

Н.А. ДЕГТЯРЕВА

**ЭКОНОМЕТРИЧЕСКИЕ МОДЕЛИ
АНАЛИЗА И ПРОГНОЗИРОВАНИЯ**

МОНОГРАФИЯ

Челябинск, 2017

УДК 33

ББК 65

Д 26

Дегтярева, Н.А. Эконометрические модели анализа и прогнозирования : монография / Н.А. Дегтярева. – Челябинск : Цицеро, 2017. – 170 с.

ISBN 978 – 5 – 91283 - 895 - 8

В данном исследовании представлены наиболее актуальные проблемы эконометрического моделирования; изучена теоретическая и методологическая основа эконометрических исследований; приведены исследования в области применения многофакторных моделей прогнозирования для анализа и прогнозирования социально-экономических явлений и процессов в обществе.

Рецензенты:

В.Н. Павленко, доктор физико-математических наук, профессор кафедры вычислительной математики ФГБОУ ВО «Челябинский государственный университет».

И.Д. Колмакова, доктор экономических наук, заведующая кафедрой «Государственное и муниципальное управление» ФГБОУ ВО «Челябинский государственный университет».

ISBN 978 – 5 – 91283 - 895 - 8

Дегтярева Н.А., 2017

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	4
ГЛАВА I. ПРОБЛЕМЫ ПОСТРОЕНИЯ ЭКОНОМЕТРИЧЕСКИХ МОДЕЛЕЙ.....	8
1.1. Методологические вопросы построения эконометрических моделей.....	8
1.2. Парный корреляционно – регрессионный анализ.....	14
1.3. Оценка качества регрессионной модели и использование ее для экономического анализа и прогнозирования.....	24
1.4. Использование парного корреляционно-регрессионного анализа на практике.....	30
1.5. Принятие эффективных управленческих решений на основе регрессионной модели.....	48
ГЛАВА 2. МОДЕЛЬ МНОЖЕСТВЕННОЙ РЕГРЕССИИ В ЭКОНОМЕТРИЧЕСКОМ ИССЛЕДОВАНИИ.....	62
2.1. Методы и модели множественного регрессионного анализа...	62
2.2. Оценка качества множественной регрессионной модели и использование ее для экономического анализа и прогнозирования.....	67
2.3. Построение регрессионных моделей в условиях мультиколлинеарности.....	78
2.4. Гетероскедастичность и автокорреляция остатков.....	88
2.5. Линейные регрессионные модели с переменной структурой.....	109
ГЛАВА 3. ИСПОЛЬЗОВАНИЕ ЭКОНОМЕТРИЧЕСКИХ МОДЕЛЕЙ В ПРОГНОЗИРОВАНИИ СОЦИАЛЬНО - ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ.....	129
3.1. Применение многофакторных моделей прогнозирования для анализа и прогнозирования демографической ситуации.....	129
3.2. Применение многофакторных моделей прогнозирования в социальных процессах.....	135
3.3. Применение многофакторных моделей прогнозирования в сфере информатизации муниципальных образовательных учреждений.....	142
3.4. Применение многофакторных моделей прогнозирования в сельском хозяйстве.....	153
ЗАКЛЮЧЕНИЕ.....	161
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ.....	164

ВВЕДЕНИЕ

В банковской, финансовой сфере, в маркетинговых и социологических исследованиях, и других социально-экономических областях приходится иметь дело с многомерными совокупностями, когда для каждого объекта регистрируется целый набор признаков.

Сегодня деятельность в любой области экономики (управлении, финансово – кредитной сфере, маркетинге, учете, аудите) требует от специалиста применения современных методов работы, знания достижений мировой экономической мысли, понимания научного языка. Большинство новых методов основано на эконометрических моделях, концепциях, приемах.

Эконометрика это наука, которая дает количественное выражение взаимосвязей экономических явлений и процессов. Эта наука возникла в результате взаимодействия и объединения трех компонент: экономической теории, статистических и математических методов. Она занимается определением наблюдаемых в экономической жизни конкретных количественных закономерностей, применяя для этой цели статистические методы – методы парной и множественной регрессии, парной, частной и множественной корреляции, выделения тренда и других компонент временного ряда, на статистическом оценивании [46].

Специфической особенностью деятельности экономиста является работа в условиях недостатка информации и неполноты исходных данных. Анализ такой информации требует специальных методов, которые составляют один из аспектов эконометрики. Центральной проблемой эконометрики являются построение эконометрической модели и определение возможностей ее использования для описания, анализа и прогнозирования реальных экономических процессов.

Современный руководитель, предприниматель должен хорошо разбираться в экономико-математических методах, уметь их практически применять для моделирования реальных экономических ситуаций.

Специфической особенностью деятельности экономиста, менеджера, инженера является работа в условиях недостатка информации и неполноты исходных данных, в условиях неопределенности [38]. Анализ такой информации требует знания специальных эконометрических методов.

Большинство эконометрических методов основано на построении эконометрических моделей и возможности их использования для описания, анализа и прогнозирования реальных экономических процессов. Эконометрическая модель это экономико-математическая модель факторного анализа, параметры которой оцениваются средствами математической статистики. Эта модель, выступающая в качестве средства анализа и прогнозирования конкретных экономических процессов [43].

Становление и развитие эконометрического метода происходили на основе высшей статистики – на методах парной и множественной регрессии, парной, частной и множественной корреляции, выделения тренда и других компонент временного ряда, на статистическом оценивании. В XIX в. были разработаны такие статистические методы как множественная регрессия, статистическая проверка гипотез, выборочные методы (Р. Фриш, К. Пирсон и др.). В первой половине XX в. в моделирование структур спроса, потребительских расходов и их эмпирической оценки внесли вклад Р. Аллен, А. Маршалл и др. В тот же период начинается изучение производственной функции (Ч. Кобб, П. Дуглас). В 1968 г. первыми лауреатами Нобелевской премии по экономике стали Я. Тинберген и Р. Фриш, внесшие основополагающий вклад в развитие эконометрических методов [18, 42].

Значительный вклад в развитие эконометрических методов внесли отечественные экономисты и математики – Е. Е. Слуцкий, Л. В. Канторович, В. С. Немчинов, С. А. Айвазян, В.С. Мхитарян, Л. Р. Магнус, П. К. Катышев, А. А. Пересецкий, Е.Е. Елисеева и др [1,25].

Область эконометрических исследований в настоящее время значительно расширилась, т.к. в наполнении эмпирическим содержанием нуждаются априорные в разных областях рассуждения [33].

Области применения эконометрических моделей напрямую связаны с целями эконометрического моделирования, основными из которых являются:

1) прогноз экономических и социально-экономических показателей, характеризующих состояние и развитие анализируемой системы;

2) имитация различных возможных сценариев социально-экономического развития анализируемой системы.

В качестве анализируемой экономической системы могут выступать страна в целом (макроэкономические системы), регионы, отрасли и корпорации (мезосистемы), а также предприятия, фирмы и домохозяйства (микроэкономические системы).

Все шире используются методы прогнозирования в деятельности плановых, аналитических, маркетинговых отделов производственных предприятий и объединений. В современных условиях управленческие решения должны приниматься лишь на основе тщательного анализа имеющейся информации, поэтому на практике все больший интерес возникает к статистическим методам, в частности, к корреляционно-регрессионному анализу, и компьютерным программам для их реализации (EconometricViews и Statistica, SPSS и др) [2, 13,36].

Развитие компьютерных систем и специальных прикладных программ по эконометрике, совершенствование методов анализа сделали

эконометрику мощным инструментом эконометрических исследований. Компьютерные эконометрические пакеты сделали эконометрические методы более доступными и наглядными, т.к. наиболее трудоемкую (рутинную) работы по расчету различных статистик, параметров, характеристик, построению таблиц и графиков в основном стал выполнять компьютер

Большинство процессов и явлений в экономике находятся в постоянной, взаимной, объективной связи. Исследование зависимостей и взаимосвязей между объективно существующими явлениями и процессами играет большую роль. Эффективная хозяйственная деятельность невозможна без оценки этих связей между различными факторами и результативными показателями, выявление их тенденций и разработки прогнозов [22].

По направлению и сложности связей между внутренними (эндогенными) переменными и внешними (экзогенными) переменными интерес вызывают регрессионные модели, т.е. экономико–математические модели факторного анализа, параметры которой оцениваются средствами математической статистики. Эта модель выступает в качестве средства анализа и прогнозирования конкретных экономических процессов.

Различают уравнения (модели) парной и множественной регрессии. Если для обозначения эндогенных переменных использовать y , а для экзогенных переменных x , то в случае линейной модели, уравнение парной регрессии имеет вид: $y_t = a_0 + a_1x_{t1} + \varepsilon_t$, а уравнение множественной регрессии имеет вид: $y = a_0 + a_1x_{t1} + a_2x_{t2} + \dots + a_mx_{tm} + \varepsilon_t$.

ГЛАВА I. ПРОБЛЕМЫ ПОСТРОЕНИЯ ЭКОНОМЕТРИЧЕСКИХ МОДЕЛЕЙ

1.1. Методологические вопросы построения эконометрических моделей

В любой эконометрической модели, в зависимости от конечных прикладных целей ее использования все участвующие в ней *переменные* подразделяются на:

- экзогенные переменные, задаваемые как бы извне, автономно, в определенной степени управляемые (планируемые);

- эндогенные переменные, значения которых формируются в процессе и внутри функционирования анализируемой социально-экономической системы под воздействием экзогенных переменных и во взаимодействии друг с другом, являются предметом объяснения в эконометрической модели;

- предопределенные переменные выступают в роли факторов-аргументов или объясняющих переменных;

- лаговые эндогенные переменные входят в уравнения анализируемой эконометрической системы, но измерены в прошлые моменты, а, следовательно, являются уже известными, заданными.

Эконометрическая модель служит для объяснения поведения эндогенных переменных в зависимости от значений экзогенных и лаговых эндогенных переменных [49].

Чтобы построить эконометрическую модель с последующим использованием ее для прогнозирования, необходимо решить следующие задачи [1]:

- собрать исходные данные;
- отобрать факторы, наиболее влияющие на результирующий показатель;

➤ выбрать тип эконометрической модели, в том числе формирование структуры и форм функциональной связи между переменными, включенными в модель;

➤ сделать статистический анализ выбранной модели, то есть оценить неизвестные параметры модели на базе имеющихся статистических данных;

➤ проверить качество модели: сопоставить реальные и модельные данные, проверить адекватность модели, оценить точность модельных данных;

➤ использовать модель для экономического анализа и прогнозирования.

Приведенная ниже схема отражает циклический характер современных эконометрических исследований (см. рис. 1).

Наиболее распространенными в эконометрическом моделировании являются следующие образующие четыре группы методы [6]:

- классическая линейная модель множественной регрессии (КЛММР) и классический метод наименьших квадратов (МНК);
- обобщенная КЛММР и обобщенный МНК;
- методы статистического анализа временных рядов;
- методы анализа систем одновременных эконометрических уравнений.

Применение этих методов делает возможным построение следующих типов эконометрических моделей:

1. Регрессионные модели с одним уравнением

В таких моделях зависимая (объясняемая) переменная y представляется в виде функции:

$$y = f(x, \beta) = f(x_1, \dots, x_k, \beta_1, \dots, \beta_k)$$

где x_1, \dots, x_k - независимые (объясняющие) переменные,

β_1, \dots, β_k - параметры.



Рисунок 1 - Схема эконометрических исследований

В зависимости от вида функции $f(x, \beta)$ модели делятся на линейные и нелинейные

Например, можно исследовать уровень дохода семьи как функцию от ряда ее экономических и социально-демографических характеристик (наличие и количество работников в семье, наличие и количество детей и прочих иждивенцев, уровень образования и квалификации главы семьи и т.д.).

2. Модели временных рядов

К этому классу относятся модели:

- *тренда*: $y(t) = T(t) + \xi_t$,

где t - время,

$T(t)$ - временной тренд заданного параметрического вида (*например*, линейный $T(t) = a + bt$),

ξ - случайная (стохастическая) компонента;

- *сезонности*: $y(t) = S(t) + \xi_t$,

где $S(t)$ - периодическая (сезонная) компонента,

ξ_t - случайная (стохастическая) компонента.

- *тренда и сезонности*; $y(t) = T(t) + S(t) + \xi_t$ (аддитивная)

или $y(t) = T(t)S(t) + \xi_t$ (мультипликативная)

где $T(t)$ - временной тренд заданного параметрического вида,

$S(t)$ - периодическая (сезонная) компонента,

ξ_t - случайная (стохастическая) компонента.

Кроме того, существуют модели временных рядов, в которых присутствует циклическая компонента, формирующая изменения анализируемого признака, обусловленные действием долговременных циклов экономической, демографической или астрофизической природы (волны Кондратьева, циклы солнечной активности и т.д.).

Модели временных рядов могут применяться для изучения и прогнозирования объема продаж туристических путевок, спроса на железнодорожные и авиабилеты, при краткосрочном прогнозировании процентных ставок и т.д.

3. Системы одновременных уравнений

Эти модели описываются системами уравнений. Системы могут состоять из тождеств и регрессионных уравнений, каждое из которых, кроме объясняющих переменных, может включать в себя объясняемые переменные из других уравнений системы. Системы одновременных уравнений требуют сложного математического аппарата и могут быть использованы для моделей национальной экономики.

Ярким примером системы одновременных уравнений служит модель спроса и предложения:

$$Q_t^S = \alpha_1 + \alpha_2 P_t + \alpha_3 P_{t-1} + \xi_t \quad (\text{предложение}),$$

$$Q_t^D = \beta_1 + \beta_2 P_t + \beta_3 Y_t + u_t \quad (\text{спрос}),$$

$$Q_t^S = Q_t^D \quad (\text{равновесие}).$$

где Q_t^D - спрос на товар в момент времени t ,

Q_t^S - предложение товара в момент времени t ,

P_t - цена на товар в момент времени t , Y_t - доход в момент t .

Цена товара P_t , и спрос на товар $Q_t = Q_t^S = Q_t^D$ определяются из уравнений модели, то есть являются эндогенными переменными. Объясняющими переменными в данной модели являются доход Y_t , и значение цены товара в предыдущий момент времени P_{t-1} .

Для эконометрического моделирования используются данные следующих трех типов.

1. Предположим, что мы располагаем результатами регистрации значений переменных (x^1, x^2, \dots, x^p) на n статически обследованных объектах. Так что если i - номер обследованного объекта, то имеющиеся исходные статистические данные состоят из n строк вида $(x_i^1, x_i^2, \dots, x_i^p)$, $i = \overline{1, n}$, где x_i^j - значение j переменной, зарегистрированное на i обследованном объекте. То есть, данные могут быть представлены в виде матрицы $n \times p$:

$$X = \begin{pmatrix} x_1^1 & x_1^2 & \dots & x_1^p \\ x_2^1 & x_2^2 & \dots & x_2^p \\ \dots & \dots & \dots & \dots \\ x_n^1 & x_n^2 & \dots & x_n^p \end{pmatrix}.$$

Такой тип данных называется пространственной выборкой или данными поперечного среза (cross-section data). Это данные по какому-либо экономическому показателю, полученные от разных однотипных объектов (фирм, регионов, но относящихся к одному и тому же моменту времени). Такие данные не имеют временного параметра, и порядок их следования

не существенен. Пример: финансовые показатели работы различных предприятий (об объеме производства, количестве работников, о доходе) за истекший год.

2. Предположим, что данные регистрируются на одном и том же объекте, но в разные периоды времени. Тогда аналогом i будет номер периода времени, к которому привязаны соответствующие данные, а n будет общим числом периодов времени. Такие данные называются временной выборкой, или временными рядами данных (timeseriesdata), или данными продольного среза. Это данные, характеризующие один и тот же объект в различные моменты времени. Для таких данных существенен порядок следования значений переменных. Пример: финансовые показатели одного предприятия (данные об инфляции, данные о средней заработной плате, данные о национальном доходе) за последние несколько лет.

3. Предположим, что отслеживается каждый из n объектов в течение T периодов времени. То есть, имеем последовательность матриц вида X , отнесенных к моментам времени $1, 2, \dots, T$:

$$X(t) = \begin{pmatrix} x_1^1(t) & x_1^2(t) & \dots & x_1^p(t) \\ x_2^1(t) & x_2^2(t) & \dots & x_2^p(t) \\ \dots & \dots & \dots & \dots \\ x_n^1(t) & x_n^2(t) & \dots & x_n^p(t) \end{pmatrix}.$$

Такие данные называются панельными, или пространственно-временной выборкой (paneldata). Данные сочетают в себе свойства как временных рядов, так и данных поперечного сечения. Как правило, значение T мало. Пример: показатели социально-экономического состояния домохозяйств за три года.

1.2. Парный корреляционно – регрессионный анализ

Большинство явлений и процессов в экономике находятся в постоянной взаимной и всеохватывающей объективной связи. Исследование зависимостей и взаимосвязей между объективно существующими явлениями и процессами играет большую роль в экономике. Оно дает возможность глубже понять сложный механизм причинно – следственных отношений между явлениями. Для исследования интенсивности, вида и формы зависимостей широко применяется корреляционно – регрессионный анализ, который является методическим инструментарием при решении задач прогнозирования, планирования и анализа хозяйственной деятельности предприятий.

Методы и модели корреляционно - регрессионного анализа занимают центральное место в математическом аппарате эконометрики [25].

Исследование корреляционных связей называют корреляционным анализом, а исследование односторонних вероятностных зависимостей – регрессионным анализом.

Задачами корреляционного анализа являются выявление статистической связи между случайными переменными, оценка ее тесноты (силы); отбор факторов, оказывающих наиболее существенное влияние на результативный признак, на основании измерения тесноты связи между явлениями; обнаружение неизвестных причинных связей с помощью логически – профессиональных рассуждений, раскрывающих механизм связей.

