения цены на топливо: плавного и резкого возрастания цены на бензин до мирового уровня – 45385 руб. за тонну, снижения импортных пошлин и увеличения импорта на 5%. Такие сценарии были выбраны исходя из того, что на переговорах реализована позиция России, в соответствии с которой либерализация тарифов по отдельным позициям осуществляется в течение переходных периодов. Эти периоды – от 1 года до 7 лет.

После присоединения к ВТО средневзвешенная ставка российских пошлин на сельскохозяйственные товары снизится примерно на 3 процентных пункта. Аналогичное трехпроцентное снижение произойдет и на тарифы на ввоз промышленных товаров. При этом:

- постепенно, за 3-4 года произойдет снижение ставок пошлин до 5-6% и частичное открытие рынка химических товаров, бумаги и изделий из нее;
- будут переведены на платежеспособный уровень ставки пошлин на обувь, бытовую электронику и электротехнику, что позволит исключить “серые” и “черные” схемы импорта в этих областях;
- будут заметно снижены или отменены ставки пошлин на технологическое, строительное, научное и измерительное оборудование;
- тарифные квоты на импорт говядины, свинины и мяса птицы будут существовать в рамках существующих параметров до 2009 года.

Прогноз численности занятых в промышленности и сельском хозяйстве на два года при плавном и резком изменении цен на бензин, при увеличении импорта на 5% и сохранении существующих закономерностей в развитии остальных показателей представлен на рисунках 5.22-5.23.

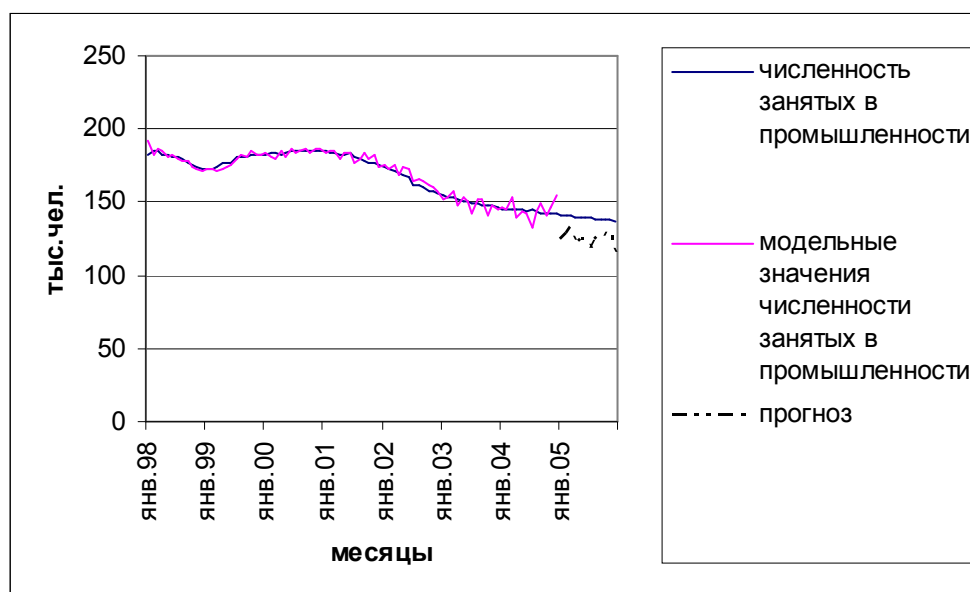


Рисунок 5.22 – Прогноз численности занятых в промышленности на первый год после вступления России в ВТО

Как можно увидеть по рисунку, ожидается снижение численности занятых в промышленности, в среднем каждый месяц снижение составит около 14300 человек. По всей видимости, основного снижения числа занятых следует ожидать в отраслях промышленности не относящихся к газовой и нефтяной промышленно-

сти (так как рост экспорта положительно повлияет на численность занятых в добывающих отраслях). Снижение численности занятых в промышленности будет идти за счет дальнейшей интенсификации промышленного производства и замедления роста выпуска промышленности. На рисунке 5.23 представлен график исходного и прогнозного рядов динамики численности занятых в сельском хозяйстве.

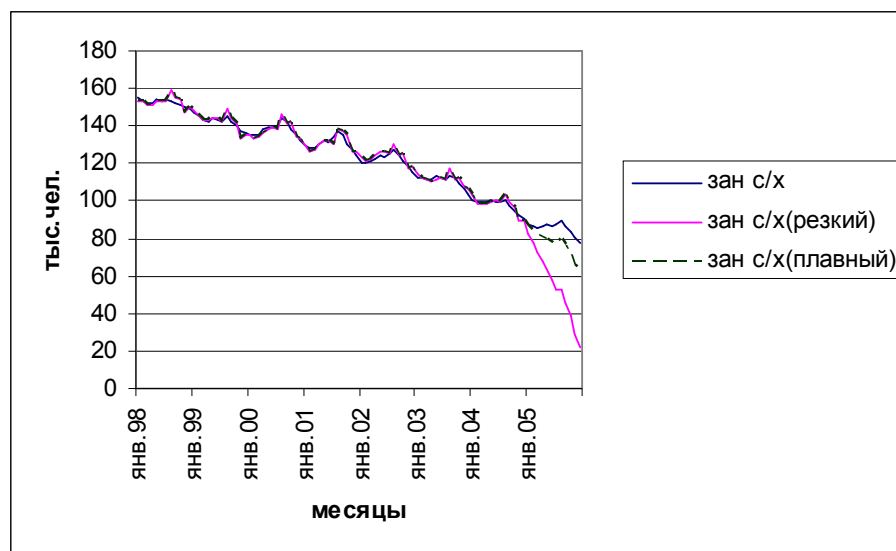


Рисунок 5.23 – Сценарный прогноз численности занятых в сельском хозяйстве на первый год после вступления России в ВТО

По графику видно, что в целом, проглядывается снижение численности занятых в сельском хозяйстве после вступления России в ВТО. В среднем каждый месяц при резком росте цен на топливо численность занятых будет снижаться на 30 тыс.чел., а при плавном росте – в среднем на 8,62 тыс.чел. ежемесячно, то есть результаты прогнозирования при резком росте цен на топливо еще более выпукло отражают негативные последствия для производства сельхозпродукции.

Аграрная ориентированность Оренбургской области может привести к значительным сокращениям занятых в этой отрасли. Хотя в сельском хозяйстве работают 18% от общего числа занятых, но проживает в сельской местности более 42% населения и 47% всех безработных региона, а это будет усиливать негативный эффект от вступления России в ВТО.

Для оценки точности прогнозов были использованы данные за 2005 г. Были рассчитаны средняя квадратичная ошибка прогноза, средняя абсолютная ошибка, средняя абсолютная процентная ошибка, которые обусловили необходимость корректировки системы одновременных регрессионных уравнений.

В модель включено еще одно уравнение, отражающее зависимость безработицы от ряда факторов и одна объясняющая переменная, характеризующая трудовую миграцию. Модель представим в виде группы регрессионных уравнений:

$$\begin{aligned}
 \tilde{Prod}_t &= \alpha_{11}Ex_t + \alpha_{12}Im_1 + \alpha_{13}Torg_t + \alpha_{14}Zarpl_t + \alpha_{15}Ind_t \\
 Prod\_cx &= \alpha_{21}Zarpl_t + \alpha_{22}Zan\_cx_t + \gamma_{21}D_t \\
 \tilde{Zan\_prom}_t &= \alpha_{31}Zarp_t + \alpha_{32}Prod_t + \alpha_{33}Inv_t + \alpha_{34}Bezr_t \\
 \tilde{Zan\_cx}_t &= \alpha_{41}Zarp_t + \alpha_{42}Prod\_cx_t + \alpha_{43}Benz_{t-1} \\
 \tilde{Bezr}_t &= \alpha_{51}Zarp_t + \alpha_{52}Zan\_torg_t + \alpha_{53}Zan\_prom_t + \alpha_{54}Prib_t
 \end{aligned} \tag{5.32}$$

где  $t = 2, 3, \dots, T$ .