Задачами регрессионного анализа являются установление формы зависимости между переменными (линейная или нелинейная, положительная или отрицательная и т.д.), оценка функции регрессии

(в виде математического уравнения того или иного типа) и, оценка неизвестных значений (прогноз значений) зависимой переменной.

При изучении взаимосвязи между переменными Y и X следует, прежде всего, установить тип зависимости (связи) (природу анализируемых переменных Y и X). Рассмотрим некоторый экономический объект (процесс, явление, систему) и выделим только две переменные, характеризующие объект. Обозначим переменные буквами Y и $X = (X_1, X_2, \dots, X_k)$ – одну и набор случайных величин. Будем предполагать, что независимая (объясняющая) переменная X оказывает воздействие на значения переменной Y , которая, таким образом, является зависимой переменной, т.е. имеет место зависимость: $Y = f(X)$.

Различают два вида зависимостей между экономическими явлениями и процессами: функциональную и статистическую (стохастическую, вероятностную).

1. Функциональные зависимости имеют место при исследовании связей между неслучайными переменными. Если Y и X являются неслучайными переменными, при этом значения Y строго зависят только от соответствующих значений X и полностью ими определяются. Если каждому значению одной переменной соответствует вполне определенное значение другой, то в этом случае говорят о функциональной зависимости, тогда Y является некоторой функцией от переменной X и верна модель $Y = f(X)$.

2. В большинстве случаев между переменными величинами существуют зависимости, когда каждому значению одной переменной, соответствует не какое – то определенное, а *множество* возможных значений другой переменной, среди которых могут быть факторы, общие для двух случайных величин, что позволяет говорить о вероятностной или статистической зависимости. Например: зависимость урожайности от

количества внесенных удобрений, производительности труда на предприятии от его энерговооруженности и т.п.

Если каждому значению одной переменной соответствует определенное (условное) распределение другой переменной, то говорят о статистической (стохастической, вероятностной) зависимости. Она возникает в следующих случаях: 1) при исследовании связей между случайными переменными Y и X . Величины X и Y подвержены действию множества различных случайных факторов; 2) при исследовании связей между случайными и неслучайными переменными Y и X , измеряемыми с некоторой случайной ошибкой (в этом случае наблюдаются не сами переменные, а искаженные, случайные величины: $Y' = Y + \varepsilon_y$ и $X'_i = X_i + \varepsilon_{x_i}, i = 1, 2, \dots, k$). 3) при анализе влияния на случайную переменную Y неслучайных факторов $X = (X_1, X_2, \dots, X_k)$. Такая связь может быть вызвана двумя причинами: а) ошибками измерения показателя Y , по отношению к которым ошибки измерения факторов X_1, X_2, \dots, X_k пренебрежительно малы; б) если зависимая переменная подвержена влиянию ряда неконтролируемых или неучтенных факторов.

Статистическая зависимость проявляется только в массовом процессе, при большом числе единиц совокупности.

В частном случае статистическая зависимость проявляется в том, что при изменении одной из величин изменяется условное математическое ожидание $M_x(Y)$ или $M_y(X)$ другой. В этом случае говорят о корреляции или корреляционной зависимости.

Если зависимость между двумя переменными такова, что каждому значению одной переменной соответствует определенное условное математическое ожидание (среднее значение) другой, то такая статистическая зависимость называется корреляционной.

Иначе, корреляционной зависимостью между двумя переменными называется функциональная зависимость между значениями одной из них и условным математическим ожиданием другой. Верна модель: $M_x(Y) = \varphi(x)$ или $M_y(X) = \psi(y)$, где $\varphi(x) \neq \text{const}$, $\psi(y) \neq \text{const}$.

Односторонняя зависимость случайной переменной Y от одной (или нескольких) неслучайной независимой переменной X называется регрессионной. Верна модель: $M_x(Y) = \varphi(x)$, где $\varphi(x) \neq \text{const}$. Пример: такая зависимость может возникнуть в случае когда при каждом фиксированном значении X соответствующие значения Y подвержены случайному разбросу за счет действия ряда неконтролируемых факторов. На самом деле, для одного X мы можем наблюдать разные значения Y : а) X – возраст индивидуума, Y – его зарплата; в) X – доход семьи, Y – расходы на питание.

При этом зависимую переменную Y называют также функцией отклика, объясняемой, выходной, результирующей, эндогенной переменной, результативным признаком, а независимую переменную X – объясняющей, входной, предсказывающей, предикторной, экзогенной переменной, фактором, регрессором, факторным признаком.

Уравнение $M_x(Y) = \varphi(x)$, где $\varphi(x) \neq \text{const}$, называется модельным уравнением регрессии Y по X (или просто уравнением регрессии).

Функция $\varphi(x)$ называется модельной функцией регрессии (или просто функцией регрессии Y по X), если она описывает изменение условного среднего значения результирующей переменной Y в зависимости от изменения значений объясняющей переменной X : $\varphi(x) = M(Y|X = x)$.

График функции $\varphi(x)$ называется модельной линией регрессии.

Для точного описания уравнения регрессии необходимо знать условный закон распределения зависимой переменной Y при условии, что

переменная X примет значение x , т.е. $X = x$. В статистической практике такую информацию получить, как правило не удастся, т.к. обычно исследователь располагает лишь выборкой пар значений (x_i, y_i) ограниченного объема n . В этом случае речь может идти об оценке (приближенном выражении, аппроксимации) по выборке функции регрессии. Такой оценкой является выборочная линия (кривая) регрессии:

$$\hat{y} = \hat{\varphi}(x, b_0, b_1, \dots, b_p)$$

где \hat{y} - условная средняя переменной Y при фиксированном значении $X = x$;
 b_0, b_1, \dots, b_p - параметры кривой.

Уравнение $\hat{y} = \hat{\varphi}(x, b_0, b_1, \dots, b_p)$ называется выборочным уравнением регрессии.

В силу воздействия неучтенных случайных факторов и причин отдельные наблюдения переменной Y будут в большей или меньшей мере отклоняться от функции регрессии $\varphi(X)$. В этом случае уравнение взаимосвязи двух переменных (парная регрессионная модель) может быть представлена в виде:

$$Y = \varphi(X) + \varepsilon,$$

где ε - случайная переменная (возмущающая переменная), характеризующая отклонение от функции регрессии и называется случайным членом или возмущением, отклонением (или ошибкой).

Присутствие в модели случайной компоненты ε , обусловлено следующими причинами:

3. Ошибки спецификации. Среди них выделяют не включение важных объясняющих переменных, агрегирование (объединение) переменных, неправильную функциональную спецификацию модели.

4. Ошибки измерения. Они связаны со сложностью сбора исходных данных и использованием в модели аппроксимирующих

переменных для учета факторов, непосредственное измерение которых невозможно.

5. Ошибки, связанные со случайностью человеческих реакций. Обусловлены тем, что поведение и непосредственное участие человека в ходе сбора и подготовки данных может быть достаточно непредсказуемым, и вносит, таким образом, свой вклад в случайный член.

Таким образом, в регрессионной модели зависимая переменная Y есть некоторая функция $\varphi(X)$ с точностью до случайного возмущения ε .

По расположению эмпирических точек *можно предположить* наличие линейной корреляционной (регрессионной) зависимости между переменными X и Y . В этом случае линейная парная регрессионная модель будет иметь вид: $y_t = \alpha + \beta x_t + \varepsilon_t$ $t=1, \dots, n$.

Уравнение зависимости y_t от x_t вида: $y_t = \alpha + \beta x_t + \varepsilon_t$ $t=1, \dots, n$

где x_t - неслучайная (детерминированная) величина;

y_t - случайная величина;

ε_t - случайная величина,

называется линейным регрессионным уравнением с двумя переменными (или моделью линейной парной регрессии),

На основе выборочного наблюдения оценивается выборочное линейное уравнение регрессии:

$$\hat{y}_t = a + bx_t,$$

где (a, b) - оценки параметров (α, β) .

В силу того, что случайные отклонения ε_t по выборке не могут быть определены, то они заменяются отклонениями e_t значений y_t переменной Y от оцененной линии регрессии \hat{y}_t , т.е. как разность между фактическим и расчетным значениями зависимой переменной: $e_t = y_t - \hat{y}_t$ - выборочная оценка отклонения ε_t или остаток регрессии. Остатки e_t , как и ошибки ε_t ,

являются случайными величинами, однако они, в отличие от ошибок ε_t , наблюдаемы.

Воздействие неучтенных случайных факторов и ошибок наблюдений в модели линейной парной регрессии $y_t = \alpha + \beta x_t + \varepsilon_t$ определяется с помощью дисперсии возмущений (ошибок) или остаточной дисперсии σ^2 . Несмещенной оценкой этой дисперсии является выборочная остаточная дисперсия:

$$S^2 = \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n-2} = \frac{\sum_{t=1}^n e_t^2}{n-2} = \hat{\sigma}^2, \quad ,$$

где $(n-2)$ - число степеней свободы;

\hat{y} - групповая средняя, найденная по уравнению регрессии;

$e_t = y_t - \hat{y}_t$ - выборочная оценка отклонения ε_t или остаток регрессии.

Величина $S = \sqrt{S^2}$ называется стандартной ошибкой регрессии.

Согласно методу наименьших квадратов неизвестные параметры a и b выбираются таким образом, чтобы сумма квадратов отклонений эмпирических значений y_t от значений \hat{y}_t , найденных по уравнению регрессии $\hat{y}_t = a + bx_t$, была минимальной. Иначе, суть МНК заключается в минимизации суммы квадратов остатков [4]:

$$Q = \sum_{t=1}^n e_t^2 = \sum_{t=1}^n (y_t - \hat{y}_t)^2 = \sum_{t=1}^n (y_t - a - bx_t)^2 \rightarrow \min$$

где (x_t, y_t) - известные значения наблюдения (числа), (a, b) - неизвестные.

На основании необходимого условия существования экстремума функции двух переменных $Q = Q(a, b)$ приравниваем к нулю ее частные производные, т.е.

$$\begin{cases} Q'_a = -2 \sum (y_t - a - bx_t) = 0 \\ Q'_b = -2 \sum (y_t - a - bx_t)x_t = 0 \end{cases} \quad \text{или} \quad \begin{cases} \sum (y_t - a - bx_t) = 0 \\ \sum (y_t - a - bx_t)x_t = 0 \end{cases}, \quad \text{или} \quad \begin{cases} a \cdot n + b \sum x_t = \sum y_t \\ a \sum x_t + b \sum x_t^2 = \sum x_t y_t \end{cases}$$

После преобразования получим следующую систему нормальных уравнений:

$$\begin{cases} a + b\bar{x} = \bar{y} \\ a\bar{x} + b\bar{x}^2 = \overline{xy} \end{cases}$$

где соответствующие средние, определяются по формулам:

$$\bar{x} = \frac{\sum_{t=1}^n x_t}{n}; \quad \bar{y} = \frac{\sum_{t=1}^n y_t}{n}; \quad \overline{xy} = \frac{\sum_{t=1}^n x_t y_t}{n}; \quad \overline{x^2} = \frac{\sum_{t=1}^n x_t^2}{n}.$$

Решение системы являются следующие оценки коэффициентов регрессии:

$$b = \frac{n \sum x_t y_t - \sum x_t \sum y_t}{n \sum x_t^2 - (\sum x_t)^2}; \quad a = \frac{1}{n} \sum y_t - \frac{1}{n} \sum x_t \cdot b \quad \text{или} \quad \begin{cases} b = \frac{\text{cov}(x, y)}{\text{var}(x)} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2} \\ a = \bar{y} - b\bar{x} \end{cases}$$

где $\text{var}(x) = S_x^2 = \sigma_x^2$ – выборочная дисперсия переменной X;

$\text{cov}(x, y)$ – выборочный корреляционный момент или выборочная ковариация.

Подставляя значение $a + b\bar{x} = \bar{y}$ из первого уравнения системы в уравнение регрессии $\hat{y} = a + bx$ получим $\hat{y} - \bar{y} = b(x - \bar{x})$. Из полученного уравнения регрессии следует, что линия регрессии проходит через точку (\bar{x}, \bar{y}) , т.е. $\bar{y} = a + b\bar{x}$.

Коэффициент b называется выборочным коэффициентом регрессии (или угловым коэффициентом регрессии или просто коэффициентом регрессии) Y по X. Он показывает, на сколько единиц в среднем изменяется переменная Y при увеличении независимой переменной X на одну единицу.

Постоянная a дает прогнозируемое значение зависимой переменной при $x = 0$. Это может иметь смысл в зависимости от того, как далеко находится $x = 0$ от выборочных значений.

Выборочные дисперсии величин y, \hat{y}, e :

$$\text{var}(y) = \frac{1}{n} \sum (y_t - \bar{y})^2 \text{ — дисперсия наблюдаемых значений } y;$$

$$\text{var}(\hat{y}) = \frac{1}{n} \sum (\hat{y}_t - \bar{y})^2 \text{ — дисперсия расчетных значений } y;$$

$\text{var}(e) = \frac{1}{n} \sum (y_i - \hat{y}_i)^2$ — дисперсия остатков.

Для того чтобы регрессионный анализ, основанный на МНК, давал наилучшие из всех возможных результаты, должны выполняться определенные условия (условия Гаусса – Маркова) или предпосылки (гипотезы) регрессионного анализа [44].

Основные гипотезы:

1. $y_t = \alpha + \beta x_t + \varepsilon_t$, $t = \overline{1, n}$ - спецификация модели.
2. x_t - неслучайная величина, y_t - случайная величина, ε_t - случайная величина (если условие о неслучайности объясняющей переменной не выполняется, то оценки коэффициентов регрессии оказываются смещенными и несостоятельными).

3. Математическое ожидание случайного отклонения ε_t в любом наблюдении равно нулю: $M(\varepsilon_t) = 0$. Данное условие означает, что случайный член не должен иметь систематического смещения. Если постоянный член включен в уравнение регрессии, то это условие выполняется автоматически.

4. Дисперсия случайного члена ε_t (или зависимой переменной y_t) постоянна для всех наблюдений: $D(\varepsilon_t) = \sigma^2$ (или $D(y_t) = \sigma^2$) – условие гомоскедастичности возмущения (требование постоянства дисперсии регрессионных возмущений) или равноизменчивости возмущения (зависимой переменной).

Данное условие означает, что дисперсия случайного члена в каждом наблюдении имеет только одно значение.

Если условие гомоскедастичности не выполняется, то оценки коэффициентов регрессии будут неэффективными, хотя и несмещенными.

5. Случайные члены ε_i и ε_j (или переменные y_i и y_j) должны быть статистически независимы между собой, т.е. $M(\varepsilon_i \varepsilon_j) = 0$, $i \neq j$. Иначе,

предполагается отсутствие систематической связи между значениями случайного члена в любых двух наблюдениях, которые должны быть абсолютно независимы друг от друга, т.е. остатки ε_i и ε_j некоррелированы.

Данное условие указывает на некоррелированность случайных членов для разных членов. Это условие часто нарушается, когда данные являются временными рядами. В случае, когда это условие не выполняется, говорят об автокорреляции остатков.

Если условие независимости случайных членов не выполняется, то оценки коэффициентов регрессии будут неэффективными, хотя и несмещенными.

6. Случайный член ε_i (или зависимая переменная y_i) есть нормально распределенная величина: $\varepsilon_i \leftrightarrow N(0; \sigma^2)$.

В этом случае модель $y_i = \alpha + \beta x_i + \varepsilon_i$ называется классической нормальной линейной регрессионной моделью.

Предположение о нормальности распределения случайного члена необходимо для проверки значимости параметров регрессии и для их интервального оценивания.

Для получения уравнения регрессии достаточно предпосылок 1-4. Требование выполнения предпосылки 5 (т.е. рассмотрение нормальной регрессии) необходимо для оценки точности уравнения регрессии и его параметров.

Теорема Гаусса – Маркова: Если регрессионная модель $y_i = \alpha + \beta x_i + \varepsilon_i$ удовлетворяет предпосылкам 2-5 регрессионного анализа, то оценки a и b , сделанные с помощью МНК, имеют наименьшую дисперсию в классе всех линейных несмещенных оценок (являются наилучшими линейными несмещенными оценками), т.е. обладают следующими свойствами:

1) несмещенности: $E(a) = \alpha$, $E(b) = \beta$ (Это означает отсутствие систематической ошибки в положении линии регрессии).

2) эффективности: имеют наименьшую дисперсию в классе всех линейных несмещенных оценок, равную

$$D(a) = \frac{\overline{x^2} \sigma^2}{n \cdot \text{var}(x)}, \quad D(b) = \frac{\sigma^2}{n \cdot \text{var}(x)}; \quad \text{где } \text{var}(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

3) состоятельности: $\lim_{n \rightarrow \infty} D(a) = 0$, $\lim_{n \rightarrow \infty} D(b) = 0$.

Это означает, что при достаточно большом n оценки (a, b) близки к (α, β)

1.3. Оценка качества регрессионной модели и использование ее для экономического анализа и прогнозирования

Определим показатель оценки тесноты корреляционной зависимости

Рассмотрим случай линейной зависимости вида: $\hat{y} - \bar{y} = b(x - \bar{x})$.

Подходящим измерителем тесноты связи Y и X является коэффициент регрессии b , ибо он показывает, на сколько единиц в среднем изменяется Y , когда X увеличивается на одну единицу. Однако b зависит от единиц измерения переменных. Поэтому нужна такая стандартная система единиц измерения, в которой данные по различным характеристикам оказались бы сравнимы между собой. Статистика знает такую систему единиц. Эта система используется в качестве единицы измерения переменной ее среднее квадратическое отклонение S [41].