Полученные модельные значения подставим вместо исходных значений ДМНК- оценки каждого из регрессионных уравнений (5.32):

$$\begin{aligned}
 \hat{Prod}_t &= -3732,81 + 7,51Ex_{t-1} + 10,31Im_1 + 1,78Torg_t + 0,82Zarpl_t + 33,69Ind_t, \\
 &\quad (1697) \quad (2,73) \quad (4,41) \quad (0,35) \quad (0,21) \quad (16,67) \\
 \hat{R}^2 &= 0,96, \quad F = 417, \\
 \hat{Prod\_cx} &= -45200,4 + 3,8Zarpl_t + 283,5Zan\_cx_t + 1774,4D_t, \\
 &\quad (6388) \quad (0,48) \quad (39,91) \quad (283,1) \\
 \hat{R}^2 &= 0,72, \quad F = 71,3, \\
 \hat{Zan\_prom}_t &= 237,4 - 0,0202Zarp_t + 0,0037Prod_t + 0,0023Inv_t - 0,3716Bezr_t, \\
 &\quad (7,81) \quad (0,002) \quad (0,0002) \quad (0,0007) \quad (0,06) \\
 \hat{R}^2 &= 0,86, \quad F = 116,9, \\
 \hat{Zan\_cx}_t &= 157,66 - 0,0081Zarp_t + 0,0021Prod\_cx_t - 0,0019Benz_{t-1}, \\
 &\quad (0,638) \quad (0,0008) \quad (0,0002) \quad (0,0003) \\
 \hat{R}^2 &= 0,98, \quad F = 1453,6, \\
 \hat{Bezr}_t &= 337,33 - 0,0149Zarp_t - 6,2035Zan\_torg_t - 0,5433Zan\_prom_t + 0,004Prib_t, \\
 &\quad (45,9) \quad (0,002) \quad (1,716) \quad (0,136) \quad (0,001) \\
 \hat{R}^2 &= 0,58, \quad F = 25,
 \end{aligned} \tag{5.33}$$

где  $t=2,3,\dots,72$ .

Подтверждена нормальность распределения остатков каждого из этих уравнений и проверены гипотезы об адекватности построенных моделей реальным данным. Все уравнения значимы ( $F_{крит} = 2,49$ ), также значимы все коэффициенты ( $t_{крит} = 1,99$ ). Регрессионные остатки каждого из уравнений гомоскедастичны.

Однако регрессионные остатки уравнений оказались автокоррелированными, поэтому возникла необходимость их отдельного моделирования. Ниже представлены АРСС-модели соответствующие уравнениям (5.33)

$$\begin{aligned}
 1) \text{ АРСС}(0,3): \hat{e}_t &= -0,407\delta_{t-1} - 0,225\delta_{t-2} - 0,243\delta_{t-3}, \\
 2) \text{ сезонная АРСС}(0,0)(1,0): \hat{e}_t &= -0,4085e_{t-s-1}, \\
 3) \text{ сезонная АРСС}(1,0)(1,0): \hat{e}_t &= 0,758e_{t-1} + 0,285e_{t-s-1}, \\
 4) \text{ АРСС}(1,0): \hat{e}_t &= 0,529e_{t-1}, \\
 5) \text{ сезонная АРСС}(1,0)(1,0): \hat{e}_t &= 0,876e_{t-1} + 0,334e_{t-s-1},
 \end{aligned} \tag{5.34}$$

где  $t = s + 1, s + 2, \dots, 72; \quad s = 12$ .

Регрессионные остатки в построенных уравнениях являются белым шумом. В окончательной форме она будет иметь вид:

$$\begin{aligned}
 \hat{Prod}_t &= -3732,81 + 7,51Ex_{t-1} + 10,31Im_1 + 1,78Torg_t + \\
 &\quad (1,697) \quad (2,73) \quad (4,41) \quad (0,35) \\
 &+ 0,82Zarpl_t + 33,69Ind_t - 0,407\delta_{t-1} - 0,225\delta_{t-2} - 0,243\delta_{t-3}, \\
 &\quad (0,21) \quad (16,67) \\
 \hat{Prod}_{cx} &= -45200,4 + 3,8Zarpl_t + 283,5Zan_{cx_t} + 1774,4D_t - 0,408e_{t-4}, \\
 &\quad (6388) \quad (0,48) \quad (39,91) \quad (283,1) \\
 \hat{Zan}_{prom}_t &= 237,4 - 0,0202Zarp_t + 0,0037Prod_t + \\
 &\quad (7,81) \quad (0,002) \quad (0,0002) \\
 &+ 0,0023Inv_t - 0,3716Bezr_t + 0,75e_{t-1} + 0,285e_{t-4}, \\
 &\quad (0,0007) \quad (0,06) \\
 \hat{Zan}_{cx}_t &= 157,66 - 0,0081Zarp_t + 0,0021Prod_{cx_t} - 0,0019Benz_{t-1} + 0,529e_{t-1}, \\
 &\quad (0,638) \quad (0,0008) \quad (0,0002) \quad (0,0003) \\
 \hat{Bezr}_t &= 337,33 - 0,0149Zarp_t - 6,2035Zan_{torg}_t - \\
 &\quad (45,9) \quad (0,002) \quad (1,716) \\
 &- 0,5433Zan_{prom}_t + 0,004Prib_t + 0,879e_{t-1} + 0,334e_{t-12-1}, \\
 &\quad (0,136) \quad (0,001)
 \end{aligned} \tag{5.35}$$

где  $t = 13, 14, \dots, 72$ .

Подтверждена нормальность распределения остатков каждого из этих уравнений и проверены гипотезы об адекватности построенных моделей реальным данным. Все уравнения значимы ( $F_{крит} = 2,49$ ), также значимы все коэффициенты ( $t_{крит} = 1,99$ ). Регрессионные остатки каждого из уравнений некоррелированы и гомоскедастичны.

Согласно полученным результатам увеличение числа прибывших увеличивает число безработных. Наибольшее снижение численности безработных вызвано трудоустройством в торговле, а также в промышленности.

Для оценки точности прогноза по скорректированной системе одновременных уравнений кроме стандартных показателей точности рассчитаем U- статистику Тейла. Данная статистика позволяет сравнивать прогностические способности моделей. Если значение статистики  $< 1$  то анализируемая модель лучше по прогностическим способностям первоначальной «эталонной» модели, если  $> 1$  - то хуже, если  $= 1$ , то анализируемая модель эквивалентна по прогностическим способностям первоначальной «эталонной» модели. Расчет статистики производится по формуле:

$$U = \sqrt{\frac{\sum \left( \frac{y_{T+L} - \hat{y}_{T+L}}{y_{T+L}} \right)^2}{\sum \left( \frac{y_{T+L} - \hat{y}_{T+L}^*}{y_{T+L}} \right)^2}}, \quad (5.36)$$

где  $y_{T+L}$  - наблюдения по контрольной выборке;

$\hat{y}_{T+L}$  - прогнозные значения по анализируемой модели;

$\hat{y}_{T+L}^*$  - прогнозные значения по предварительной «эталонной» модели.

Расчеты показали, что значение статистики Тейла превышает 1 ( $U = 1,1$ ), то есть прогнозы по скорректированной модели точнее.