Представим уравнение $\hat{y} - \bar{y} = b(x - \bar{x})$ в эквивалентном виде:

$$\frac{\hat{y} - \bar{y}}{S_y} = b \frac{S_x}{S_y} \frac{x - \bar{x}}{S_x}$$

В этой системе величина $r = b \frac{S_x}{S_y}$ показывает, на сколько величин S_y

изменится в среднем Y , когда X увеличится на одно S_x .

Величина r является показателем тесноты связи и называется выборочным коэффициентом корреляции (или коэффициентом корреляции).

Если $r > 0$, то корреляционная связь между переменными называется прямой; если $r < 0$ - обратной. При прямой (обратной) связи увеличение одной из переменных ведет к увеличению (уменьшению) условной (групповой) средней другой.

Учитывая, что $b = \frac{\overline{xy} - \bar{x}\bar{y}}{x^2 - \bar{x}^2} = \frac{Cov(X, Y)}{S_x^2}$, формулу для r представим в

виде:

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{S_x \cdot S_y} \quad \text{или} \quad r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{x^2 - (\bar{x})^2} \sqrt{y^2 - (\bar{y})^2}}$$

$$\text{или} \quad r_{xy} = \frac{n \sum_{t=1}^n x_t y_t - (\sum_{t=1}^n x_t)(\sum_{t=1}^n y_t)}{\sqrt{n \sum_{t=1}^n x_t^2 - (\sum_{t=1}^n x_t)^2} \sqrt{n \sum_{t=1}^n y_t^2 - (\sum_{t=1}^n y_t)^2}}.$$

Выборочный коэффициент корреляции r (при достаточном большом объеме выборки n) также как и коэффициент корреляции двух случайных величин обладает следующими свойствами:

1. Коэффициент корреляции принимает значения на отрезке $[-1; 1]$ т.е. $-1 \leq r \leq 1$. Чем ближе $|r|$ к единице, тем теснее связь.

2. При $r = \pm 1$ корреляционная связь представляет линейную функциональную зависимость. При этом все наблюдаемые значения располагаются на прямой линии.

3. При $r = 0$ линейная корреляционная связь отсутствует. При этом линия регрессии параллельна оси Ox .

4. В случае парной линейной регрессионной модели коэффициент детерминации равен квадрату коэффициента корреляции: $R^2 = r^2$.

Для нелинейной регрессии тесноту связи Y и X оценивает индекс корреляции ρ_{xy} ($0 \leq \rho_{xy} \leq 1$):

$$\rho_{xy} = \sqrt{1 - \frac{\sigma_{оцм}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\sum (y_t - \hat{y}_t)^2}{\sum (y_t - \bar{y})^2}}.$$

Рассмотрим вариацию (разброс) $\sum (y_t - \bar{y})^2$ значений y_t вокруг среднего значения \bar{y} . Разобьем эту вариацию на две части: объясненную регрессионным уравнением и не объясненную (т.е. связанную с ошибками ε_t).

Пусть $\hat{y}_t = a + bx_t$ предсказанное значение y_t , тогда $y_t - \bar{y} = (y_t - \hat{y}_t) + (\hat{y}_t - \bar{y})$ и вариация y_t представляется в виде трех слагаемых:

$$\sum (y_t - \bar{y})^2 = \sum (y_t - \hat{y}_t)^2 + \sum (\hat{y}_t - \bar{y})^2 + 2 \sum (y_t - \hat{y}_t) \cdot (\hat{y}_t - \bar{y}).$$

Так как $(\sum e_t (\hat{y}_t - \bar{y}) = \sum e_t (a + bx_t - \bar{y}) = (a + b\bar{x} - \bar{y}) \sum e_t + b \sum e_t x_t = 0)$, поэтому верно равенство:

$$\sum_{TSS} (y_t - \bar{y})^2 = \sum_{ESS} (y_t - \hat{y}_t)^2 + \sum_{RSS} (\hat{y}_t - \bar{y})^2.$$

где $TSS = var(y)$ - вся дисперсия, $ESS = var(e)$ - необъясненная часть всей дисперсии (остаточная дисперсия), $RSS = var(\hat{y})$ - объясненная часть всей дисперсии.

Таким образом, дисперсия $var(y)$ разложена на две части:

$$var(y) = var(\hat{y}) + var(e),$$

объясненную регрессионным уравнением и необъясненную часть.

Коэффициентом детерминации R^2 , или долей объясненной дисперсии (или долей вариации зависимой переменной, объясненной с помощью уравнения регрессии) называется:

$$R^2 = \frac{\text{var}(\hat{y})}{\text{var}(y)} = 1 - \frac{\text{var}(e)}{\text{var}(y)}, \text{ причем } 0 \leq R^2 \leq 1.$$

Отношение дисперсии $\frac{\text{var}(e)}{\text{var}(y)}$ представляет собой долю необъясненной дисперсии.

Коэффициент детерминации обладает следующими свойствами [50]:

1. Если $R^2 = 1$, то подгонка точная т.е. все точки наблюдений лежат на регрессионной прямой и между переменными Y и X существует линейная функциональная зависимость: $\text{var}(y) = \text{var}(\hat{y})$, $\text{var}(e) = 0$, $y_t = \hat{y}_t$, $t = \overline{1, n}$.

2. Если $R^2 = 0$, то регрессия ничего не дает (вариация зависимой переменной полностью обусловлена воздействием неучтенных в модели переменных, и линия регрессии параллельна оси абсцисс), т.е. переменная x не улучшает качества предсказания y по сравнению с предсказанием (горизонтальной прямой) $\hat{y} = \bar{y}$.

3. Чем ближе к единице R^2 , тем лучше качество подгонки, т.е. регрессия \hat{y} более точно аппроксимирует наблюдаемые значения y (регрессия лучше аппроксимирует эмпирические данные, тем теснее наблюдения примыкают к линии регрессии).

Коэффициент детерминации является одной из наиболее эффективных оценок адекватности регрессионной модели, мерой качества уравнения регрессии (или мерой качества подгонки регрессионной модели к наблюдаемым значениям y_t), характеристикой прогностической силы анализируемой регрессионной модели.

Оценку качества построенной модели даст также средняя ошибка аппроксимации – среднее отклонение расчетных значений от фактических:

$$\bar{E} = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \cdot 100\%$$

Допустимый предел значений \bar{E} - не более 8-10%.

Оценка значимости уравнения регрессии позволяет установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальными данными и достаточно ли включенных в уравнение объясняющих переменных (одной или нескольких) для описания зависимой переменной. Для этого проверяется гипотеза H_0 : о статистической незначимости уравнения регрессии и показателя тесноты связи R^2 .

В качестве критерия проверки гипотезы принимают случайную величину:

$$F = \frac{\sum_{t=1}^n (\hat{y}_t - \bar{y})^2 / m}{\sum_{t=1}^n (y_t - \hat{y}_t)^2 / (n - m - 1)} = \frac{R^2}{1 - R^2} (n - 2),$$

которая имеет распределение Фишера с $v_1 = 1$, $v_2 = n - 2$ степенями свободы.

Вычисляется наблюдаемое (фактическое) значение критерия $F_{факт}$ (где n – объем выборки, m – число параметров в уравнении регрессии) и сравнивается с критическим (табличным) значением $F_{кр}$ значений F - критерия Фишера.

- если $F_{факт} < F_{кр}$, то H_0 принимается и признается статистическая незначимость, ненадежность уравнения регрессии (или R^2 незначим, или уравнение регрессии незначимо);

- если $F_{факт} > F_{кр}$, то H_0 отклоняется и признается статистическая значимость, надежность уравнения регрессии (или R^2 значим, или уравнение регрессии значимо)

Для оценки статистической значимости коэффициентов регрессии проверяется гипотеза о случайной природе показателей, т.е. о незначимом их отличии от нуля: $H_0: \beta = 0$ (или $H_0: \alpha = 0$)

В качестве критерия проверки гипотезы H_0 принимают случайную величину: $t = \frac{b-0}{S_b} = \frac{b}{S_b}$ (или $t = \frac{a}{S_a}$), которая имеет распределение Стьюдента с $\nu = n - 2$ степенями свободы.

Вычисляется фактическое (наблюдаемое) значение критерия $t_{факт}$. По таблице критических точек распределения Стьюдента, по заданному уровню значимости α и числу степеней свободы $\nu = n - 2$ находят критическую (табличную) точку $t_{кр}$. Вычисленный критерий $t_{факт}$ сравнивается с критическим (табличным) значением $t_{кр}$:

- если $|t| < t_{кр}$, то H_0 принимается и признается случайная природа формирования коэффициента b (или a);

Если H_0 принимается, то есть основания считать, что величина Y не зависит от X . Говорят, что коэффициент b статистически незначим (он слишком близок к нулю).

- если $|t| > t_{кр}$, то H_0 отвергается, т.е. коэффициент b (или a) не случайно отличаются от нуля и сформировались под влиянием систематически действующего фактора x .

Если H_0 отклоняется, то есть основания считать, что между Y и X существует определенная линейная зависимость. Говорят, что коэффициент b статистически значим.

На основе точной и адекватной модели выполняется прогнозирование в регрессионных моделях, т.е. построение оценки

зависимой переменной для некоторого набора независимых переменных, которых нет в исходных наблюдениях.

При точечном прогнозировании оценка - некоторое число, а при интервальном прогнозировании - интервал, в котором находится истинное значение зависимой переменной с заданным уровнем значимости.

Предсказанным (прогнозным) значением является оценка \hat{y}_p (точечный прогноз), которая определяется путем подстановки в выборочное уравнение регрессии соответствующего (прогнозного) значения $x = x_p$: $\hat{y}_p = a + bx_p$.

Ошибка предсказания равна разности между предсказанным и действительным значениями: $\Delta_p = \hat{y}_p - y_p$.

Доверительный интервал прогноза для действительных (индивидуальных) значений зависимой переменной y_p определяется выражением:

$$\hat{y}_p - t_{1-\alpha;n-2} S_{\hat{y}_p} \leq y_p^* \leq \hat{y}_p + t_{1-\alpha;n-2} S_{\hat{y}_p}.$$

где $S_{\hat{y}} = \sqrt{S_{\hat{y}}^2}$ - стандартная ошибка групповой средней \hat{y} и равна

$$S_{\hat{y}} = S \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum_{t=1}^n (x_t - \bar{x})^2}}.$$

1.4. Использование парного корреляционно-регрессионного анализа на практике

Эконометрическое исследование на основе парной модели регрессии состоит в изучении принципов количественного анализа реальных экономических процессов и явлений во времени и в пространстве; в получении знаний по эмпирическому выводу экономических зависимостей, закономерностей и законов; в приобретении навыков

построения и использования эконометрических моделей, а также в умении оценивать их параметры для объяснения поведения исследуемых экономических явлений; в умении проверять выдвигаемые гипотезы о свойствах экономических показателей и формах их связи; в умении оценивать и использовать результаты экономического анализа для прогноза и принятия обоснованных экономических решений; в формировании навыков эконометрического анализа на практике с использованием реальных массивов экономических данных и современного эконометрического программного обеспечения: Eviews, Statistica, MSExcel при решении задач.

Вопросы построения и использования эконометрических моделей рассмотрим на примере линейной двухфакторной модели [8].

В табл. 1 приведены статистические данные [39] результативного показателя $y(t)$ и данные факторов $x_1(t)$ и $x_2(t)$:

Таблица 1 – Статистические данные

$y(t)$	30	28	33	38	40	42	44	49	47
$x_1(t)$	5	7	10	12	15	18	20	23	24
$x_2(t)$	12	15	16	19	17	20	24	25	28

Проведем эконометрическое исследование.

Для определения меры степени линейной связи между двумя переменными используется выборочный коэффициент корреляции:

$$r_{yx} = \frac{\sum (y_t - \bar{y})(x_t - \bar{x})}{\sqrt{\sum (y_t - \bar{y})^2} \sqrt{\sum (x_t - \bar{x})^2}}$$

Если факторов несколько, то из них выбирается в модель один или два фактора, с помощью парных коэффициентов корреляции. Составляется матрица парных коэффициентов корреляции, измеряющих тесноту связи каждого из факторов-признаков с результативным фактором и между собой (табл.2):

Таблица 2 – Матрица парных коэффициентов корреляции

Факторы	y	x_1	x_2	...	x_m
y	1	r_{yx1}	r_{yx2}	...	r_{yxm}
x_1	r_{yx1}	1		...	r_{x1xm}
x_2	r_{yx2}	r_{x1x2}	1	...	r_{x2xm}
...
x_m	r_{yxm}	r_{x1xm}	r_{x2xm}	...	1

Чем ближе его значение к 1, тем теснее связь. Считается, что связь достаточно сильная, если коэффициент корреляции по абсолютной величине превышает 0,7, то есть $0,7 < |r_{yx}| < 1$; слабая, если меньше 0,3, то есть $|r_{yx}| < 0,3$; при равенстве его нулю, связь полностью отсутствует.

В модель включают те факторы, связь которых с зависимой переменной наиболее сильная.

По данным задачи построим матрицу коэффициентов парной корреляции $y(t)$ с $x_1(t)$ и $x_2(t)$ и выберем фактор наиболее тесно связанный с зависимой переменной $y(t)$.

Найдем средние значения результативного признака и наблюдаемых факторов:

$$\bar{y} = \frac{\sum_{t=1}^9 y_t}{n} = \frac{351}{9} = 39,0 \quad \bar{x}_1 = \frac{\sum x_1}{n} = \frac{134}{9} = 14,9 \quad \bar{x}_2 = \frac{\sum x_2}{n} = \frac{176}{9} = 19,6$$

Вычислим парные коэффициенты корреляции, используя промежуточные вычисления, приведённые в табл.3:

$$r_{yx_1} = \frac{\sum (y - \bar{y})(x_1 - \bar{x}_1)}{\sqrt{\sum (y - \bar{y})^2} \sqrt{\sum (x_1 - \bar{x}_1)^2}} = \frac{397}{\sqrt{438} \sqrt{376,8}} = \frac{397}{405,8} \approx 0,98$$

$$r_{yx_2} = \frac{\sum (y - \bar{y})(x_2 - \bar{x}_2)}{\sqrt{\sum (y - \bar{y})^2} \sqrt{\sum (x_2 - \bar{x}_2)^2}} = \frac{288,2}{\sqrt{438} \sqrt{218,24}} = \frac{288,2}{308,8} \approx 0,93$$

$$r_{x_1x_2} = \frac{\sum (x_1 - \bar{x}_1)(x_2 - \bar{x}_2)}{\sqrt{\sum (x_1 - \bar{x}_1)^2} \sqrt{\sum (x_2 - \bar{x}_2)^2}} = \frac{274,6}{\sqrt{376,8} \sqrt{218,24}} = \frac{274,6}{287,3} \approx 0,96$$

Из полученных парных коэффициентов корреляции составим матрицу парных коэффициентов корреляции:

	y	x_1	x_2
y	1	0,98	0,93
x_1	0,98	1	0,96
x_2	0,93	0,96	1

Так как $r_{yx_1} = 0,97 > 0$, то связь прямая; $|r_{yx_1}| > 0,7$ – высокая

Так как $r_{yx_2} = 0,91 > 0$, то связь прямая; $|r_{yx_2}| > 0,7$ – высокая

Так как $|r_{x_1x_2}| = 0,97 > 0,8$, то имеет место мультиколлинеарность, оба фактора в модель включать нельзя

Так как $|r_{yx_1}| > |r_{yx_2}|$, то в модель выбираем фактор x_1 , потому что он наиболее тесно связан с зависимой переменной y .

Построение уравнения регрессии сводится к оценке ее параметров. Для оценки параметров регрессии используют метод наименьших квадратов (МНК). МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака y от теоретических y_p минимальна.

Для линейного уравнения вида: $\hat{y}_t = a_0 + a_1x_1$ коэффициенты найдём по формулам, используя расчетную табл.3:

$$a_1 = \frac{\sum (y - \bar{y})(x_1 - \bar{x}_1)}{\sum (x_1 - \bar{x}_1)^2} = \frac{397}{376,8} \approx 1,1$$

$$a_0 = \bar{y} - a_1 \cdot \bar{x}_1 = 39,0 - 1,1 \cdot 14,9 = 22,6$$

Линейная однопараметрическая модель регрессии для выбранного x_1 примет вид: $\hat{y}_t = 22,6 + 1,1x_1$.

Таблица 3 – Расчетная таблица

	y	x_1	x_2	$y - \bar{y}$	$x_1 - \bar{x}_1$	$(y - \bar{y})(x_1 - \bar{x}_1)$	$(y - \bar{y})^2$	$(x_1 - \bar{x}_1)^2$	$(x_2 - \bar{x}_2)$	$(x_2 - \bar{x}_2)^2$	$(y - \bar{y})(x_2 - \bar{x}_2)$	$(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)$
1	30	5	12	-9	-9,9	89,1	81	98,0	-7,6	57,76	68,4	75,24
2	28	7	15	-11	-7,9	86,9	121	62,4	-4,6	21,16	50,6	36,34
3	33	10	16	-6	-4,9	29,4	36	24	-3,6	12,96	21,6	17,64
4	38	12	19	-1	-2,9	2,9	1	8,4	-0,6	0,36	0,6	1,74
5	40	15	17	1	0,1	0,1	1	0,01	-2,6	6,76	2,6	-0,26
6	42	18	20	3	3,1	9,3	9	9,61	0,4	0,16	1,2	1,24
7	44	20	24	5	5,1	25,5	25	26,0	4,4	19,36	22,0	22,44
8	49	23	25	10	8,1	81,0	100	65,6	5,4	29,16	54,0	43,74
9	47	24	28	8	9,1	72,8	64	82,8	8,4	70,56	67,2	76,44
Σ	351	134	176			397,0	438	376,8		218,24	288,2	274,6

Качество модели оценивается по адекватности и точности на основе анализа остатков регрессии: $E(t) = y_t - \hat{y}(t)$ - остаточный ряд.

Анализ остатков позволяет получить представление, насколько хорошо подобрана сама модель и насколько правильно выбран метод оценки коэффициентов.

Под адекватностью модели понимается соответствие ее реальному экономическому процессу. Модель адекватна, если ее остаточный ряд $E(t)$ является случайной компонентой. Это значит, что для остаточного ряда, как для случайной компоненты, должны выполняться четыре ее свойства [50]:

1. математическое ожидание значений остаточного ряда близко или равно нулю;
2. значения остаточного ряда случайны;
3. уровни остаточного ряда независимые;
4. подчинены нормальному закону распределения.

Проверим модель на адекватность. Для этого в табл. 4 приведём промежуточные вычисления и результаты использования линейной модели.

Таблица 4 – Промежуточные вычисления

t	y_t	x_1	\hat{y}_p	$E(t)$	Поворот - ная точка	$E(t)^2$	$E(t)-E(t-1)$	$(E(t)-E(t-1))^2$	$(E(t) /y_t)100\%$
1	30	5	28,1	1,9	–	3,61	–	–	6,3
2	28	7	30,3	-2,3	1	5,29	-4,2	17,6	8,2
3	33	10	33,6	-0,6	0	0,36	1,7	2,9	1,8
4	38	12	35,8	2,2	1	4,84	2,8	7,6	5,8
5	40	15	39,1	0,9	0	0,81	-1,3	1,7	2,3
6	42	18	42,4	-0,4	0	0,16	-1,3	1,7	1,0
7	44	20	44,6	-0,6	1	0,36	-0,2	0,04	1,4
8	49	23	47,9	1,1	1	1,21	1,7	2,9	2,3
9	47	24	49,0	-2,0	–	4,0	-3,1	9,6	4,3
Σ	351	134		0,2	4	20,6		44,24	33,4

Проверка равенства математического ожидания уровней ряда остатков нулю осуществляется с использованием t критерия Стьюдента:

$$t_p = \frac{|\bar{E}|}{\sigma_E} \cdot \sqrt{n}$$

где $\sigma_E = \sqrt{\frac{\sum_{t=1}^n (E(t) - \bar{E})^2}{n-1}}$ - среднеквадратическое отклонение уровней

остаточного ряда; \bar{E} - среднее значение уровней остаточного ряда, т.е. $\bar{E} = \frac{\sum_{t=1}^n E(t)}{n}$

Если $t_p < t_\alpha$, то $M(E) \approx 0$, где t_α – критерий Стьюдента, α - уровень значимости.

На практике, если $|\bar{E}| < 0,5$, то критерий Стьюдента выполняется и, значит, гипотеза о равенстве математического ожидания значений остаточного ряда нулю выполняется.

Вычислим:

$$M(E) = \bar{E} = \frac{\sum E(t)}{n} = \frac{0,2}{9} = 0,02$$

Так как $|\bar{E}| = 0,02 < 0,5$, то $t_p < t_\alpha$ – критерий Стьюдента выполняется, то есть $M(E) \approx 0$.

Проверка случайного характера отклонения может осуществляться различными методами: методом серий знаков; поворотных точек (пиков) и др.

Рассмотрим метод поворотных точек:

Значение $E(t)$ - является поворотной точкой, если $E(t) > E(t-1)$ и $E(t) > E(t+1)$ или $E(t) < E(t-1)$ и $E(t) < E(t+1)$:

Число поворотных точек P должно удовлетворять строгому неравенству:

$$P > \left[2 \cdot \frac{(n-2)}{3} - 1,96 \sqrt{\frac{16n-29}{90}} \right]$$

(Квадратные скобки означают, что берётся целая часть числа, т. е. наибольшее целое число не превосходящее данное; числом 1 отмечаются поворотные точки).

Проверку случайности уровней ряда остатков проведём на основе критерия поворотных точек:

$$P > \left[2 \cdot \frac{(n-2)}{3} - 1,96 \sqrt{\frac{16n-29}{90}} \right] = \left[\frac{2 \cdot (9-2)}{3} - 1,96 \sqrt{\frac{16 \cdot 9 - 29}{90}} \right] = [2,4] = 2$$

Так как $P = 4 > 2$, то свойство случайности выполняется.

Уровни остаточного ряда независимы, если отсутствует автокорреляция. Наличие автокорреляции устанавливается с помощью d -критерия Дарбина-Уотсона:

$$d = \frac{\sum_{t=2}^n (E(t) - E(t-1))^2}{\sum_{t=1}^n E(t)^2} = \frac{44,24}{20,6} \approx 2,15$$

Полученное значение d сравнивается с двумя критическими табличными значениями d_1 и d_2 . (если $n \leq 10$; $d_1 = 1,08$, $d_2 = 1,36$.)

Если $d > 2$, то находят $d' = 4 - d$ и к найденным d и d' применяют критерий Дарбина-Уотсона.

Так как $d = 2,15$ не попадает в интервал от d_2 до 2, т.е. в $[1,36; 2]$, то найдем $d' = 4 - 2,15 = 1,85 \in [1,36; 2]$, то есть уровни ряда являются независимыми (автокорреляции нет).

Наличие нормального закона распределения в остаточном ряду линейной функции устанавливается с помощью R/S -критерия:

$$RS = \frac{E_{\max} - E_{\min}}{S_E}$$

где E_{\max} - максимальный уровень ряда остатков;

E_{\min} - минимальный уровень ряда остатков;

$$S_E = \sqrt{\frac{\sum E(t)^2}{n-1}} \text{ - «исправленное» среднее квадратическое отклонение}$$

Вычисленное значение сравнивают с табличными верхней и нижней границами (фрагмент таблицы для уровня значимости $\alpha=0,05$):

n	Нижняя граница	Верхняя граница
10	2,67	3,68
20	3,18	4,49
30	3,47	4,89

Если значение этого критерия попадает между табулированными границами с заданным уровнем значимости, то гипотеза о нормальном законе распределения ряда остатков принимается.

Вычислим значение RS–критерия:

$$S_E = \sqrt{\frac{20,6}{8}} = \sqrt{2,58} \approx 1,6, \text{ тогда } RS = \frac{2,2 - (-2,3)}{1,6} = 2,8$$

Так как $RS = 2,8 \in [2,7; 3,7]$, то есть значение RS–критерия попадает между табулированными границами с заданным уровнем вероятности (для $n \leq 10, p = 70\%$), то гипотеза о нормальном распределении ряда остатков принимается, что позволяет строить доверительный интервал прогноза.

Таким образом, в ходе исследования было установлено, что все четыре условия выполняются, значит, построенная линейная модель парной регрессии адекватна.

Для характеристики точности модели воспользуемся средней относительной ошибкой: $\overline{E_{\text{оми}}} = \frac{1}{n} \sum \frac{|E(t)|}{y_t} \cdot 100\%$

Для нашего примера: $\overline{E_{\text{оми}}} = \frac{33,4}{9} = 3,7\% < 5\%$, что свидетельствует о "хорошей" точности модели.

Оценку качества построенной модели дает коэффициент детерминации:

$$R^2 = 1 - \frac{S_E^2}{S_y^2} = 1 - \frac{\sum E(t)^2}{\sum (y - \bar{y})^2}, \quad R^2 = 1 - \frac{20,6}{438} \approx 0,95$$

Если $R \approx 1$, то модель качественная. Коэффициент R^2 показывает долю вариации результативного признака под воздействием изучаемых факторов.

Коэффициент $R^2 = 0,95$ показывает, что более 95% вариации зависимой переменной (y) учтено в модели и происходит под влиянием фактора, включённого в модель, а остальные 5% приходится на другие факторы, не включённые в модель.

Так как $R = \sqrt{0,95} = 0,98$, то есть $R \approx 1$, то модель качественная.

Используя коэффициент детерминации можно проверить значимость уравнения регрессии с помощью критерия Фишера.

Для этого выполняется сравнение фактического $F_{\text{факт}}$ и критического (табличного) $F_{\text{табл}}$ значений F-критерия Фишера. $F_{\text{факт}}$ определяется из соотношения:

$$F_{\text{факт}} = \frac{R^2 / m}{(1 - R^2) / (n - m - 1)}$$

Если $F_{\text{табл}} < F_{\text{факт}}$, то гипотеза о случайной природе оцениваемых характеристик отклоняется и признается их статистическая значимость и надежность.

Важную роль при оценке влияния факторов играют коэффициенты регрессионной модели. Однако непосредственно с их помощью нельзя сопоставить факторы по степени их влияния на результативный признак из-за различия единиц измерения и разной степени колеблемости. Для устранения таких различий применяются средние частные коэффициенты эластичности и бета-коэффициент.

Для построенной модели регрессии рассчитаем коэффициент эластичности и бетта коэффициент.

Вычислим коэффициент эластичности модели: $\mathcal{E}_{yx_1}^{\wedge} = a_1 \frac{\bar{x}_1}{\bar{y}}$,

В нашем примере $a_1 = 1,1$, $\bar{x}_1 = 14,9$, $\bar{y} = 39$, поэтому

$$\mathcal{E}_{yx_1}^{\wedge} = \frac{1,1 \cdot 14,9}{39} \approx 0,42$$

Вычисленный коэффициент эластичности показывает, что при увеличении x_1 на 1% результативный признак y в среднем увеличивается на 0,42%.

Вычислим бетта β - коэффициент по формуле:

$$\beta_{yx_i}^{\wedge} = a_i \frac{S_{x_i}}{S_y} = a_i \sqrt{\frac{\sum (x_i - \bar{x}_i)^2}{\sum (y_t - \bar{y})^2}}$$

где S_{x_i} и S_y – средние квадратические ошибки выборки величин x_i и y .

Используя табл.3 получим:

$$\beta_{yx_1}^{\wedge} = a_1 \sqrt{\frac{\sum (x_1 - \bar{x}_1)^2}{\sum (y - \bar{y})^2}} = 1,1 \sqrt{\frac{376,8}{438}} = 1,1 \sqrt{0,96} = 1,1 \cdot 0,93 \approx 1,02$$

Бетта β – коэффициент показывает, на какую часть величины своего среднего квадратического отклонения изменится в среднем значение результативного признака y при изменении факторного признака x_i на величину его среднеквадратического отклонения при фиксированном на постоянном уровне значений остальных независимых переменных.

Указанные коэффициенты позволяют проранжировать факторы по степени влияния факторов на зависимую переменную.

При использовании построенной модели для прогнозирования делается предположение о сохранении в период прогнозирования и на период упреждения существовавших ранее взаимосвязей переменных [3].

Прогноз состоит из двух частей: точечный и интервальный прогнозы.

Для прогнозирования зависимой переменной на k шагов вперед необходимо знать прогнозные значения всех входящих в модель факторов. Их оценки могут быть получены методами экспертных оценок или непосредственно заданы исследователем или могут быть получены на основе экстраполяционных методов, например, с использованием средних абсолютных приростов (САП) факторных признаков [20]:

$$x(n+k) = x(n) + \text{САП} \cdot k, \quad \text{где } \text{САП} = (x_n - x_1)/(n-1)$$

Вычислим САП и получим прогнозные оценки фактора на основе величины его среднего прироста: $x(n+k) = x(n) + \text{САП} \cdot k$,

$$\text{САП} = (24-5)/(9-1) = 19/8 \approx 2,4$$

$$\text{Если } k=1, \text{ то } x(9+1) = x(10) = x(9) + 2,4 \cdot 1 = 24 + 2,4 = 26,4$$

$$\text{Если } k=2, \text{ то } x(9+2) = x(11) = x(9) + 2,4 \cdot 2 = 24 + 4,8 = 28,8$$

Для получения прогнозных (точечных) оценок зависимой переменной по модели: $\hat{y}(n+k) = a_0 + a_1 x(n+k)$, подставим в неё найденные прогнозные значения фактора: $\hat{y}(10) = 22,6 + 1,1 \cdot 26,4 = 51,6$

$$\hat{y}(11) = 22,6 + 1,1 \cdot 28,8 = 54,3$$

Для линейной модели регрессии область возможных значений результативного показателя при известных значениях факторов, то есть доверительный интервал прогноза рассчитывается следующим образом. Оценивается величина отклонения от линии регрессии (обозначим ее U):

Величина $U(k)$ для парной модели регрессии имеет вид:

$$U(k) = t_\alpha \cdot \overline{S}_E \cdot \sqrt{1 + \frac{1}{n} + \frac{(x(n+k) - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$$

где \overline{S}_E – ошибка аппроксимации, которая вычисляется по формуле: $\overline{S}_E = \sqrt{\frac{\sum E(t)^2}{n - m - 1}}$

n - число наблюдений, m –число факторов в модели;
 k - количество шагов;
 x_t – наблюдаемое значение факторного признака в момент t ;
 $x(n+k)$ - прогнозное значение фактора на k шагов вперед;
 t_α -табличное значение t -статистики Стьюдента при заданном уровне значимости α и числа наблюдений n .

Если исследователь задает уровень вероятности попадания прогнозируемой величины внутрь доверительного интервала, равный 70%, то $t_\alpha = 1,05$; если 95%, то $t_\alpha = 1,96$; если 99%, то $t_\alpha = 2,65$.

Доверительный интервал прогноза для шага прогнозирования k имеет следующие границы:

верхняя граница прогноза: $\hat{y}(n+k)+U(k)$,

нижняя граница прогноза: $\hat{y}(n+k)-U(k)$, причем $U(1) < U(2)$.

Доверительный интервал прогноза имеет следующие границы:

при $k = 1$:

Верхняя граница прогноза: $y_p(9+1)+U(1)=51,6+3,7=55,3$

Нижняя граница прогноза: $y_p(9+1)-U(1)=51,6-3,7=47,9$

Величина $U(k)$ для найденной модели регрессии имеет вид при $k=1$:

$$U(1) = 1,05 \cdot 2,9 \sqrt{1 + \frac{1}{9} + \frac{(26,4 - 14,9)^2}{376,8}} = 3,04 \cdot 1,2 \approx 3,7$$

Доверительный интервал прогноза имеет следующие границы:

при $k = 2$

Верхняя граница прогноза: $y_p(9+4) + U(2) = 54,3+3,9 = 58,2$

Нижняя граница прогноза: $y_p(9+2) - U(2)=54,3-3,9 = 50,4$

Величина $U(k)$ для найденной модели регрессии имеет вид при $k = 2$:

$$U(2) = 1,05 \cdot 2,9 \sqrt{1 + \frac{1}{9} + \frac{(28,8 - 14,9)^2}{376,8}} = 3,04 \cdot 1,3 \approx 3,9$$

Сведем результаты точечного и интервального прогнозирования в таблицу прогноза:

k	x(n+k)	$\hat{y}(n+k)$	Нижняя граница	Верхняя граница
1	26,4	51,6	47,9	55,3
2	28,8	54,3	50,4	58,2

Если построенная регрессионная модель адекватна и прогнозные оценки факторов достаточно надежны, то с выбранной исследователем вероятностью можно утверждать, что при сохранении сложившихся закономерностей развития прогнозируемая величина попадет в интервал, образованный нижней и верхней границами.

Отообразим на графиках фактические данные, модель и результаты прогнозирования (рис.2).

Для построения модели достаточно двух точек:

при $t = 1$ ($x_1 = 5$), $\hat{y}(1) = 22,6 + 1,1 \cdot 5 = 28,1$

при $t = 9$ ($x_1 = 24$), $\hat{y}(9) = 22,6 + 1,1 \cdot 24 = 49,0$

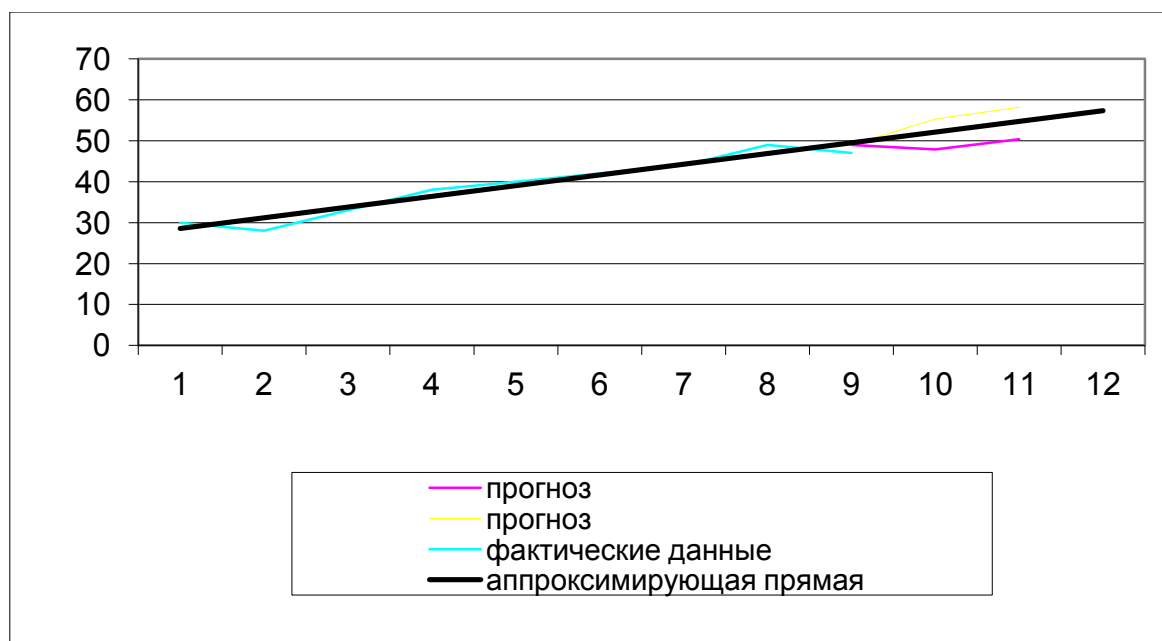


Рисунок 2 - Результаты прогнозирования по линейной модели регрессии

С вероятностью 0,7 можно гарантировать, что при сохранении сложившихся закономерностей развития прогнозируемая величина y в

следующем году примет значение из интервала (47,9;55,3), а ещё через год из интервала (50,4;58,2).