На основе построенных моделей была произведена оценка последствий присоединения России к ВТО для численности безработных Оренбургской области. Прогнозирование последствий от вступления в ВТО производилось исходя из двух вариантов изменения цены на топливо: плавного и резкого возрастания цены на бензин до мирового уровня – 45385 руб. за тонну, снижения импортных пошлин и увеличения импорта на 5%, также увеличения числа прибывших в Оренбургскую область. Показатель число прибывших в Оренбургскую область был взят как фактор, отражающий внутреннюю и внешнюю трудовую миграцию. Предполагается, что в результате вступления России в ВТО ожидается усиление межотраслевой мобильности рабочей силы в региональном разрезе. При этом важно учитывать и внутреннюю трудовую миграцию, как переток рабочей силы в конкурентные отрасли, более развитые в данном регионе, чем в соседних. Однако если число прибывших из регионов России более или менее поддается учету, то число прибывших из соседних стран занижено, вследствие того, что большинство из них проникает на территорию региона нелегально.

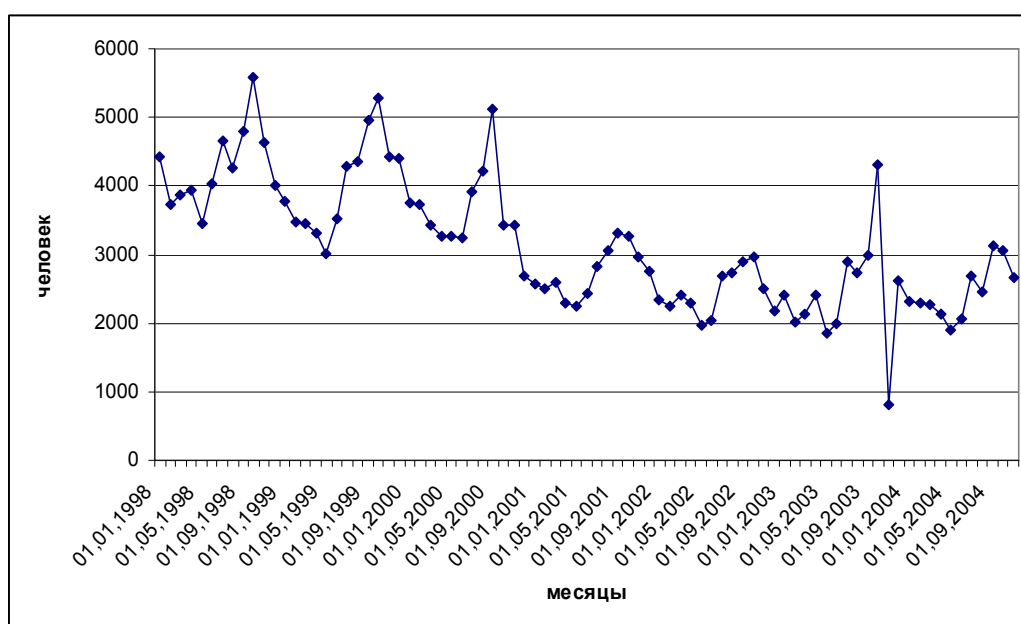


Рисунок 5.24 – Динамика числа прибывших в Оренбургскую область



Прогнозирование последствий от вступления в ВТО будем производить также исходя из двух вариантов изменения числа прибывших: плавного (сохранение существующей положительной тенденции) и резкого (увеличения числа прибывших в два раза в течение года после снятия барьеров) [42]. На рисунке 5.6.3.9 представлена динамика числа прибывших в Оренбургскую область за 1998-2004 гг.

Как видно из рисунка 5.24, наблюдается снижение числа прибывших вплоть до начала 2004 г. затем наблюдается тенденция к росту. Также просматривается наличие сезонной составляющей. Для прогнозирования числа прибывших использовано экспоненциальное сглаживание [33, 36]. Оценка полученной модели:

$$\begin{aligned} \hat{y}_t &= \hat{f}_t + \hat{g}_t \\ \hat{f}_t &= 0,225(y_t - \hat{g}_{t-12}) + 0,775(\hat{f}_{t-1} + \hat{c}_{t-1}) \\ \hat{g}_t &= \hat{g}_{t-12} \\ \hat{c}_t &= \hat{c}_{t-1} \\ t &= 13, 14, \dots, 72. \end{aligned}$$

Остатки некоррелированы и нормально распределены. На рисунке 5.25 представлен прогноз числа прибывших в Оренбургскую область на 2005 г.

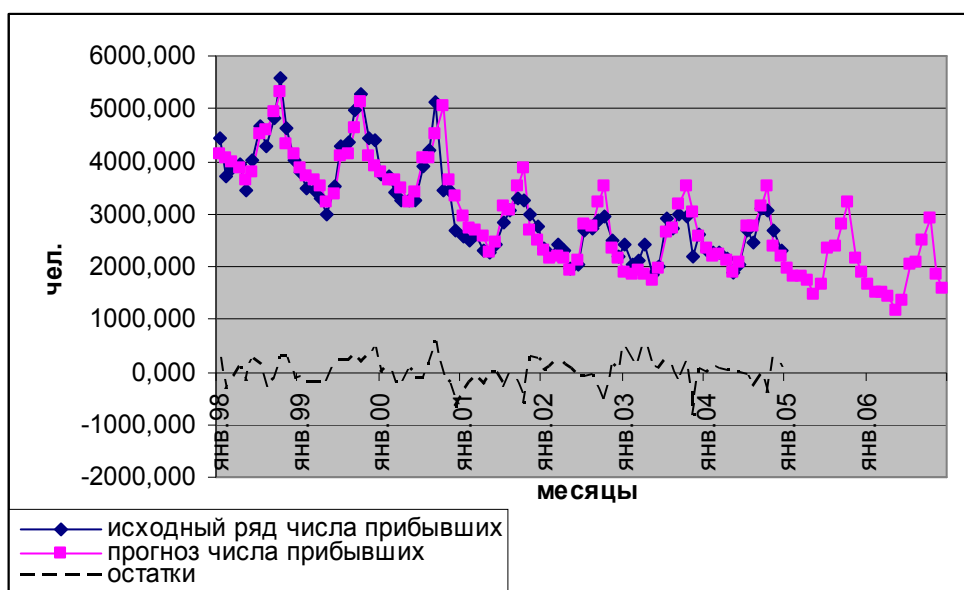


Рисунок 5.25 – Прогноз числа прибывших в Оренбургскую область на 2005 г.

Согласно прогнозу максимальное число прибывших, более 3200 чел., ожидается в октябре 2005 г. Минимальное число прибывших ожидается в мае – 1483 чел.

На рисунке 5.26 представлен сценарий резкого увеличения числа прибывших.

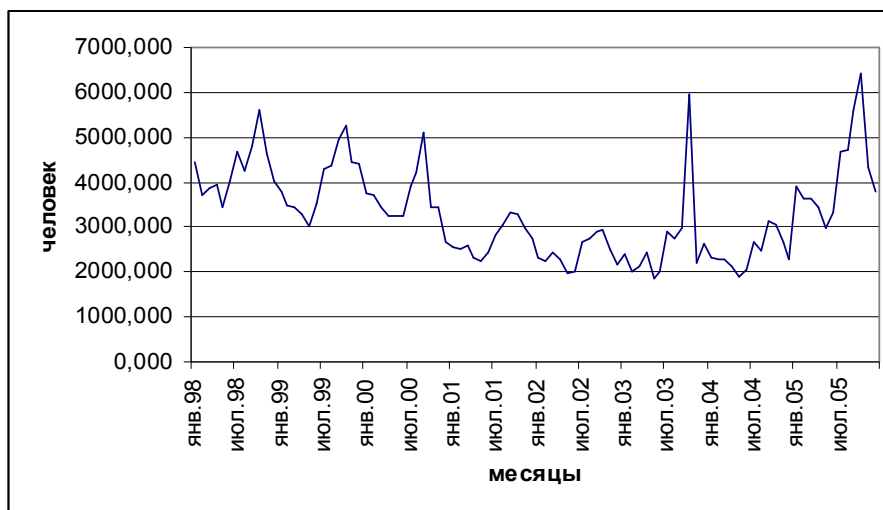


Рисунок 5.26 – Сценарий резкого увеличения числа прибывших в Оренбургскую область

На рисунке 5.27 представлен прогноз численности безработных при резком увеличении числа прибывших и при сохранении тенденции развития остальных показателей.