Чтобы построить нелинейную парную регрессию: $\hat{y}_t = a \cdot b^x$ необходимо привести ее к линейной [31]. Для этого прологарифмируем обе части равенства по основанию 10: $lgy = lga + xlg b$.

Обозначим: $lg y = Y$, $lg a = A$, $lg b = B$

Запишем предыдущее равенство в виде: $Y = A + B \cdot x$; таким образом мы получили линейное уравнение регрессии, где $B = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2}$, $A = \bar{y} - B \cdot \bar{x}$

Для вычисления коэффициентов A и B, составим табл. 5, в которой приведем промежуточные вычисления:

Таблица 5 – Расчетная таблица для вычисления коэффициентов

t	y_t	x	$y = lgy$	$x \cdot y$	x^2	$y - \bar{y}$	$(y - \bar{y})^2$
1	30	5	1,4771	7,3855	25	-0,1054	0,0111
2	28	7	1,4472	10,1304	49	-0,1353	0,0183
3	33	10	1,5185	15,185	100	-0,064	0,0041
4	38	12	1,5682	18,8184	144	-0,143	0,0002
5	40	15	1,6021	24,0315	225	0,0196	0,0004
6	42	18	1,6232	29,218	324	0,0407	0,0017
7	44	20	1,6435	32,87	400	0,061	0,0037
8	49	23	1,6902	38,875	529	0,1077	0,0116
9	47	24	1,6721	40,1304	576	0,0896	0,0080
	351	134	14,2421	216,64	2372		0,0591

$$\overline{x \cdot y} = \frac{\sum xy}{n} = \frac{216,64}{9} = 24,07 \quad \overline{x^2} = 2372 : 9 = 263,6 \quad \bar{x} = 134 : 9 = 14,9 \quad \bar{y} = 14,2421 : 9 = 1,5825$$

$$\text{Вычислим } B = \frac{24,07 - 14,9 \cdot 1,5825}{263,66 - 14,9^2} = \frac{0,491}{41,65} = 0,012$$

$$A = 1,5825 - 0,012 \cdot 14,9 = 1,401$$

Так как $A = lga = m$, то $a = 10^m$; $B = lgb = p$, то $b = 10^p$

Отсюда, $\hat{y}(t) = 10^{1,401} \cdot (10^{0,012})^x$ - показательная модель.

Проверим качество этой модели, с помощью коэффициента

детерминации:
$$R^2 = 1 - \frac{\sum E(t)^2}{\sum (y - \bar{y})^2}.$$

Составим табл. 6 промежуточных вычислений:

Таблица 6 – Промежуточные вычисления

t	y _t	x	\hat{y}_p	E(t)	E(t) ²	E(t) /y _t ·100%
1	30	5	28,9	1,1	1,21	3,7
2	28	7	30,5	-2,5	6,25	8,9
3	33	10	33,2	-0,2	0,04	0,6
4	38	12	35,1	2,9	8,41	7,6
5	40	15	38,1	1,9	3,61	4,8
6	42	18	41,4	0,6	0,36	1,4
7	44	20	43,7	0,3	0,09	0,7
8	49	23	47,5	1,5	2,25	3,1
9	47	24	51,4	-4,4	19,36	9,4
Σ	351	134			41,58	40,2

$$\hat{y}(t) = 10^{1,401} \cdot (10^{0,012})^x; E(1) = 30 - 28,9 = 1,1$$

$$\hat{y}_p(1) = 10^{1,401} \cdot (10^{0,012})^5 = 10^{1,401} \cdot 10^{0,06} = 10^{1,461} = 28,9; E(2) = 28 - 30,5 = -2,5$$

$$\hat{y}_p(2) = 10^{1,401} \cdot (10^{0,012})^7 = 10^{1,485} = 30,5; E(3) = 33 - 33,2 = -0,2$$

$$\hat{y}_p(3) = 10^{1,401} \cdot (10^{0,012})^{10} = 10^{1,521} = 33,2; E(4) = 38 - 35,1 = 2,9$$

$$\hat{y}_p(4) = 10^{1,401} \cdot (10^{0,012})^{12} = 10^{1,545} = 35,1; E(5) = 40 - 38,1 = 1,9$$

$$\hat{y}_p(5) = 10^{1,401} \cdot (10^{0,012})^{15} = 10^{1,581} = 38,1; E(6) = 42 - 41,4 = 0,6$$

$$\hat{y}_p(6) = 10^{1,401} \cdot (10^{0,012})^{18} = 10^{1,617} = 41,4; E(7) = 44 - 43,7 = 0,3$$

$$\hat{y}_p(7) = 10^{1,401} \cdot (10^{0,012})^{20} = 10^{1,641} = 43,7; E(8) = 49 - 47,5 = 1,5$$

$$\hat{y}_p(8) = 10^{1,401} \cdot (10^{0,012})^{23} = 10^{1,677} = 47,5; E(9) = 47 - 51,4 = -4,4$$

$$\hat{y}_p(9) = 10^{1,401} \cdot (10^{0,012})^{24} = 10^{1,689} = 48,9$$

$$\text{Коэффициент детерминации: } R^2 = 1 - \frac{36,78}{438} = 0,92$$

Так как $R = \sqrt{0,92} = 0,96 \approx 1$, то модель качественная.

Вычислим среднюю относительную ошибку аппроксимации:

$$\overline{E_{\text{отн}}} = \frac{1}{n} \sum \frac{|E(t)|}{y_t} \cdot 100\% = \frac{1}{9} \cdot 40,2 = 4,47\% < 5\%, \text{ что свидетельствует о "хорошей"}$$

точности модели; по ней можно строить прогноз.

Проверим значимость модели регрессии с использованием F–критерия Фишера:

$$F_p = \frac{R^2 / m}{(1 - R^2) / (n - m - 1)} = \frac{0,92 / 1}{(1 - 0,92) / (9 - 1 - 1)} = \frac{0,92}{0,08 : 7} = 80,5$$

Если $F_p = 80,5 > F_{\text{табл}} = 5,59$, то уравнение регрессии значимое

Для построения точечного прогноза [49], сначала получим прогнозные оценки фактора на основе величины: $x_{\text{прогн}} = 110\% \cdot x_n$ с помощью формулы: $x(n+k) = 1,1 \cdot x(n)$

$$\text{Если } k = 1; \quad x(9+1) = x(10) = 1,1 \cdot x(9) = 1,1 \cdot 24 = 26,4$$

$$\text{Если } k = 2; \quad x(9+2) = x(11) = 1,1 \cdot x(10) = 1,1 \cdot 26,4 = 29,1$$

Для получения прогнозных оценок зависимой переменной по модели, подставим в неё найденные прогнозные значения фактора.

$$\hat{y}(n+k) = 10^{1,401} \cdot (10^{0,012})^{x(n+k)}$$

$$\hat{y}(10) = 10^{1,401} \cdot (10^{0,012})^{26,4} = 10^{1,401} \cdot 10^{0,3168} = 10^{1,7178} = 52,2$$

$$\hat{y}(11) = 10^{1,401} \cdot (10^{0,012})^{29,1} = 10^{1,401} \cdot 10^{0,3492} = 10^{1,7502} = 56,3$$

Доверительный интервал прогноза имеет следующие границы:

$$\text{Верхняя граница: } \hat{y}_p(n+k) + U(k)$$

$$\text{Нижняя граница: } \hat{y}_p(n+k) - U(k)$$

$$\text{Найдём } U(1) = 1,05 \cdot 2,43 \sqrt{1 + \frac{1}{9} + \frac{(26,4 - 14,9)^2}{376,8}} = 2,55 \cdot 1,21 = 3,1$$

$$\overline{S_E} = \sqrt{\frac{\sum E(t)^2}{n-m-1}} = \sqrt{\frac{41,58}{9-1-1}} \approx 2,43$$

$$U(2) = 1,05 \cdot 2,43 \sqrt{1 + \frac{1}{9} + \frac{(29,1 - 14,9)^2}{376,8}} = 2,55 \cdot 1,28 \approx 3,3$$

Таблица прогноза:

k	$x(n+k)$	$\hat{y}(n+k)$	Нижняя граница	Верхняя граница
1	26,4	52,2	59,1	55,2
2	29,1	56,3	53	59,6

Отобразим (рис.3) на графиках фактические данные, модель и результаты прогнозирования.

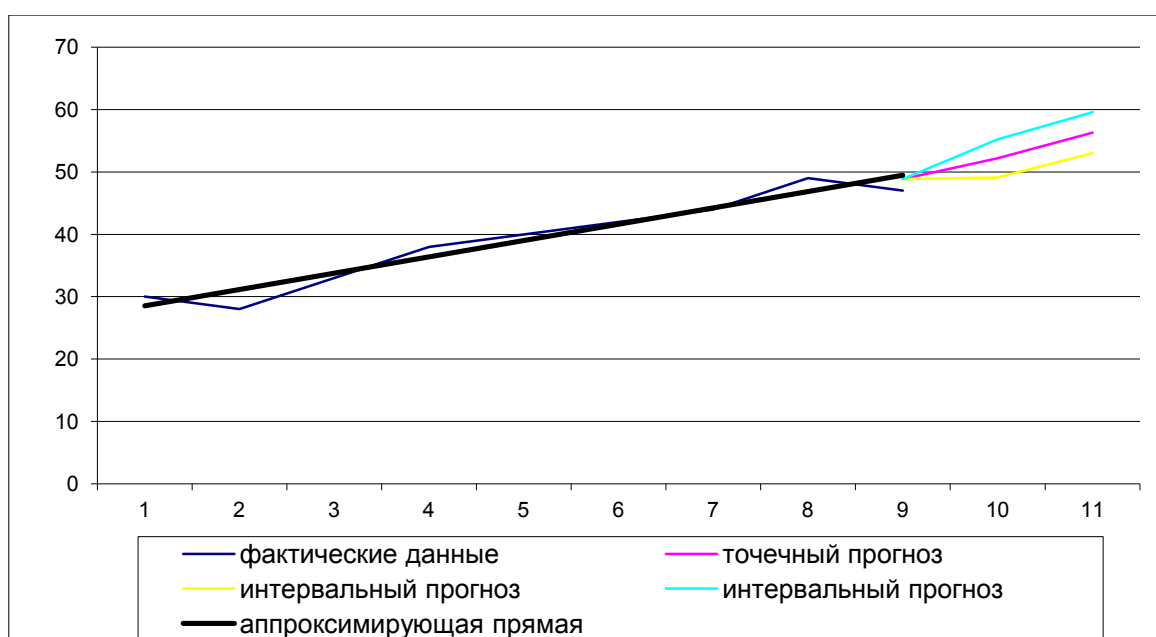


Рисунок 3 – Результаты прогнозирования по показательной модели

Результаты эконометрического исследования показали, что с вероятностью 0,7 можно гарантировать, что в следующем году результативный признак у примет значение из интервала (50,9;57,1), а ещё через год из интервала (53,7;60,3).

1.5. Принятие эффективных управленческих решений на основе регрессионной модели

Рыночная экономика требует улучшения использования статистической и экономической информации, проведения эконометрических расчетов будущих результатов хозяйственной деятельности и обоснования путей их достижения. Такие расчеты содействуют правильной оценке влияния факторов на различные показатели, характеризующие эффективное развитие их взаимосвязи. Рыночная экономика требует от руководителя - предпринимателя умения использовать статистические данные с целью построения, обоснования моделей и использования их для прогнозирования. Возможность прогнозирования ситуации дает возможность руководителю принимать обоснованные управленческие решения [5].

Сегодня деятельность в любой области экономики требует от специалиста применения современных математических методов и моделей [21]. Современный руководитель, предприниматель должен хорошо разбираться в экономико-математических методах, уметь их практически применять для моделирования реальных экономических ситуаций .

Все шире используются методы прогнозирования в деятельности плановых, аналитических, маркетинговых отделов производственных предприятий и объединений. В современных условиях управленческие решения должны приниматься лишь на основе тщательного анализа имеющейся информации, поэтому широкое применение в настоящее время имеют статистические программные пакеты Statistica, Eviews и Excel и др. для решения эконометрических задач проведения анализа данных, вычисления параметров модели, анализа модели [29,34].

Большинство процессов и явлений в экономике находятся в постоянной, взаимной, объективной связи. Исследование зависимостей и взаимосвязей между объективно существующими явлениями и процессами

играет большую роль в повышении экономической эффективности функционирования предприятия в условиях рынка [18]. Эффективная хозяйственная деятельность невозможна без оценки этих связей, выявление их тенденций и разработки прогнозов [49].

По направлению и сложности связей между внутренними (эндогенными) переменными и внешними (экзогенными) переменными интерес вызывают регрессионные модели.

Различают уравнения (модели) парной и множественной регрессии. Если для обозначения эндогенных переменных использовать y , а для экзогенных переменных x , то в случае линейной модели, уравнение парной регрессии имеет вид: $y_t = a_0 + a_1x_{t1} + \varepsilon_t$, а уравнение множественной регрессии имеет вид: $y = a_0 + a_1x_{t1} + a_2x_{t2} + \dots + a_mx_{tm} + \varepsilon_t$.

Чтобы построить эконометрическую модель с последующим использованием ее для прогнозирования, необходимо решить следующие задачи [4,51]:

- собрать исходные данные;
- отобрать факторы, наиболее влияющие на результирующий показатель;
- выбрать тип эконометрической модели, в том числе формирование структуры и форм функциональной связи между переменными, включенными в модель;
- сделать статистический анализ выбранной модели, то есть оценить неизвестные параметры модели на базе имеющихся статистических данных;
- проверить качество модели: сопоставить реальные и модельные данные, проверить адекватность модели, оценить точность модельных данных;
- использовать модель для экономического анализа и прогнозирования.

Рассмотрим исследование зависимости издержек производства от следующих факторов: основных производственных фондов и численности занятых в производстве [20].

На одном из предприятий города поквартально произведена выборка [20] - издержек производства (тыс. руб.): 1815,48; 807,53; 1462,61; 1307,49; 1710,40; 1430,44; 1165,42; 1309,35; 1464,13; 1059,87; 1010,98; 1515,41; 1746,04; 2116,98; 709,53; 1481,01; 1648,89; 1158,94; 1924,64; 1050,83, выборка - основных производственных фондов (тыс. руб.): 1123,56; 442,16; 1038,71; 791,07; 1137,65; 908,92; 681,17; 863,25; 965,68; 607,24; 541,30; 1060,70; 1108,75; 1460,49; 453,78; 968,32; 1094,29; 682,56; 1268,93; 666,37 и выборка - численности занятых в производстве(чел.): 204, 180, 189, 144, 207, 165, 124, 157, 176, 110, 108, 193, 202, 226, 146, 176, 110, 108, 193, 202 за пять лет.

Пусть результирующий фактор $y(t)$ - издержки производства; факторные признаки: $x_1(t)$ - основные производственные фонды; $x_2(t)$ - численность занятых в производстве.

Наибольшие трудности в использовании аппарата множественного регрессионного анализа возникают при наличии мультиколлинеарности факторов [48]. Чем сильнее мультиколлинеарность факторов, тем менее надежна оценка распределения суммы объясненной вариации по отдельным факторам с помощью метода наименьших квадратов.

Для оценки мультиколлинеарности факторов была использована матрица парных коэффициентов корреляции между факторами $y(t)$, $x_1(t)$, $x_2(t)$ (табл. 7).

Таблица 7 - Матрица парных коэффициентов корреляции

	y	x_1	x_2
y	1,000	0,985	0,808
x_1	0,985	1,000	0,839
x_2	0,808	0,839	1,000

Так как $r_{yx_1} = 0,985 > 0$, то связь прямая; $|r_{yx_1}| = 0,985 \approx 1$, то связь тесная. Так как $r_{yx_2} = 0,808 > 0$, то связь прямая; $|r_{yx_2}| = 0,808 \approx 1$ – связь тесная. Так как $|r_{x_1x_2}| = 0,839 > 0,75$, парная корреляция между малозначимыми объясняющими переменными достаточно высока, т.е. имеет место мультиколлинеарность, оба фактора в модель включать нельзя. В модель включим фактор x_1 , потому что он наиболее тесно связан с зависимой переменной y , т.е. $|r_{yx_1}| > |r_{yx_2}|$.

Построение уравнения регрессии сводится к оценке ее параметров. Для оценки параметров регрессии используют метод наименьших квадратов (МНК). МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака от теоретических минимальна, то есть

$$\sum_{t=1}^n (y_t - (a_0 - a_1x_1))^2 \rightarrow \min.$$

Для того, чтобы полученные МНК оценки параметров обладали желательными свойствами, делают следующие предположения об отклонениях ε_t [34]:

- 1) величина ε_t есть величина случайная;
- 2) математическое ожидание ε_t равно нулю: $M(\varepsilon_t) = 0$;
- 3) дисперсия ε_t постоянна: $D(\varepsilon_t) = D(\varepsilon_j) = \sigma^2$ для всех t и j ;
- 4) значения ε_t независимы между собой;
- 5) величина ε_t есть нормально распределенная случайная величина.

Если регрессионная модель удовлетворяет перечисленным предположениям, то оценки параметров, полученные с помощью МНК, обладают следующими свойствами [23]:

- 1) оценки являются несмещенными, т.е. математическое ожидание оценки каждого параметра равно истинному значению параметра;

2) оценки состоятельны, надежность оценки при увеличении выборки растет;

3) оценки эффективны, они имеют наименьшую дисперсию в классе всех линейных несмещенных оценок.

Построим линейную регрессионную модель $\hat{y}_t = a_0 + a_1 x_t$ для выбранного $x(t) = x_1(t)$. Оценки коэффициентов регрессии найдем по формулам, с помощью метода наименьших квадратов [4,16]:

$$a_1 = \frac{\sum (y - \bar{y})(x_1 - \bar{x}_1)}{\sum (x_1 - \bar{x}_1)^2}; a_0 = \bar{y} - a_1 \cdot \bar{x}_1,$$

т.е. $a_1 = \frac{1936416,234}{1477323,421} \approx 1,311$, а $a_0 = 1394,799 - 1,311 \cdot 893,245 = 223,969$

Выборочное уравнение регрессии, для выбранного x_1 примет вид:

$$\hat{y}_t = 223,969 + 1,311x_t$$

Для решения проблемы спецификации модели была построена степенная модель: $\hat{y}_t = 1331 \times x^{0,0017}$, показательная модель: $\hat{y}_t = 555 \times 1,00099^{x_1}$ и линейная модель: $\hat{y}_t = 223,969 + 1,311x_t$. Оценки коэффициентов указанных регрессий также были найдены методом наименьших квадратов (при этом нелинейные модели были приведены к линейным методом линеаризации) [21,38].

Основные результаты статистического анализа по каждой модели приведены в сводной табл. 8.

Таблица 8 - Результаты статистического анализа

Модель	Коэффициент детерминации	Критерий Фишера	Средняя относительная ошибка
Линейная	0,97	582	3,97
Степенная	0,31	8,9	23,89
Показательная	0,94	313	5,28

Очевидно, лучшей моделью является линейная модель:

$$\hat{y}_t = 223,969 + 1,311x_t$$

Качество модели оценивается по адекватности и точности модели на основе анализа остатков регрессии e_t : $e_t = y_t - \hat{y}_t$ - остаточный ряд.

Анализ остатков позволяет получить представление, насколько хорошо подобрана сама модель и насколько правильно выбран метод оценки коэффициентов.

Соответствует (адекватна) ли эта модель реальному экономическому процессу? Модель адекватна, если остаточная компонента удовлетворяет четырем свойствам случайной компоненты остаточного ряда, а именно [51]:

- равенство математического ожидания уровней ряда остатков нулю;
- случайность уровней ряда остатков;
- независимость уровней ряда остатков;
- соответствие ряда остатков нормальному закону распределения.

Проверка равенства математического ожидания уровней ряда остатков нулю, осуществляется с использованием формулы:

$$M(e) = \bar{e} = \frac{\sum e(t)}{n}$$

$$\text{т.е. } M(e) = \frac{0,000000000001}{20} \approx 0,0000000000003$$

Так как $|\bar{e}| = 0,0000000000003 < 0,5$, то $M(e) \approx 0$.

Проверка случайности уровней ряда остатков, проводится на основе критерия поворотных точек:

$$P > \left[2 \cdot \frac{(n-2)}{3} - 1,96 \sqrt{\frac{16n-29}{90}} \right]$$

$$\text{т.е. } P > \left[\frac{2 \cdot (20 - 2)}{3} - 1,96 \sqrt{\frac{16 \cdot 20 - 29}{90}} \right] = [8,5] = 8$$

Так как число поворотных точек $P = 10 > 8$, то свойство случайности выполняется.

Независимость уровней ряда остатков проверяется с помощью d -критерия Дарбина-Уотсона:

$$d = \frac{\sum_{t=2}^n (e(t) - e(t-1))^2}{\sum_{t=1}^n e(t)^2}$$

$$\text{т.е. } d = \frac{171274,527}{76720,140} \approx 2,232$$

Так как $d = 2,232$ не попадает в интервал от d_2 до 2, то найдем $d' = 4 - 2,232 = 1,768 \in [1,411; 2]$, то есть уровни ряда являются независимыми.

Соответствие ряда остатков нормальному закону распределения определяется при помощи RS -критерия:

$$RS = \frac{e_{\max} - e_{\min}}{S_e},$$

$$\text{т.е. } RS = \frac{118,794 - (-122,858)}{63,544} = 3,803$$

Так как $3,803 \in [3,18; 4,49]$, то есть значение RS -критерия попадает между табулированными границами с заданным уровнем вероятности (для $n \leq 20$, $p = 70\%$), то гипотеза о нормальном распределении ряда остатков принимается.

Все четыре условия выполнены, то модель адекватна.

Для характеристики точности модели вычисляют среднюю относительную ошибку аппроксимации [17]:

$$\overline{E_{\text{отн}}} = \frac{1}{n} \sum \frac{|e(t)|}{y_t} \cdot 100\%$$

т.е. $\overline{E_{\text{отн}}} = \frac{1}{20} \times 79,332 \approx 3,967 \% < 5\%$, что свидетельствует о хорошей

точности модели. В среднем расчетные значения \hat{y}_t для линейной модели отличаются от фактических значений y_t на 3,97 %.

Оценку качества построенной модели дает коэффициент детерминации:

$$R^2 = 1 - \frac{\sum e(t)^2}{\sum (y_t - \bar{y})^2}$$

т.е. $R^2 = 1 - \frac{\sum e(t)^2}{\sum (y_t - \bar{y})^2} = 1 - \frac{76720,140}{2614896,803} \approx 0,971$

Коэффициент $R^2 = 0,97$ показал, что более 97% вариации зависимой переменной учтено в модели и происходит под влиянием фактора (основные производственные фонды), включённого в модель, а остальные 3% приходится на другие факторы, не включённые в модель.

Так как параметры уравнения регрессии значимы, уравнение значимо, показатели тесноты значимы, ошибка аппроксимации равна 3,97 %, коэффициент детерминации равен 0,97, уравнение регрессии значимо, то было сделано заключение, что построенная регрессионная модель зависимости издержек производства от основных производственных фондов - $\hat{y}_t = 223,969 + 1,311x_t$ может быть использована для анализа и прогноза.

При использовании построенной точной и адекватной модели линейной регрессии для прогнозирования делается предположение о

сохранении в период прогнозирования и на период упреждения существовавших ранее взаимосвязей переменных.

Различают точечное и интервальное прогнозирование [6]. В первом случае – это конкретное число, во втором – интервал, покрывающий истинное значение переменной с заданным уровнем доверия. Выделяют также условное и безусловное прогнозирование в зависимости от того, известны ли интересующие нас объясняющие переменные точно или приближенно. В нашем случае прогнозирование безусловно, т.к. значения регрессора известны точно [33].

Прогнозирование экономических показателей основано на идее экстраполяции. Под экстраполяцией обычно понимают распространение закономерностей, связей и соотношений, действующих в изучаемом периоде, за его пределы. В более широком смысле слова ее рассматривают как получение представлений о будущем на основе информации, относящейся к прошлому и настоящему. В процессе построения прогнозных моделей в их структуру иногда закладываются элементы будущего предполагаемого состояния объекта или явления, но в целом эти модели отражают закономерности, наблюдаемые в прошлом и настоящем, поэтому достоверный прогноз возможен лишь относительно таких объектов и явлений, которые в значительной степени детерминируются прошлым и настоящим [10].

Прогноз состоит из двух частей: точечный и интервальный прогнозы.

Для прогнозирования зависимой переменной на k шагов вперед необходимо знать прогнозные значения всех входящих в модель факторов. Их оценки могут быть получены методами экспертных оценок или непосредственно заданы исследователем или могут быть получены на основе экстраполяционных методов, например, с использованием средних абсолютных приростов (САП) факторных признаков [20]:

$$x(n+k) = x(n) + \text{САП} \cdot k, \quad \text{где САП} = (x_n - x_1) / (n-1)$$

Для построения точечного прогноза на два шага вперед по модели регрессии $\hat{y}_t = 223,969 + 1,311x_t$ вычисляют средний абсолютный прирост по формуле:

$$\text{САП} = (666,370 - 1123,560) / (20-1) = -457,19/19 \approx -24,06$$

$$\text{при } k = 1, \quad x(20+1) = x(21) = x(20) + (-24,06) \cdot 1 = 642,307$$

$$\text{при } k = 2, \quad x(20+2) = x(22) = x(20) + (-24,06) \cdot 2 = 618,245$$

Для получения точечных прогнозных оценок зависимой переменной по модели:

$$\hat{y}(n+k) = a_0 + a_1 x(n+k)$$

подставляют в неё найденные прогнозные значения фактора:

$$\hat{y}(21) = 223,969 + 1,311 \cdot 642,307 = 1065,880$$

$$\hat{y}(22) = 223,969 + 1,311 \cdot 618,245 = 1034,339$$

Очевидно, что точное совпадение фактических данных в будущем и прогностических точечных оценок маловероятно. Поэтому точечный прогноз должен сопровождаться двусторонними границами, т.е. указанием интервала значений, в котором с достаточной долей уверенности можно ожидать появления прогнозируемой величины. Установление такого интервала называется интервальным прогнозом.

Для линейной модели регрессии область возможных значений результативного показателя при известных значениях факторов, то есть доверительный интервал прогноза рассчитывается следующим образом [4]. Оценивается величина отклонения от линии регрессии - U:

Величина $U(k)$ для парной модели регрессии имеет вид:

$$U(k) = t_\alpha \cdot \overline{S_E} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x(n+k) - \bar{x})^2}{\sum (x_t - \bar{x})^2}}$$

где $\overline{S_E}$ - ошибка аппроксимации, которая вычисляется по формуле: $\overline{S_E} = \sqrt{\frac{\sum E(t)^2}{n-1}}$;

n - число наблюдений;

k - количество шагов;

x_t - наблюдаемое значение факторного признака в момент t ;

$x(n+k)$ - прогнозное значение фактора на k шагов вперед;

t_α -табличное значение t -статистики Стьюдента при заданном уровне значимости α и числа наблюдений n .

Доверительный интервал прогноза для шага прогнозирования k имеет следующие границы: верхняя граница прогноза: $\hat{y}(n+k) + U(k)$, нижняя граница прогноза: $\hat{y}(n+k) - U(k)$, причем $U(1) < U(2)$.

Итак, Доверительный интервал прогноза имеет следующие границы: при $k = 1$ верхняя граница прогноза: $y_p(20+1) + U(1) = 1065,880 + 71,654 = 1137,534$; нижняя граница прогноза: $y_p(20+1) - U(1) = 1065,880 - 71,654 = 994,225$.

$$\text{При } k = 1, U(1) = 1,05 \cdot 65,286 \cdot \sqrt{1 + \frac{1}{20} + \frac{(642,3075 - 893,245)^2}{1477323,421}} \approx 71,654$$

Доверительный интервал прогноза имеет следующие границы: при $k = 2$ верхняя граница прогноза: $y_p(20+2) + U(2) = 1034,339 + 71,935 = 1106,274$; нижняя граница прогноза: $y_p(20+2) - U(2) = 1034,339 - 71,935 = 962,404$.

$$\text{При } k = 2, U(2) = 1,05 \cdot 65,286 \cdot \sqrt{1 + \frac{1}{20} + \frac{(618,254 - 893,245)^2}{1477323,421}} \approx 71,935 ,$$

$$\text{где } \overline{S_E} = \sqrt{\frac{76720,140}{20-1-1}} = \sqrt{4262,23} \approx 65,286$$

Результаты прогнозирования покажем в табл. 9.

Таблица 9 - Результаты прогнозирования

k	$x(n+k)$	$\hat{y}(n+k)$	Нижняя граница	Верхняя граница
1	642,307	1065,880	994,225	1137,534
2	618,245	1034,339	962,404	1106,274

На графике (рис.4) покажем фактические данные, модель и результаты прогнозирования.

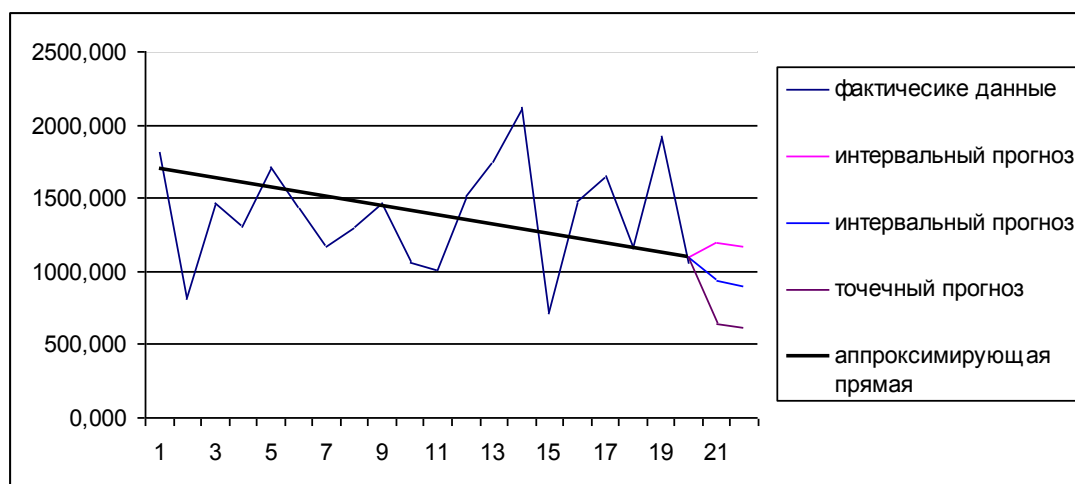


Рисунок 4 - Результаты аппроксимации и прогнозирования

Если построенная регрессионная модель адекватна и прогнозные оценки факторов достаточно надежны, то с выбранной вероятностью ($1 - \alpha = 0,7$) можно утверждать, что при сохранении сложившихся закономерностей развития прогнозируемая величина издержек производства в первом квартале следующего года примет значение из интервала (994,225; 1137,534) тыс. руб., а во втором - (962,404; 1106,274) тыс. руб. [16].

Полученный прогноз на основе точной регрессионной модели создает возможность для принятия обоснованных эффективных управленческих решений руководителем предприятия, тем самым, контролируя величину издержек на предприятии, уменьшая риск принятия неверных, необоснованных и субъективных решений. Также можно

учесть нежелательные тенденции, выявленные при прогнозировании и принять необходимые меры для их предупреждения, тем самым помешать осуществлению прогноза.

Таким образом, следует отметить, что развитие эконометрических прогнозных исследований в значительной степени обусловлено именно относительной простотой процедур разработки прогнозов, ясностью и определенностью использования их результатов в практике управления общественным развитием.

При разработке прогнозов важную роль играет процедура их верификации. Верификация предполагает обоснование достоверности прогноза, оценки его точности, качества. Одним из важнейших этапов верификации является выявление (или не выявление) систематической ошибки при формализованном описании (экстраполяции) тенденций развития исследуемого процесса. Такая ошибка может быть порождена, неправильно выбранной формой основного функционала эконометрической модели, ошибками при выборе состава входящих в нее факторов, погрешностями в оценках коэффициентов модели. Появление систематической ошибки в общем случае может быть вызвано и неверным подбором “прогнозного фона”.

Если прогноз разрабатывается на основе неуправляемых факторов то и сам процесс является неуправляемым. Прогнозы таких процессов часто называют поисковыми (исследовательскими). В этом случае система управления имеет возможность только приспособиться к его тенденциям прогнозируемого процесса, учесть их при обосновании управляющих мер для соответствующего объекта.

Если факторы являются управляемыми, то система управления может сознательно выбирать, формировать их уровни, определяя тем самым наиболее рациональную, “оптимальную” для объекта тенденцию

развития процесса в прогнозном периоде. Такие прогнозы обычно называют нормативными.

При частично управляемых факторах, возможности регулирования развития процесса в прогнозный период являются ограниченными.

В случае управляемых и частично управляемых факторов заметим, что эконометрические модели предоставляют исследователю фактически всю информацию относительно границ управления (диапазонах изменения факторов), эффективности их использования в управлении. При этом, показатель эффективности в некоторой степени может быть определен на основании значений коэффициентов эластичности переменной y по факторам x_{ik} (в части определения реакции y на изменения x_{ik}).

Другие составляющие эффективности (стоимость затрат на реализацию управления, результаты, выгоды, к которым оно приводит) выявляются на основе экономического анализа рассматриваемой проблемы [18].

Одной из важнейших характеристик качества прогноза является величина его доверительного интервала. Очевидно, что при прочих равных условиях, чем уже этот интервал, тем более обоснованным представляется и сам прогноз, и мероприятия по управлению рассматриваемым процессом.

:

ГЛАВА 2. МОДЕЛЬ МНОЖЕСТВЕННОЙ РЕГРЕССИИ В ЭКОНОМЕТРИЧЕСКОМ ИССЛЕДОВАНИИ

2.1. Методы и модели множественного регрессионного анализа

Экономические явления, как правило, определяются большим числом одновременно и совокупно действующих факторов. В связи с этим часто возникает задача исследования зависимости одной зависимой переменной Y от нескольких объясняющих переменных x_1, x_2, \dots, x_k . Эта задача решается с помощью множественного регрессионного анализа [1,25,47].

Например: Спрос на некоторый товар определяется не только ценой данного блага, но и ценами на заменяющие и дополняющие товары, доходом потребителей и многими другими факторами [21]. В этом случае рассматривается множественная регрессия:

$$M(Y|x_1, x_2, \dots, x_k) = f(x_1, x_2, \dots, x_k)$$

Множественная регрессия представляет собой регрессию результативного признака с двумя или более числом независимых переменных вида:

$$Y = f(x_1, x_2, \dots, x_k) + \varepsilon.$$

Уравнение зависимости y_t от $x_{t1}, x_{t2}, \dots, x_{tk}$ вида:

$$y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_k x_{tk} + \varepsilon_t, t = 1, \dots, n \quad (*)$$

или

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t, t = 1, \dots, n$$

где x_{tk} - значения регрессора x_k в наблюдении t , а $x_{t0} = 1, t = 1, \dots, n$;

y_t - t -ое наблюдение зависимой переменной y_t ;

β_j - неизвестные коэффициенты (параметры), $j = \overline{0, k}$;

ε_t - случайный член.