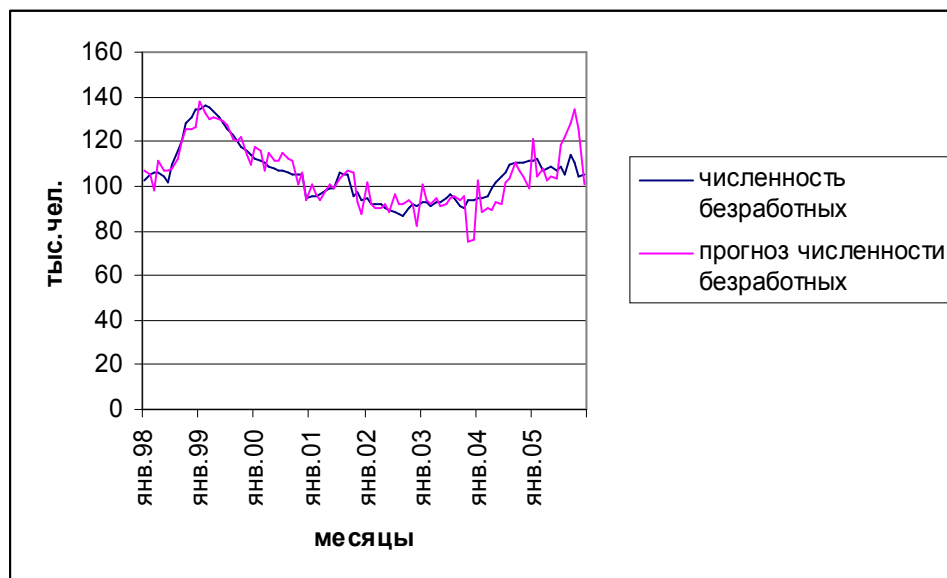


Рисунок 5.27 – Прогноз численности безработных Оренбургской области на первый год после вступления России в ВТО при резком увеличении числа прибывших

Как можно увидеть из рисунка 5.27 снятие барьеров при вступлении России в ВТО должно привести к увеличению числа безработных вследствие увеличения числа прибывших, то есть трудовой миграции. Если учесть значительное высвобождение работников в сельском хозяйстве, то к концу 2005 г. при условии, что в 2004 г. Россия вступила бы в ВТО, следует ожидать увеличение числа безработных от 120 до 160 тыс. человек.

Вступление в ВТО положительно скажется на экспортно-ориентированных отраслях, однако здесь не ожидается значительного увеличения числа занятых, что связано с высокой технологичностью производств. Основным негативным эффектом будет связан с ростом импорта и ростом цен на топливо. Особо это коснется сельского хозяйства. Меры должны быть приняты до вступления России в ВТО, из-за требования отмены протекционистской политики после вступления. Смягчение отрицательных эффектов возможно при пересмотре структуры сельскохозяйственного производства – увеличения производства пшеницы твердых сортов, развития животноводства и птицеводства, так как именно здесь следует ожидать наибольшего иностранного конкурентного давления. В настоящее время выращивание животных на мясо осуществляется, как правило, на базе скота мясного направления, эффективность отрасли невысокая. поголовье скота мясных пород имеет тенденцию к уменьшению. Эта проблема может особо обостриться при вступлении в ВТО Казахстана. Практически 100% импорта пшеницы приходится на Казахстан [29]. Однако это не единственная опасность, следует также учитывать и трудовую миграцию (здесь следует учитывать миграцию также из других государств Азии и стран СНГ). Приграничный характер региона, а также совпадение специализации экономики Оренбургской области и Казахстана могут усилить негативный эффект. Вступление в ВТО снимет многочисленные барьеры, таким образом, возрастет трудовая миграция. Реальная численность трудовых мигрантов неизвестна. Приток рабочей силы, который трудно спрогнозировать усилит и без этого напряженную ситуацию на рынке труда, которая возникнет вследствие высвобождения значительной доли рабочей силы из проблемных отраслей экономики региона.

#### **5.6.4 Моделирование и прогнозирование основных индикаторов демографической безопасности на основе системы одновременных уравнений**

Рассмотрим построение системы одновременных уравнений для моделирования и прогнозирования основных индикаторов демографической безопасности. Демографический прогноз – это научно обоснованное предвидение будущей демографической ситуации и основных параметров движения населения. На демографические прогнозы в большей степени опирается планирование всего народного хозяйства: производства товаров и услуг, жилищного и коммунального строительства, трудовых ресурсов, подготовки кадров специалистов, школ и детских учреждений, дорог и средств транспорта, военно-призывного контингента. На данные демографического прогноза опирается прогнозирование расходов по социальному обеспечению. Изменение возрастного состава населения существенным образом сказывается на масштабах и структуре заболеваемости населения. Прогноз позволяет определить перспективную потребность населения в конкретных формах медицинского обслуживания, правильно ориентировать инвестиции в развитие здравоохранения. Прогноз численности и состава семей, а также их доходов и потребностей необходим для оценки перспектив жилищного строительства. Таким образом, демографические прогнозы охватывают фактически весь хозяйственный и военный потенциал страны.

Демографические процессы протекают под влиянием множества факторов как эндогенного, так и экзогенного типа. Поскольку характеризующие их показатели представлены временными рядами, которые к тому же в большинстве случаев являются нестационарными, то для выявления факторов, оказывающих влияние на основные индикаторы демографической безопасности, использован нетрадиционный корреляционный анализ.

Информационной базой служат ряды годовой динамики за период с 1970 по 2005 годы следующих показателей:

- $y_1$  - общий коэффициент рождаемости;
- $y_2$  - общий коэффициент смертности;
- $y_3$  - уровень брачности населения;
- $y_4$  - показатель общей заболеваемости населения;
- $y_5$  - коэффициент миграционного прироста;
- $x_1$  - уровень разводимости населения;
- $x_2$  - коэффициент младенческой смертности;
- $x_3$  - удельный вес населения в трудоспособном возрасте;
- $x_4$  - удельный вес населения старше трудоспособного возраста;
- $x_5$  - ожидаемая продолжительность жизни;
- $x_6$  - обеспеченность населения больничными койками;
- $x_7$  - показатель заболеваемости системы кровообращения;
- $x_8$  - охват детей дошкольными учреждениями;
- $x_9$  - удельный вес учащихся, занимающихся во вторую и третью смены;
- $x_{10}$  - площадь жилищ, приходящаяся в среднем на одного жителя;
- $x_{11}$  - благоустройство жилищного фонда газом;
- $x_{12}$  - обеспеченность населения собственными легковыми автомобилями;
- $x_{13}$  - уровень официально зарегистрированной безработицы;
- $x_{14}$  - число пострадавших с утратой трудоспособности на 1 рабочий день и более;
- $x_{15}$  - число зарегистрированных преступлений;
- $x_{16}$  - численность лиц, состоящих на диспансерном учете по поводу алкоголизма;
- $x_{17}$  - объем использования свежей воды;
- $x_{18}$  - выбросы загрязняющих веществ в атмосферный воздух от стационарных источников.