называется линейным регрессионным уравнением с k переменными или классической нормальной линейной моделью множественной регрессии (КЛНММР), если ε_t удовлетворяет предпосылкам регрессионного анализа.

Матричное описание регрессии облегчает как теоретический анализ, так и необходимые расчетные процедуры [25].

Введем обозначения:

$Y = (y_1 y_2 \dots y_n)'$ - матрица-столбец или вектор, значений зависимой переменной размера n ;

$\beta = (\beta_0 \beta_1 \dots \beta_k)'$ - $(k+1) \times 1$ вектор коэффициентов (параметров);

$\varepsilon = (\varepsilon_1 \dots \varepsilon_n)'$ - $n \times 1$ вектор ошибок (возмущений, остатков);

$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}$ - $n \times (k+1)$ матрица значений объясняющих

переменных.

Тогда в матричной форме модель (*) примет вид:

$$Y = X\beta + \varepsilon \quad (**)$$

Запишем предпосылки множественного регрессионного анализа в матричной форме:

1) $Y = X\beta + \varepsilon$ - спецификация модели.

2) ε - случайный вектор; X - неслучайная (детерминированная) матрица, столбцы которой должны быть линейно независимые, т.е. ранг матрицы X - максимальный, $r(X) = k+1$, т.е. ранг матрицы X равен числу ее столбцов.

3) а) $M(\varepsilon) = 0_n$, б) $M(\varepsilon\varepsilon') = \sigma^2 E_n$, где $E_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}$ - $n \times n$ единичная

матрица.

б) ε - нормально распределенный случайный вектор, т.е. $\varepsilon \approx N_n(0; \sigma^2 E_n)$.

4) Случайное отклонение должно быть независимо от объясняющих переменных: $\sigma_{\varepsilon x} = 0$.

Модель (*) или (**), удовлетворяющая приведенным предпосылкам (1-4), кроме 3) с о нормальном законе вектора возмущений ε , называется классической линейной моделью множественной регрессии (КЛММР).

На основе n наблюдений оценивается выборочное уравнение множественной регрессии:

$$\hat{y}_t = b_0 + b_1 x_{t1} + b_2 x_{t2} + \dots + b_k x_{tk} + e_t, t = \overline{1, n}$$

$$\text{или в матричной форме: } \hat{Y} = Xb + e,$$

где $b = (b_0 b_1 \dots b_k)'$ - оценки параметров $(\beta_0 \dots \beta_k)'$;

$e = (e_1 e_2 \dots e_n)$ - оценки возмущений $(\varepsilon_1 \dots \varepsilon_n)'$.

Воздействие неучтенных случайных факторов и ошибок наблюдений в модели множественной регрессии определяется с помощью дисперсии возмущений (случайных ошибок) или остаточной дисперсии σ^2 . Несмещенной оценкой этой дисперсии является выборочная остаточная дисперсия:

$$S^2 = \frac{1}{n - k - 1} \sum_{t=1}^n e_t^2,$$

где n - число наблюдений;

k - число независимых переменных.

Величина $S = \sqrt{S^2}$ называется стандартной ошибкой регрессии.

В соответствии с МНК минимизируется сумма квадратов остатков:

$$S = \sum_{t=1}^n e_t^2 = \sum (y_t - \hat{y}_t)^2 \rightarrow \min$$

$$\text{или } S = \sum_{t=1}^n e_t^2 = e'e = (Y - Xb)'(Y - Xb) \rightarrow \min,$$

после раскрытия скобок получим: $S = Y'Y - 2b'X'Y + b'X'Xb \rightarrow \min$.

Необходимым условием ее минимума является равенство нулю вектора частных производных $\frac{\partial S}{\partial b} = \left(\frac{\partial S}{\partial b_0} \frac{\partial S}{\partial b_1} \dots \frac{\partial S}{\partial b_k} \right) = 0$, т.е.

$\frac{\partial S}{\partial b} = -2X'Y' + 2X'Xb = 0$, откуда получаем систему нормальных уравнений в

матричной форме для определения вектора b : $X'Xb = X'Y$.

Решением этого уравнения является вектор b :

$$b = (X'X)^{-1} X'Y,$$

где $(X'X)^{-1}$ - матрица, обратная матрице $(X'X)$ коэффициентов системы;

$X'Y$ - матрица-столбец, или вектор ее свободных членов.

Таким образом, оценки параметров модели множественной регрессии в матричной форме определяются выражением:

$$b = (X'X)^{-1} X'Y,$$

а несмещенная оценка $S_{b_j}^2$ дисперсии $\sigma_{b_j}^2$ коэффициента регрессии b_j определится по формуле:

$$S_{b_j}^2 = S^2 (X'X)^{-1}_{jj},$$

где S^2 - несмещенная оценка параметра σ^2 (остаточная дисперсия);

$(X'X)^{-1}_{jj}$ - диагональный элемент матрицы $(X'X)^{-1}$.

Среднее квадратическое отклонение или стандартная ошибка коэффициента регрессии b_j примет вид:

$$S_{b_j} = S \sqrt{(X'X)^{-1}_{jj}},$$

Если предпосылки относительно случайного члена ε выполняются (кроме предпосылки о требовании нормальности закона распределения вектора возмущений ε), оценки параметров множественной регрессии являются несмещенными, состоятельными и эффективными.

Теорема Гаусса–Маркова: При выполнении предпосылок множественного регрессионного анализа оценка метода наименьших квадратов $b = (X'X)^{-1} X'Y$ является наиболее эффективной, т.е. обладает наименьшей дисперсией в классе линейных несмещенных оценок.

Рассмотрим интерпретацию воздействия отдельных факторов множественной регрессии на результирующий показатель.

Коэффициент регрессии β_i при переменной x_i выражает предельный прирост зависимой переменной Y при изменении переменной x_i , при условии постоянства других переменных. То есть β_i показывает на сколько увеличится в среднем Y при увеличении x_i на единицу.

На практике часто бывает необходимо сравнение влияния на зависимую переменную различных объясняющих переменных, когда последние выражаются разными единицами измерения. В этом случае используют стандартизированные коэффициенты регрессии (бета - коэффициенты β_j) и средние коэффициенты эластичности $\bar{\epsilon}_j (j = \overline{1, n})$.

Коэффициент β учитывает влияние анализируемых факторов на функцию y с учетом различий в уровне их колеблемости и рассчитывается по формуле:

$$\beta_j = b_j \cdot \frac{S_{x_j}}{S_y},$$

где β_j – коэффициент β j -го фактора;

$$S_{x_j} - \text{среднее квадратическое отклонение } j\text{-го фактора; } S^2_{x_j} = \frac{1}{n-1} \sum_{t=1}^n (x_t - \bar{x}_t)^2;$$

$$S_y - \text{среднее квадратическое отклонение функции; } S^2_y = \frac{1}{n-1} \sum_{t=1}^n (y_t - \bar{y}_t)^2;$$

b_j - коэффициент регрессии при j -м факторе.

Стандартизированный коэффициент регрессии β_j показывает, на сколько величин S_y изменится в среднем зависимая переменная Y при

увеличении только j -й объясняющей переменной x_j на величину своего S_{x_j} .

Из двух факторов x_i и x_j может быть исключен тот фактор, который имеет меньшее значение β – коэффициента.

Допустим, исключению подлежит один из мультиколлинеарных факторов x_i или x_j . Оба фактора управляемы на уровне предприятия, коэффициенты регрессии b_i и b_j статистически значимы. Фактор x_i более тесно связан с y , т.е. $r_{x_i,y} > r_{x_j,y}$, но при этом $\beta_{x_i} < \beta_{x_j}$. В этом случае обычно исключению подлежит фактор x_j .

Средние коэффициенты эластичности рассчитывается по формуле:

$$\bar{\varepsilon}_{yx_j} = b_j \frac{\bar{x}_j}{\bar{y}}$$

Коэффициент эластичности показывает, на сколько процентов (от средней) изменится в среднем функция Y при увеличении только аргумента x_j на 1%.

Указанные коэффициенты позволяют упорядочить факторы по степени влияния факторов на зависимую переменную.

2.2. Оценка качества множественной регрессионной модели и использование ее для экономического анализа и прогнозирования

В случаях, когда имеется одна независимая и одна зависимая переменные, естественной мерой зависимости (связи) между ними является выборочный (парный) коэффициент корреляции.

Использование множественной регрессии позволяет обобщить это понятие на случай, когда имеется несколько независимых переменных. В этом случае необходима корректировка, так как высокое значение коэффициента корреляции между зависимой и какой-либо независимой переменной может означать высокую степень линейной зависимости, но

может означать и то, что третья переменная оказывает значительное влияние на две первых и, что именно она служит основной причиной их высокой корреляции. Поэтому необходимо найти «чистую» корреляцию между двумя переменными, исключив влияние других факторов путем расчета коэффициента частной корреляции.

Коэффициенты частной корреляции для уравнения регрессии с двумя независимыми переменными рассчитываются как:

$$r_{yx_1(x_2)} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{(1 - r_{yx_2}^2) \cdot (1 - r_{x_1x_2}^2)}},$$

$$r_{yx_2(x_1)} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{(1 - r_{yx_1}^2) \cdot (1 - r_{x_1x_2}^2)}},$$

$$r_{x_1x_2(y)} = \frac{r_{x_1x_2} - r_{yx_1} \cdot r_{yx_2}}{\sqrt{(1 - r_{yx_1}^2) \cdot (1 - r_{yx_2}^2)}},$$

где $r_{yx_1(x_2)}$ - коэффициент частной корреляции между y и x_1 при исключенном влиянии x_2 ;

$r_{yx_2(x_1)}$ - коэффициент частной корреляции между y и x_2 при исключенном влиянии x_1 ;

$r_{x_1x_2(y)}$ - коэффициент частной корреляции между x_1 и x_2 , исключаящий влияние y .

Частные коэффициенты корреляции изменяются в пределах от -1 до 1.

Коэффициенты частной корреляции более высоких порядков можно определить через коэффициенты частной корреляции более низких порядков по следующей рекуррентной формуле:

$$r_{yx_i(x_1x_2\dots x_k)} = \frac{r_{yx_i(x_1x_2\dots x_{k-1})} - r_{yx_k(x_1x_2\dots x_{k-1})} r_{x_i x_k(x_1x_2\dots x_{k-1})}}{\sqrt{(1 - r_{yx_k(x_1x_2\dots x_{k-1})}^2) \cdot (1 - r_{x_i x_k(x_1x_2\dots x_{k-1})}^2)}}$$

Коэффициенты частной корреляции широко используются на стадии формирования модели, при отборе факторов.

Так, например, при построении многофакторной модели применяется метод исключения переменных, в ходе которого строится уравнение регрессии с полным набором переменных, затем рассчитывается матрица частных коэффициентов корреляции. Далее проверяется

статистическая значимость каждого из коэффициентов согласно t - критерию Стьюдента. Независимая переменная, имеющая наименьшую и несущественную корреляцию с зависимой переменной, исключается. Затем строится новое уравнение регрессии, и процедура продолжается до тех пор, пока не окажется, что все частные коэффициенты корреляции статистически значимы, т.е. существенно отличаются от нуля.

Тесноту совместного влияния факторов на результат оценивает индекс множественной корреляции:

$$R_{yx_1x_2\dots x_k} = \sqrt{1 - \frac{\sigma_{ocm.}^2}{\sigma_y^2}}$$

Значение $R_{yx_1x_2\dots x_k}$ лежит в пределах от 0 до 1 и должно быть больше или равно максимальному парному индексу корреляции:

$$R_{yx_1x_2\dots x_k} \geq r_{yx_i} \quad (i = \overline{1, k})$$

Проверка статистической значимости частного коэффициента корреляции суть проверки гипотезы о том, что он равен нулю:

$$H_0 : r_{yx_i(x_1x_2\dots x_k)} = 0$$

Рассчитывается статистика:

$$t = \frac{r_{yx_i(x_1x_2\dots x_k)}}{\sqrt{1 - (r_{yx_i(x_1x_2\dots x_k)})^2}} \cdot \sqrt{n - (k + 1)}$$

Вывод о значимости частного коэффициента корреляции делается при $|t| > t_\varepsilon$, где t_ε - соответствующее табличное значение t - распределения с $(n - (k + 1))$ степенями свободы.

В уравнении множественной регрессии содержится k объясняющих переменных. Дисперсию зависимой переменной можно разложить на объясненную и необъясненную составляющие:

$$\text{var}(y) = \text{var}(\hat{y}) + \text{var}(e).$$

Используя определение выборочной дисперсии, это уравнение можно представить в виде:
$$\sum_{t=1}^n (y_t - \bar{y})^2 = \sum_{t=1}^n (\hat{y}_t - \bar{y})^2 + \sum_{t=1}^n e^2_t .$$

Обозначим:

$TSS = \sum_{t=1}^n (y_t - \bar{y})^2$ - общий разброс зависимой переменной;

$RSS = \sum_{t=1}^n (\hat{y}_t - \bar{y})^2$ - разброс, объясненный регрессией;

$ESS = \sum_{t=1}^n e^2_t$ - разброс, не объясненный регрессией.

Тогда
$$TSS = RSS + ESS$$
 или
$$Q = Q_R + Q_E .$$

 $(n-1)(k)(n-k-1)$

(в скобках указано число степеней свободы, соответствующее каждому члену уравнения).

Замечание. Любая сумма квадратов связана с числом степеней свободы, т.е. с числом независимого варьирования переменной. Существует равенство между числами степеней свободы для этого уравнения. Отнесение каждой суммы квадратов этого уравнения на одну степень свободы приводит их к сравнимому виду.

Коэффициент детерминации является одной из наиболее эффективных оценок адекватности регрессионной модели, мерой качества уравнения регрессии, характеристикой его прогностической силы.

Коэффициент детерминации (или множественный коэффициент детерминации) R^2 определяется по формуле:

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS} \quad \text{или} \quad R^2 = 1 - \frac{\sigma^2_{ocm.}}{\sigma^2_y}$$

и характеризует долю вариации зависимой переменной, обусловленной регрессией или изменчивостью объясняющих переменных.

Коэффициент R^2 показывает качество подгонки регрессионной модели к наблюдаемым значениям y_t .

Если $R^2 = 0$, то регрессия y на x_1, \dots, x_k не улучшает качество предсказания y_t по сравнению с предсказанием $\hat{y} = \bar{y}$.

Если $R^2 = 1$, то это означает точную подгонку: все $e_i = 0$, т.е. все точки наблюдений удовлетворяют уравнению регрессии.

Чем ближе R^2 к единице, тем лучше регрессия описывает зависимость между объясняющими и зависимой переменными.

Вместе с тем использование только одного коэффициента детерминации R^2 для выбора наилучшего уравнения регрессии может оказаться недостаточным. На практике встречаются случаи, когда плохо определенная модель регрессии может дать сравнительно высокий коэффициент R^2 .

Недостатками коэффициента детерминации R^2 является то, что он:

3. увеличивается при добавлении новых объясняющих переменных, хотя это и не обязательно означает улучшение качества модели

4. R^2 изменяется даже при простейшем преобразовании зависимой переменной. В этом смысле предпочтительнее использовать скорректированный (адаптированный), поправленный (adjusted) коэффициент детерминации $R^2 (R^2_{adj})$.

Для того чтобы получить адекватную оценку того, насколько хорошо вариация результирующего признака объясняется вариацией нескольких факторных признаков, применяют скорректированный коэффициент детерминации.

Скорректированным (адаптированным, поправленным) коэффициентом детерминации R^2_{adj} называется величина, определяемая по

формуле:

$$R^2_{adj} = \hat{R}^2 = 1 - \frac{n-1}{n-k-1} (1 - R^2)$$

где n – число наблюдений;

k – число факторов.

Свойства \hat{R}^2 :

1) $R^2 \geq R^2_{adj}, k > 1$ 2) $R^2_{adj} \leq 1$, но может принимать значения меньше 0.

Таким образом, использование «скорректированного» коэффициента детерминации R^2_{adj} более корректно для сравнения регрессий при изменении количества регрессоров.

Чем больше число объясняющих переменных k , тем меньше \hat{R}^2 по сравнению с R^2 . В отличие от R^2 скорректированный коэффициент \hat{R}^2 может уменьшаться при введении в модель новых объясняющих переменных, не оказывающих существенного влияния на зависимую переменную. Однако даже увеличение \hat{R}^2 при введении в модель новой объясняющей переменной не всегда означает, что ее коэффициент регрессии значим. Другими словами, увеличение \hat{R}^2 еще не означает улучшения качества регрессионной модели.

Проверка качества модели множественной регрессии также может быть осуществлена и с помощью дисперсионного анализа [14].

В модели множественной регрессии общая вариация Q - сумма квадратов отклонений зависимой переменной от средней может быть разложена на две составляющие:

$$Q = Q_R + Q_e$$

где Q_R - сумма квадратов отклонений, обусловленная регрессией;

Q_e - остаточная сумма квадратов, характеризующая влияние неучтенных факторов.