Учитывая результаты нетрадиционного корреляционного анализа, для основных индикаторов демографической безопасности предложены модели в форме системы одновременных регрессионных уравнений:

$$\begin{cases} \tilde{y}_{1,t} = \beta_{11}y_{3,t-1} + \beta_{12}y_{5,t} + \beta_{13}x_{5,t} + \beta_{14}x_{10,t-1} + \beta_{15}x_{13,t} \\ \tilde{y}_{2,t} = \beta_{21}y_{4,t} + \beta_{22}x_{4,t} + \beta_{23}x_{14,t} + \beta_{24}x_{16,t-1} \\ \tilde{y}_{3,t} = \beta_{31}y_{5,t} + \beta_{32}x_{3,t} + \beta_{33}x_{10,t} + \beta_{34}x_{13,t} \\ \tilde{y}_{4,t} = \beta_{41}y_{5,t} + \beta_{42}x_{1,t} + \beta_{43}x_{9,t-1} + \beta_{44}x_{17,t} + \beta_{45}x_{18,t-2} \\ \tilde{y}_{5,t} = \beta_{51}y_{1,t} + \beta_{52}y_{3,t} + \beta_{53}x_{3,t} + \beta_{54}x_{13,t} \end{cases} \quad (5.37)$$

где  $t = 3, 4, \dots, T$ .

Построение систем одновременных уравнений на основе рядов динамики социально-экономических показателей также сопровождается трудностями, связанными с проблемой ложной регрессии. Как и в предыдущем случае производится проверка на коинтегрированность с помощью расширенного критерия Дики-Фуллера.

Проверка на коинтегрированность осуществлялась с помощью расширенного критерия Дики-Фуллера, в результате выяснено, что временные ряды перечисленных показателей являются коинтегрированными.

Система одновременных уравнений является сверхидентифицируемой, поэтому для оценивания структурной модели использовали двухшаговый метод наименьших квадратов.

ДМНК- оценки каждого из регрессионных уравнений:



Согласно полученным результатам уровень рождаемости Оренбургской области складывается под влиянием уровня брачности населения, коэффициента миграционного прироста, ожидаемой продолжительности жизни, обеспеченности населения жильем и уровня официально зарегистрированной безработицы. Увеличение уровня брачности в предыдущий момент приводит к росту рождаемости в среднем на 0,44 ‰. Увеличение площади жилищ, приходящегося в среднем на одного жителя в предыдущий момент времени на 1 м<sup>2</sup> влечет рост рождаемости в среднем на 0,477 ‰. Увеличение уровня официально зарегистрированной безработицы на 1 % ведет к снижению рождаемости в среднем на 0,19 ‰.

На общий коэффициент смертности Оренбургской области влияют такие процессы, как заболеваемость, старение населения. Увеличение показателя общей заболеваемости приводит к росту смертности в среднем на 1,15 ‰, а увеличение удельного веса населения старше трудоспособного возраста на 1% влечет за собой рост смертности населения Оренбургской области в среднем на 1,76 ‰. Существенное влияние на уровень смертности оказывает проблема алкоголизма в нашей области. Увеличение численности лиц, состоящих на диспансерном учете по поводу алкоголизма в предыдущий момент на 1 единицу, приводит к росту смертности в среднем на 0,84 ‰.

Увеличение в среднем уровня брачности населения Оренбургской области происходит за счет увеличения коэффициента миграционного прироста, а также за счет увеличения лиц в трудоспособном возрасте. С ростом уровня официально зарегистрированной безработицы на 1% уровень брачности в среднем снижается на 0,98‰, а коэффициент миграционного прироста – на 0,132 ‰. При увеличении удельного веса населения в трудоспособном возрасте на 1 %, происходит рост уровня брачности в среднем на 0,37‰ и увеличение миграционного прироста в среднем на 0,02 ‰.

Показатель общей заболеваемости населения складывается в большей части под влиянием экологических факторов. Увеличение объема использованной чистой воды на 1 м<sup>3</sup>/г приводит к снижению заболеваемости в среднем на 0,143 случая на 1000 населения. В то время как увеличение выбросов загрязняющих



веществ в атмосферный воздух от стационарных источников в предыдущие моменты влечет за собой рост заболеваемости в среднем на 0,063 случая на 1000 населения.

На основе построенной системы одновременных уравнений произведем прогноз для основных индикаторов демографической безопасности на пять лет вперед. Для моделирования рядов динамики экзогенных переменных применялись адаптивные методы краткосрочного прогнозирования.

Применим метод экспоненциального сглаживания для ряда динамики уровня разводимости населения. Исходя из предварительного анализа компонентного состава исследуемого ряда динамики, будем строить модель, в которой присутствует линейный тренд и отсутствует сезонность. Такая модель в общем виде имеет вид:

Оценка модели экспоненциального сглаживания для динамики уровня разводимости населения (5.38):

$$\begin{aligned} \hat{x}_t &= \hat{f}_t, \\ \hat{f}_t &= 0,4x_t + 0,6(\hat{f}_t + \hat{c}_t), \\ \hat{c}_t &= 0,2(\hat{f}_t - \hat{f}_{t-1}) + 0,8\hat{c}_t, \end{aligned} \quad (5.38)$$

где  $f_t$  отражает тенденцию развития процесса;

$\alpha_f, \alpha_c$  , - параметры сглаживания, причем  $0 < \alpha_f, \alpha_c < 1$ ;

$c_{t-1}$  - абсолютный прирост, характеризующий изменение среднего уровня процесса, или аддитивный коэффициент роста;

$$t = 3, 4, \dots, T.$$

Построенная модель адекватна, остатки распределены нормально и некоррелированы.

Для прогнозирования можно применить другие методы адаптивного прогнозирования, например методологию прогнозирования Бокса-Дженкинса [48]. Выявлено наличие трендовой составляющей во временном ряде, характеризую-

щем динамику площади жилищ, приходящихся в среднем на одного жителя. Ряд является нестационарным, поэтому была использована модель ARIMA(p,d,q). Для идентификации модели был проведен графический анализа выборочных автокорреляционных и частных автокорреляционных функций. Также были рассчитаны значения специальных информационных критериев.

$$\text{Оценка сезонной АРПСС}(1,1,0)(1,0,0): \Delta \hat{x}_t = -0,4585 \Delta x_{t-1} - 0,4739 \Delta x_{t-4-1}$$

Аналогичным образом провели моделирование рядов динамики остальных экзогенных переменных. Прогнозные значения экзогенных переменных приведены в таблице 5.15.

Таблица 5.15 – Прогнозные значения экзогенных переменных

Год	$x_1$	$x_3$	$x_4$	$x_5$	$x_9$	$x_{10}$	$x_{13}$	$x_{14}$	$x_{16}$	$x_{17}$	$x_{18}$
2006	5,2	63,6	18,9	64,7	12,3	19,5	3,8	3,7	185,9	1712,7	936,6
2007	5,3	64,0	18,9	64,4	11,6	19,4	3,9	4,0	178,9	1636,7	966,9
2008	5,3	64,5	18,8	64,1	10,9	19,2	4,0	4,3	173,7	1687,1	997,2
2009	5,4	65,0	18,7	63,7	10,3	19,1	4,1	4,4	169,7	1653,6	1027,5
2010	5,4	65,5	18,6	63,4	9,6	19,2	4,2	4,1	166,7	1675,8	1057,8

Подставив соответствующие прогнозные значения в систему одновременных регрессионных уравнений, получим прогноз по основным индикаторам демографической безопасности (таблица 5.16).

Таблица 5.16 – Прогнозные значения основных индикаторов демографической безопасности на предстоящие пять лет

Год	Общий коэффициент рождаемости	Общий коэффициент смертности	Уровень брачности населения	Показатель общей заболеваемости	Коэффициент миграционного прироста
2006	10,7	15,7	7,4	872,73	-1,12
2007	10,6	15,2	7,4	798,14	-1,79
2008	10,7	15,9	7,3	865,42	-1,76
2009	10,5	16,1	6,3	861,80	-1,72
2010	10,8	16,3	6,9	858,18	-1,69

Для оценки точности прогнозов были использованы данные за 2006 г. Были рассчитаны средняя квадратичная ошибка прогноза, средняя абсолютная

ошибка. Графики прогнозных значений для общего уровня рождаемости и смертности представлены на рисунке 5.28.

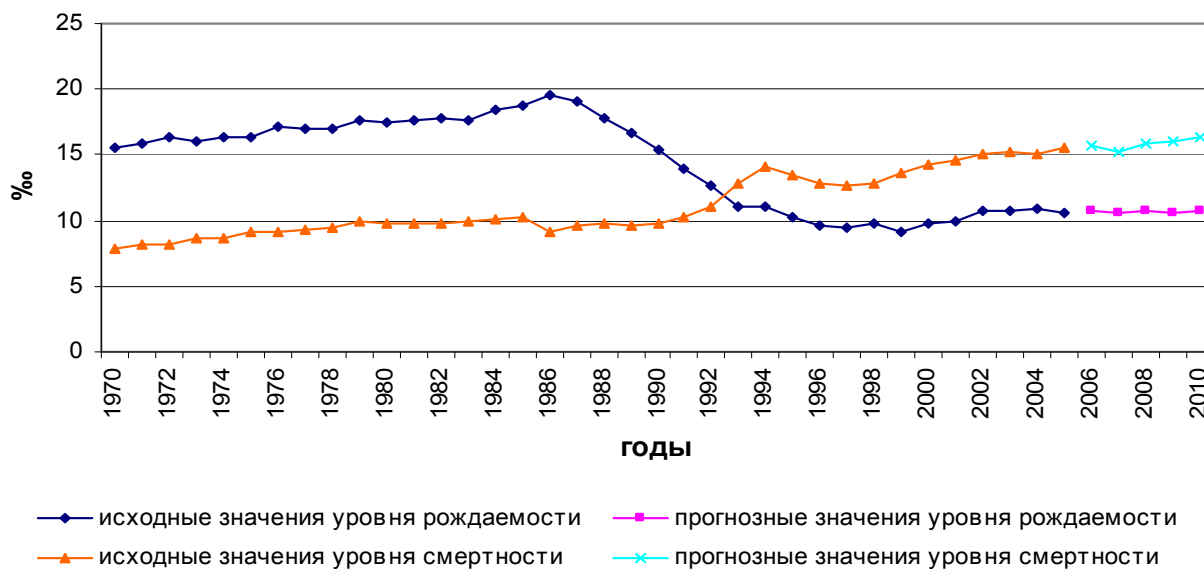


Рисунок 5.28 - График прогнозных значений уровня рождаемости и смертности населения Оренбургской области на пять лет вперед

Графики прогнозных значений остальных индикаторов демографической безопасности приведены на рисунках 5.29-5.31.

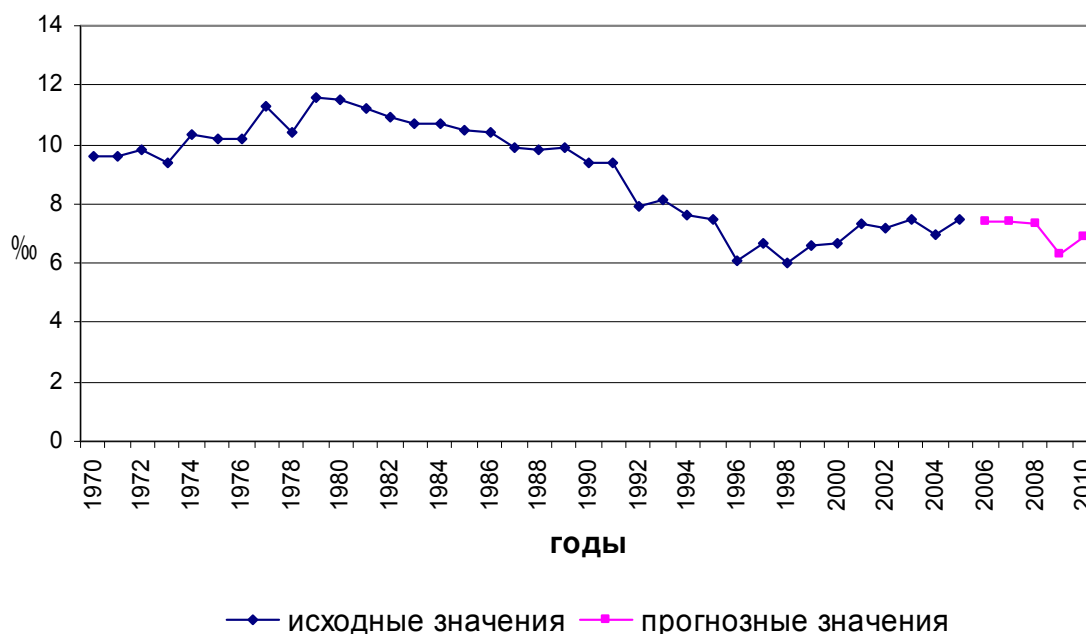


Рисунок 5.29- График прогнозных значений уровня брачности населения Оренбургской области



Рисунок 5.30 - График прогнозных значений показателя общей заболеваемости населения Оренбургской области



Рисунок 5.31 - График прогнозных значений коэффициента миграционного прироста

Таким образом, можно сделать вывод, что в ближайшее десятилетие в Оренбургской области сохранится процесс депопуляции: общий коэффициент рождаемости к 2010 году составит 10,8‰, смертности – 16,3‰. Даже незначительный рост в отдельные годы уровня рождаемости не позволит области выйти из демографического кризиса, поскольку депопуляция – явление долгосрочное и нет оснований, как показывают расчеты прогноза, предполагать быстрый выход из нее по мере устранения негативного воздействия политических, социально-экономических и демографических факторов. Сложившиеся к настоящему времени параметры воспроизводства населения будут и в дальнейшем обуславливать сокращение численности населения.

Уровень брачности, который активно влияет на воспроизводство населения, также будет снижаться. К концу прогнозируемого периода его значение составит 6,9‰. Будет также наблюдаться рост общей заболеваемости населения, хотя это увеличение весьма незначительное. К 2010 году на 1000 населения будет приходиться 858,18 случаев. Коэффициент миграционного прироста остается отрицательным, к концу прогнозируемого периода его величина составит -1,69.

## **Вопросы и задания по теоретическому материалу раздела 5**

1. В чем преимущество построения уравнения регрессии в стандартизованном масштабе?
2. В чем преимущество рекуррентного метода наименьших квадратов? В каких задачах лучше всего реализовать данный метод?
3. В чем необходимость применения условных коэффициентов автокорреляции? Могут ли положительные приросты одного временного ряда позволять строить более точные прогнозы, чем отрицательные?
4. Как можно проследить динамику связи во времени? Есть ли необходимость в этом?
5. В чем преимущество и особенность парциальных коэффициентов корреляции?
6. Какая процедура применяется при определении ранга коинтеграции нескольких временных рядов? В чем ее суть?
7. В каком случае применяется метод Зельнера и в чем его преимущество?
8. Поясните, как можно установить структуру связи между исследуемыми показателями для построения модели в форме системы одновременных уравнений.
9. В каких случаях для прогнозирования социально-экономических показателей следует применять векторные модели авторегрессии?

10. Как можно осуществлять прогнозирование факторов при построении многофакторной модели?
11. Какие многофакторные модели прогнозирования можно строить при условии, что показатели будут стационарными?
12. В чем различия между группой уравнений и системой одновременных регрессионных уравнений?
13. В чем преимущество использования модели корректировки ошибок по сравнению с моделью в форме СОУ?
14. Выберите показатели, характеризующие на Ваш взгляд, демографическое развитие региона? Поясните и обоснуйте свой выбор.
15. Выберите показатели, характеризующие на Ваш взгляд, миграционные процессы в регионе? Поясните и обоснуйте свой выбор.
16. Предложите набор факторов исходя из содержательного анализа, который можно рассматривать при моделировании миграционных процессов.
17. По каким характеристикам можно проводить изучение динамики структуры внешней миграции?
18. Какими способами можно изучать динамику структуры социально-экономических показателей?
19. Какие факторы на Ваш взгляд целесообразно включать в многофакторную модель при прогнозировании последствий вступления России в ВТО для Оренбургской области?
20. Как Вы считаете, влияют ли миграционные процессы на демографическую безопасность региона? Как это можно установить?
21. Каким образом Вы будете проводить исследование на коинтегрированность рядов динамики социально-экономических показателей, учитывая влияние на их развитие кризиса 2008 года?
22. Есть ли необходимость учета влияния мирового кризиса при моделировании и прогнозировании социально-экономических показателей региона? Если да, то, как это будет осуществляться при моделировании на основе одномерных и многомерных временных рядов?

23. Достаточно ли при прогнозировании показателей демографического развития основываться на моделях только одномерных временных рядов? Поясните и обоснуйте свой выбор.

24. Какая особенность характерна для компонентного состава ряда динамики инвестиций в основной капитал Оренбургской области? Как эта особенность проявляется при моделировании?

25. Как можно осуществлять верификацию прогнозов при отсутствии реальных будущих значений?

26. Каково преимущество использования адаптивных моделей прогнозирования?

27. В каких многофакторных моделях кроме СОУ прослеживается явное преимущество предварительного применения нетрадиционного корреляционного анализа?

### **Задание к исследовательской работе**

Исходная информационная база – данные по основным социально-экономическим показателям Оренбургской области за период с 2005 по 2010 гг (файл «Данные \_ Временные ряды.xls»).

#### **Задание 1. Моделирование и прогнозирование объемов производства по видам экономической деятельности при вступлении России в ВТО**

По данным, представленным в Приложении:

- определить какие виды экономической деятельности выступают базовыми (виды специализации) для экономики Оренбургской области;

- осуществить отбор факторов, оказывающих влияние на объемы производства по видам экономической деятельности, являющихся базовыми для Оренбургской области;

- оценить зависимость объемов производств по базовым ВЭД от отобранных факторов (многофакторные модели в форме группы регрессионных уравне-

ний, системы одновременных регрессионных уравнений, векторной модели авторегрессии, векторной модели корректировки ошибок).

- предложить сценарии развития ситуации, исходя из достигнутых соглашений вступления в ВТО;

- осуществить сценарное прогнозирование объемов производств по ВЭД при вступлении России в ВТО на основе предложенных многофакторных моделей.

Замечание. В случае если удалось подобрать модель в виде группы регрессионных уравнений, то проверить коррелированность остатков уравнений и в случае необходимости применить метод Зельнера.

## **Задание 2 Моделирование и прогнозирование показателей рынка труда при вступлении России в ВТО**

По данным, представленным в Приложении:

- определить какие виды экономической деятельности выступают базовыми (виды специализации) для экономики Оренбургской области;

- осуществить отбор факторов, оказывающих влияние на численность занятых по видам экономической деятельности, являющихся базовыми для Оренбургской области;

- оценить зависимость численность занятых по базовым ВЭД и численности безработных от отобранных факторов (многофакторные модели в форме группы регрессионных уравнений, системы одновременных регрессионных уравнений, векторной модели авторегрессии, векторной модели корректировки ошибок).

- предложить сценарии развития ситуации, исходя из достигнутых соглашений вступления в ВТО;

- осуществить сценарное прогнозирование численности занятых и безработных при вступлении России в ВТО на основе предложенных многофакторных моделей.



### **Задание 3 Моделирование и прогнозирование демографического развития региона**

По данным, содержащимся в статистических сборниках Территориального органа Федеральной службы государственной статистики по Оренбургской области:

- осуществить отбор факторов, оказывающих влияние на демографическое развитие в Оренбургской области;
- осуществить прогнозирование основных показателей демографического развития на основе одномерных и многомерных временных.

### **Задание 4 Моделирование и прогнозирование миграционных процессов**

По данным, содержащимся в статистических сборниках Территориального органа Федеральной службы государственной статистики по Оренбургской области, а также в сборниках УФМС России по Оренбургской области:

- осуществить отбор факторов, оказывающих влияние на миграционный прирост в Оренбургской области;
- осуществить прогнозирование основных показателей миграционного движения на основе одномерных и многомерных временных рядов.

### **Задание 5 Моделирование динамики структуры социально-экономических показателей по видам экономической деятельности**

По данным, содержащимся в статистических сборниках Территориального органа Федеральной службы государственной статистики по Оренбургской области, осуществить моделирование динамики структуры по предложенным социально-экономическим показателям по видам экономической деятельности либо другим специфичным для данного показателя категориям.

## Список использованных источников к разделу 5

- 1 D. McDowall, R. McCleary, E. E. Meidiger, R.A. Hay, Jr Interrupted time series analysis.- Sage Publications, Inc. University of Iowa. 1980.
- 2 Dickey, D., Fuller, W. Distribution of the Estimators for Autoregressive Time Series with a Unit Root / D. Dickey, W. Fuller // Journal of the American Statistical Association, 1979. №74 - P. 427-431.
- 3 Elliot, G. Efficient Tests for an Autoregressive Unit Root / G Elliot et al. // Econometrica, 1996. №64. - P. 813-836.
- 4 Engle, R.F. Cointegration and Error Correction: Representation, Estimation and Testing / R.F. Engle, C.W.J. Granger C.W.J. // Econometrica. 1987. Vol. 55. -№2. - p. 251–276.
- 5 Johansen, J. Statistical Ananalysis of Cointegrating Vectors // Journal of Economic Dynamics and Control. 1998. Vol. 12. p. 231–254.
- 6 Kenneth R. Stiltner, David R. Barton. Econometric models and construction forecasting // Construction Review. -1990. - №1.
- 7 Kwiatkowski, D. Testing the Null Hypothesis of Stationary against the • Alternative of a Unit Root / D. Kwiatkowski et al. // Journal of Econometrics, 1992. №54. - P. 159 – 178.
- 8 Watson M.W. Vector Avtoregression and Cointegration / M.W. Watson// Handbook of Econometrics.1994. Vol. 4. Amsterdam: North-Holland. p. 2844–2915.
- 9 Арженовский, С. В. Методы социально-экономического прогнозирования: учеб. пособие / С. В. Арженовский. - М. : Дашков и К, 2009. - 236 с. : ил. - Прил.: с. 221-235. - Библиогр.: с. 218-220. - ISBN 978-5-91131-941-0.
- 10 Ахмедов, А. Вступление в ВТО и рынок труда в России. рабочие материалы Московского центра Карнеги / А. Ахмедов, Е. Бессонова.-№3, 2004 -33с.
- 11 Бантикова, О.И. Построение интегрального показателя, характеризующего уровень демографической безопасности / О.И. Бантикова // Вестник ОГУ. – 2007. – № 2. – С. 98 – 104.

12 Большаков, А. А. Методы обработки многомерных данных и временных рядов: учеб. пособие для вузов / А. А. Большаков, Р. Н. Каримов. - М.: Горячая линия - Телеком, 2007. - 522 с.

13 Буреш, О.В. Статистический анализ и моделирование общей миграции населения сквозь призму демографической безопасности (на примере Оренбургской области) / О.В. Буреш, О.И. Бантикова // Вестник экономической интеграции. – 2008. – № 5(10). – С. 111 – 121.

14 Васянина, В.И. Внешняя трудовая миграция в Оренбургской области. Проблемы демографии, медицины и здоровья населения России: история и современность: сборник материалов II Международной научно-практической конференции / В.И. Васянина. – Пенза: РИО ПГСХА, 2006. - С.193 -196. ISBN 5-94338-192-9

15 Васянина, В.И. Сравнительный анализ административно-территориальных образований региона по уровню экономической безопасности. Вестник Оренбургского государственного университета / В.И. Васянина. - Оренбург: ГОУ ОГУ. – 2008. – № 10. – С. 76-80. ISSN 1814-6457

16 Всемирная торговая организация и национальные экономические интересы/ под ред. И.С. Королева; Ин-т мировой экономики и междунар. отношений М.: Наука,2003.- 327 с.

17 ВТО – всемирная торговая организация. Женева, Секретариат ВТО, 1995. -300 с.

18 Джонстон, Дж. Эконометрические методы / Дж. Джонстон. – М.: Статистика, 1980. – 444 с.

19 Доугерти, К. Введение в эконометрику / К. Доугерти. - М.: ИНФРА – М., 1997.

20 Жемчужникова, Ю.А. Анализ динамики процесса инвестирования в Оренбургской области / Ю.А. Жемчужникова // Взаимодействие реального и финансового секторов в трансформационной экономике: материалы Международной научной конференции. – Оренбург: ИПК ГОУ ОГУ, 2008. – С. 379 – 382.

- 21 Жемчужникова, Ю.А. Классификация муниципалитетов Оренбургской области по уровню инвестиционной привлекательности / Ю.А. Жемчужникова // Воспроизводственный потенциал региона: сборник материалов III Международной научно-практической конференции. – Уфа: РИЦ БашГУ, 2007. – С. 357 – 360.
- 22 Жемчужникова, Ю.А. Моделирование зависимостей между показателями, характеризующими инвестиционную привлекательность / Ю.А. Жемчужникова // Вестник ОГУ. – 2008. – №9. – С. 118 – 121.
- 23 Зубаревич, Н.В. Социальные последствия вступления в ВТО для регионов России / Н.В. Зубаревич // Вестник Московского университета. Сер.5, География. - 2003. - N2. - С.60-65.
- 24 Канторович, Г. Г. Анализ временных рядов / Г.Г. Канторович // Экономический журнал ВШЭ – 2002 г. – №2. - С. 251-273.
- 25 Канторович, Г. Г. Анализ временных рядов / Г.Г. Канторович // Экономический журнал ВШЭ – 2002 г. – №3. - С. 379-401.
- 26 Канторович, Г. Г. Анализ временных рядов / Г.Г. Канторович // Экономический журнал ВШЭ – 2002 г. – №4. - С. 498-523.
- 27 Канторович, Г. Г. Анализ временных рядов / Г.Г. Канторович // Экономический журнал ВШЭ – 2003 г. – №5. - С. 79-103.
- 28 Канторович, Г. Г. Анализ временных рядов / Г.Г. Канторович // Экономический журнал ВШЭ – 2002 г. – №1. - С.85-116.
- 29 Кауц, Е.В. ВТО и российский АПК / Е.В. Кауц // Пищевая промышленность. - 2003. - N2. - С.78.
- 30 Ковалева, Л. Н. Многофакторное прогнозирование на основе рядов динамики / Л. Н. Ковалева . - М. : Статистика, 1980. - 104 с.
- 31 Ливенцев, Н.Н. Актуальные проблемы присоединения России к ВТО / Н.Н. Ливенцев, Я.Д. Лисоволик. -М. : Экономика, 2002. - 383 с.
- 32 Лисоволик Я.Д. Отраслевые и региональные аспекты присоединения России к ВТО: Дис. канд. экон. Наук / Я.Д. Лисоволик. - М., 1999. - С.228.

- 33 Лукашин, Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов: учебное пособие / Ю. П. Лукашин. – М.: Финансы и статистика, 2003. – 416 с.
- 34 Магнус Я.Р. Эконометрика: Начальный курс / Я.Р. Магнус, П.К. Каттышев, А.А. Пересецки. - М.: Дело, 2005. – 504 с.
- 35 Носко, В.П. Эконометрика: введение в регрессионный анализ временных рядов / В.П. Носко. – М., 2002.- 273 с.
- 36 Носко, В. П. Эконометрика : элемент. методы и введ. в регрес. анализ врем. рядов / В. П. Носко; Ин-т экономики переходного периода. - М. : [Б. и.], 2004. - 501 с. - Библиогр.: с. 485-492. - Предм. указ.: с. 493-501. - ISBN 5-93255-141-0
- 37 Айвазян, С. А. Прикладная статистика. Основа эконометрики : в 2 т.: учеб. для вузов / С. А. Айвазян, В. С. Мхитарян. - М. : ЮНИТИ-ДАНА, 2001. – 432 с.
- 38 Реннер, А.Г. Математические методы моделирования социально-экономических процессов (региональный аспект) / А. Г. Реннер [и др.] . - Самара: Изд-во СамНЦ РАН, 2008. - 182 с.
- 39 Реннер, А.Г. Отраслевые аспекты присоединения России к ВТО (на примере Оренбургской области) / А.Г. Реннер, Л.М. Бакирова // Вестник ОГУ.- 2006.- №8.- С. 223-227.
- 40 Реннер, А.Г. Моделирование уровня демографической безопасности Оренбургской области с учетом неоднородности данных / А.Г. Реннер, О.И. Бантикова // Взаимодействие реального и финансового секторов в трансформационной экономике: материалы Международной конференции. – Оренбург, 2008.
- 41 Реннер, А.Г. Об эконометрическом подходе к моделированию структуры в динамике /А.Г. Реннер, Ю.А. Реннер // Взаимодействие реального и финансового секторов в трансформационной экономике: материалы Международной конференции. – Оренбург, 2008, С. 377-379.

42 Реннер, А.Г. Моделирование и сценарное прогнозирование занятости при вступлении России в ВТО с учетом специализации экономики региона / А.Г. Реннер, Л.М. Туктамышева // Вестник ОГУ.- 2006.- №10.- С.250-255.

43 Реннер, А.Г. Анализ структуры бюджетных инвестиций в основной капитал в отрасли экономики Оренбургской области за период 1970 – 2005гг. /А.Г. Реннер, Ю.А. Жемчужникова // Вестник ОГУ. – 2007. – №8. – С. 73 – 76.

44 Тихомиров, Н.П. Эконометрика: учебник / Н.П. Тихомиров, Е.Ю. Дорохина – М.: Издательство «Экзамен», 2003. – 512 с.

45 Туктамышева, Л.М. Учет специализации экономики региона при прогнозировании последствий вступления России в ВТО для рынка труда /Л.М. Туктамышева // Проблемы социально-экономической устойчивости региона: материалы IV Международной научно-практической конференции. – Пенза: РИО ПГСХА, 2007. – С. 122-126.

46 Ханк, Дж.Э. Бизнес-прогнозирование: [пер. с англ.] / Джон Э. Ханк, Дин У. Уичерн, Артур Дж. Райтс.- 7-е изд. - М.; СПб.; Киев : Вильямс, 2003. - 656 с.

47 Чураков, Е.П. Математические методы обработки экспериментальных данных в экономике: учеб. Пособие / Е.П. Чураков. – М.: Финансы и статистика, 2004. – 240 с.

48 Чураков, Е.П. Прогнозирование эконометрических временных рядов: учебное пособие / Е.П. Чураков. -М.: Фин. и стат., 2008. (Заказ-2008).

49 Шурыгин, А.М. Математические методы прогнозирования: учеб. пособие для вузов / А. М. Шурыгин. - М. : Горячая линия-Телеком, 2009. - 180 с.

50 Юдаева, К. Секторальный и региональный анализ последствий вступления России в ВТО: оценка издержек и выгод: рабочие материалы Московского центра Карнеги / К. Юдаева, Е. Бессонова.-№3, 2003. - 43с.