Для множественной регрессии в матричной форме:

$$Q = Y'Y - n \cdot \bar{y}^2, \quad Q_R = b'X'Y - n\bar{y}^2, \quad Q_e = X'Y - b'X'Y.$$

Приведем таблицу дисперсионного анализа ниже (табл.10):

Таблица 10 – Таблица дисперсионного анализа

Источник вариации	Сумма квадратов отклонений	Число степеней свободы	Дисперсия на одну степень свободы
x_1, x_2, \dots, x_k	$Q_R = \sum (\hat{Y} - \bar{Y})^2 = n\sigma_y^2 R^2$	k	$D_1 = \frac{Q_R}{k}$
Остаток	$Q_e = \sum (Y - \hat{Y})^2 = n\sigma_y^2 (1 - R^2)$	$n - k - 1$	$D_2 = \frac{Q_e}{n - (k + 1)}$
Общая вариация	$\sum (Y - \bar{Y})^2 = Q_R + Q_e = n\sigma_y^2$	$n - 1$	

Для определения того, действительно ли объясненный разброс RSS больше случайного ESS, используется F-тест.

F – тест - оценивания качества (значимости) уравнения регрессии – состоит в проверке гипотезы H_0 : о статистической незначимости уравнения регрессии (или незначимости показателя тесноты связи R^2).

В качестве критерия проверки гипотезы принимают случайную величину:

$$F = \frac{RSS / k}{ESS / (n - k - 1)} = \frac{R^2}{1 - R^2} \cdot \frac{n - k - 1}{k}$$

Величина F имеет распределение Фишера с $\nu_1 = k, \nu_2 = n - k - 1$ степенями свободы. Вычисляется наблюдаемое (фактическое) значение критерия $F_{факт}$ и сравнивается с критическим (табличным) значением $F_{кр}$ значений F- критерия Фишера.

Если $F_{факт} < F_{кр}$, то H_0 принимается, т.е. R^2 незначим, или уравнение регрессии незначимо;

Если $F_{факт} > F_{кр}$, то H_0 отклоняется, т.е. R^2 значим, или уравнение регрессии значимо; иначе, гипотеза H_0 об отсутствии связи между переменными x_1, x_2, \dots, x_k и y отклоняется.

Для оценки статистической значимости коэффициентов регрессии проверяется гипотеза о незначимом отличии их от нуля: $H_0 : \beta_j = 0$

В качестве критерия проверки гипотезы H_0 принимают случайную величину:

$$t_j = \frac{b_j}{S_{b_j}},$$

которая имеет t-распределение Стьюдента с $\nu = n - k - 1$ степенями свободы.

Вычисляется фактическое значение критерия $t_{\text{факт}}$. По таблице критических точек распределения Стьюдента, по заданному уровню значимости α и числу степеней свободы находят критическую (табличную) точку $t_{\text{кр}}$.

Вычисленный критерий $t_{\text{факт}}$ сравнивается с критическим (табличным) значением $t_{\text{кр}}$:

Если $|t_j| > t_{\alpha; n-k-1}$, то гипотеза H_0 отвергается, коэффициент β_j значим.

Если $|t_j| < t_{\alpha; n-k-1}$, то гипотеза H_0 принимается, коэффициент β_j незначим.

В ходе проведения множественного регрессионного анализа целесообразно проводить и следующие проверки гипотез.

Проверка гипотезы об одновременном равенстве нулю части коэффициентов регрессии: $H_0 : \beta_{k+1} = \beta_{k+2} = \dots = \beta_{k+m} = 0$

Пусть сначала была оценена регрессия sk объясняющими переменными: $\hat{y} = b_0 + b_1x_1 + \dots + b_kx_k$

и объясненная сумма квадратов составляет RSS_k .

Затем добавлено еще m переменных и по тем же данным оценено уравнение: $\hat{y} = b_0 + b_1x_1 + \dots + b_{k+m}x_{k+m}$,

при этом объясненная сумма квадратов возрастает до RSS_{k+m} .

Таким образом, объяснили дополнительно величину $(RSS_{k+m} - RSS_k)$, используя для этого m степеней свободы.

Требуется выяснить, превышает ли данное увеличение объясненной части то увеличение, которое может быть получено случайно (ESS_{k+m}).

В качестве критерия проверки этой гипотезы принимают случайную величину:

$$F = \frac{(RSS_{k+m} - RSS_k) / m}{ESS_{k+m} / (n - k - m - 1)} = \frac{R^2_{k+m} - R^2_k}{1 - R^2_{k+m}} \frac{n - k - m - 1}{m},$$

которая имеет распределение Фишера с $\nu_1 = m$, $\nu_2 = n - k - m - 1$ степенями свободы.

Вычисляется наблюдаемое (фактическое) значение критерия $F_{факт}$ (где n – объем выборки, m – число добавленных переменных в уравнение регрессии) и сравнивается с критическим (табличным) значением $F_{кр}$ значений F - критерия Фишера.

Если $F_{факт} > F_{кр}$, то H_0 отклоняется, т.е. дополнительное включение в модель m переменных оправдано.

Если $F_{факт} < F_{кр}$, то H_0 принимается, т.е. дополнительное включение в модель m переменных неоправданно.

Проверка гипотезы $H_0 : \beta' = \beta''$ (тест Чоу).

Пусть имеются две выборки объема n_1 и n_2 . Для каждой из этих выборок оценено уравнение регрессии sk объясняющими переменными:

$$(\hat{y})' = b'_0 + b'_1 x_1 + \dots + b'_k x_k$$

с необъясненной суммой квадратов $ESS_1 (\nu = n_1 - k - 1)$;

$$(\hat{y})'' = b''_0 + b''_1 x_1 + \dots + b''_k x_k$$

с необъясненной суммой квадратов $ESS_2 (\nu = n_2 - k - 1)$.

Необходимо выяснить, действительно ли две выборки однородны в регрессионном смысле? Другими словами, можно ли объединить две выборки в одну и рассматривать единую модель регрессии Y по X ?

В этом случае проверяется нулевая гипотеза $H_0 : \beta' = \beta''$, т.е. что все соответствующие коэффициенты этих уравнений равны друг другу.

Если нулевая гипотеза верна, то две регрессионные модели можно объединить в одну объемом $n = (n_1 + n_2)$, сразу для всех наблюдений с необъясненной суммой квадратов $ESS_0 (v = n_1 + n_2 - k - 1)$, причем $ESS_0 = ESS_1 + ESS_2$.

В качестве критерия проверки этой гипотезы принимают случайную величину:

$$F = \frac{ESS_0 - (ESS_1 + ESS_2) \frac{n_1 + n_2 - 2k - 2}{k + 1}}{ESS_1 + ESS_2},$$

которая имеет распределение Фишера с $v_1 = k + 1$ и $v_2 = n_1 + n_2 - 2k - 2$ степенями свободы.

Вычисляется фактическое значение критерия $F_{факт}$ и сравнивается с критическим значением $F_{кр}$ значений F - критерия Фишера.

Если $F_{факт} > F_{кр}$, то H_0 отклоняется, т.е. нельзя построить единое уравнение регрессии для обеих выборок.

Если $F_{факт} < F_{кр}$, то H_0 принимается, т.е. построить единое уравнение регрессии для обеих выборок можно.

Этот тест даст ответ на вопрос, можно ли за весь рассматриваемый период времени построить единое уравнение регрессии (рис. 5а), или же нужно разбить временной интервал на части и на каждой из них строить свое уравнение регрессии (рис. 5б).

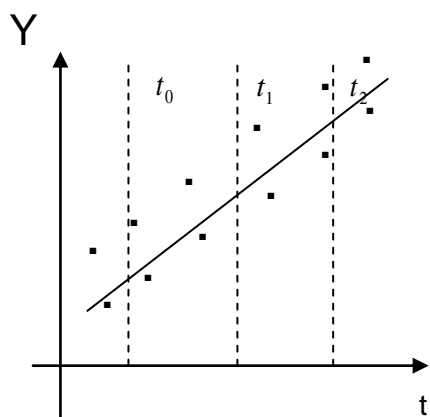


Рисунок 5а

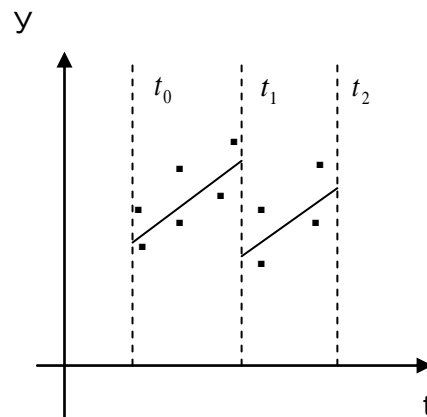


Рисунок 5б

Критерий Г. Чою может быть использован при построении регрессионных моделей при воздействии качественных признаков, когда имеется возможность разделения совокупности наблюдений по степени воздействия этого фактора на отдельные группы и требуется установить возможность использования единой модели регрессии.

Точная и адекватная модель множественной регрессии используется исследователем для прогнозирования.

Предсказанным (прогнозным) значением зависимой переменной (точечным прогнозом) является оценка \hat{y}_0 зависимой переменной y_t , которая определяется путем подстановки в выборочное уравнение множественной регрессии соответствующих (прогнозных) значений независимых переменных, задаваемых вектором $X'_0 = (1 \ x_{10} \ x_{20} \ \dots \ x_{k0})$.

Важным для оценки точности определения зависимой переменной (прогноза) является построение доверительного интервала для функции регрессии и для индивидуальных значений зависимой переменной.

Доверительный интервал для функции регрессии или для условного математического ожидания зависимой переменной $M_x(Y)$:

$$\hat{y} - t_{1-\alpha;k} \cdot S_{\hat{y}} \leq M_x(Y) \leq \hat{y} + t_{1-\alpha;k} \cdot S_{\hat{y}},$$

где \hat{y} - групповая средняя, определяемая по уравнению регрессии;

$S_{\hat{y}} = S\sqrt{X_0'(X'X)^{-1}X_0}$ - стандартная ошибка уравнения регрессии.

Доверительный интервал y_0^* для индивидуальных значений зависимой переменной имеет вид:

$$\hat{y}_0 - t_{\alpha;n-k-1} \cdot S_{\hat{y}_0} \leq y_0^* \leq \hat{y}_0 + t_{\alpha;n-k-1} \cdot S_{\hat{y}_0},$$

где \hat{y}_0 - точечный прогноз зависимой переменной;

$S_{\hat{y}_0} = S\sqrt{1 + X_0'(X'X)^{-1}X_0}$ - стандартная ошибка точечной оценки зависимой переменной.

2.3. Построение регрессионных моделей в условиях мультиколлинеарности

На практике исследователю нередко приходится сталкиваться с ситуацией, когда полученная им регрессия является «плохой», т.е. t -статистики большинства оценок малы, что свидетельствует о незначимости соответствующих регрессоров. В то же время F -статистика может быть достаточно большой, что говорит о значимости регрессии в целом. Одна из возможных причин такого явления носит название мультиколлинеарности и возникает при наличии высокой корреляции между регрессорами [4].

Мультиколлинеарность – это коррелированность двух или нескольких объясняющих переменных в уравнении регрессии.

Мультиколлинеарность является одним из основных препятствий эффективного применения множественного регрессионного анализа. Мультиколлинеарность может проявляться в функциональной (явной) и стохастической (скрытой) формах.

При функциональной форме мультиколлинеарности по крайней мере одна из парных связей между объясняющими переменными является линейной функциональной зависимостью.

Если объясняющие переменные связаны строгой функциональной зависимостью, то говорят о совершенной мультиколлинеарности.

Например, пусть уравнение регрессии имеет вид:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

Пусть также между объясняющими переменными существует строгая линейная зависимость:

$$X_2 = \gamma_0 + \gamma_1 X_1.$$

Подставим в уравнение регрессии, получим:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 (\gamma_0 + \gamma_1 X_1) + \varepsilon$$

или

$$Y = (\underbrace{\beta_0 + \beta_2 \gamma_0}_a) + (\underbrace{\beta_1 + \beta_2 \gamma_1}_b) X_1 + \varepsilon,$$

отсюда получаем уравнение парной линейной регрессии:

$$Y = a + b \cdot X_1 + \varepsilon.$$

По МНК нетрудно определить коэффициенты a и b . Тогда получим систему двух уравнений:

$$\begin{cases} \beta_0 + \beta_2 \gamma_0 = a, \\ \beta_1 + \beta_2 \gamma_1 = b. \end{cases}$$

В эту систему входят три неизвестных величины $\beta_0, \beta_1, \beta_2$. Такая система имеет бесконечно много решений.

Таким образом, совершенная мультиколлинеарность не позволяет однозначно определить коэффициенты регрессии исходного уравнения и разделить вклады объясняющих переменных X_1 и X_2 в их влиянии на зависимую переменную Y . В этом случае невозможно сделать обоснованные статистические выводы об этих коэффициентах. Следовательно, в случае совершенной мультиколлинеарности выводы по коэффициентам и по самому уравнению регрессии будут ненадежными.

Однако в экономических исследованиях мультиколлинеарность чаще проявляется в стохастической форме, когда между хотя бы двумя объясняющими переменными существует тесная корреляционная связь.

Зависимость, в которой между объясняющими переменными существует довольно сильная корреляционная зависимость называется несовершенной мультиколлинеарностью. Она характеризуется высоким коэффициентом корреляции ρ между соответствующими объясняющими переменными. Если значение $|\rho| \approx 1$, то говорят о почти совершенной мультиколлинеарности.

В любом случае мультиколлинеарность затрудняет разделение влияния объясняющих факторов на поведение зависимой переменной и делает оценки коэффициентов регрессии ненадежными. Данный вывод наглядно подтвердим с помощью диаграммы Венна (рис.б):

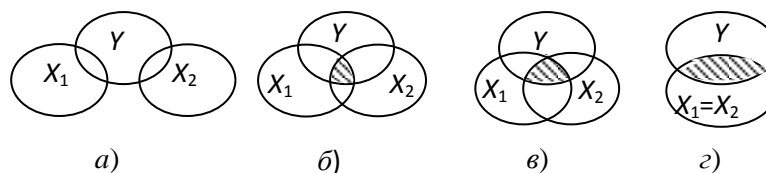


Рисунок 6 – Диаграммы Венна

На рис. ба) коррелированность между объясняющими переменными X_1 и X_2 отсутствует и влияние каждой из них на Y находит отражение в наложении кругов X_1 и X_2 на круг Y . По мере усиления линейной зависимости между X_1 и X_2 соответствующие круги все больше накладываются друг на друга. Заштрихованная область отражает совпадающие части влияния X_1 и X_2 на Y . На рис. бг) при совершенной мультиколлинеарности невозможно разграничить степени индивидуального влияния объясняющих переменных X_1 и X_2 на зависимую переменную Y .

Последствия мультиколлинеарности:

1. Стандартные ошибки коэффициентов увеличиваются, следовательно, вычисленные t-статистики занижены. Это может привести

к неоправданному выводу о существенности влияния соответствующей объясняющей переменной на зависимую переменную.

2. Оценки коэффициентов по МНК и их стандартные ошибки становятся очень чувствительными к малейшим изменениям данных, т.е. они становятся неустойчивыми. Добавление или исключение малого количества наблюдений может привести к очень сильному изменению оценок коэффициентов. При этом резко уменьшается и точность предсказания по модели.

3. Затрудняется определение вклада каждой из объясняющих переменных в объясняемую уравнением регрессии дисперсию зависимой переменной.

4. Возможно получение неверного знака у коэффициентов регрессии.

По каким признакам можно определить мультиколлинеарность.

Точных количественных критериев для определения наличия или отсутствия мультиколлинеарности не существует. Тем не менее, имеются некоторые эвристические подходы по ее выявлению. Применение этих методов позволяет выделить несколько признаков, по которым может быть установлено наличие мультиколлинеарности:

1. коэффициент детерминации R^2 и F - статистика высокие, но некоторые (или все) коэффициенты незначимы (имеют низкие t -статистики);

2. парная корреляция между малозначимыми объясняющими переменными достаточно высока (обычно больше $r_{ij} \geq 0,75; 0,8$).

3. высокие частные коэффициенты корреляции.

4. добавление или удаление наблюдений из выборки сильно изменяют значения оценок.

5. неправильные с экономической точки зрения знаки оценок регрессии;

6. неоправданно большие значения оценок коэффициентов.

Рассмотрим методы устранения мультиколлинеарности.

В зависимости от целей исследования, наличие мультиколлинеарности может и не сказываться на прогнозных качествах модели.

Если целью исследования является определение степени влияния каждой из объясняющих переменных на зависимую переменную, то наличие мультиколлинеарности, приводящее к увеличению стандартных ошибок, скорее, исказит истинные зависимости между переменными.

Единого метода устранения мультиколлинеарности, годного в любом случае, не существует. Это связано с тем, что причины и последствия мультиколлинеарности неоднозначны и во многом зависят от результатов выборки.

Для устранения или уменьшения мультиколлинеарности используется ряд методов:

1) метод исключения переменной (ых) из модели.

Из двух объясняющих переменных, имеющих высокий коэффициент корреляции (больше 0,8), одну переменную исключают из рассмотрения. При этом, какую переменную оставить, а какую удалить из анализа, решают в первую очередь на основании экономических соображений. Если с экономической точки зрения ни одной из переменных нельзя отдать предпочтение, то оставляют ту из двух переменных, которая имеет больший коэффициент корреляции с зависимой переменной.

2) метод, заключается в переходе от несмещенных оценок коэффициентов регрессии к смещенным оценкам, обладающим меньшим рассеянием относительно оцениваемого параметра.

Иногда для этого достаточно увеличить объем выборки. *Например,* при использовании ежеквартальных данных можно перейти к поквартальным данным. Увеличение количества данных сокращает

дисперсии коэффициентов регрессии и тем самым увеличивает их статистическую значимость.

3) метод, заключается в изменении спецификации модели.

Согласно этому методу либо изменяется форма модели, либо добавляются объясняющие переменные, неучтенные в первоначальной модели, но существенно влияющие на зависимую переменную.

4) метод заключается в преобразовании переменных.

В этом методе используется переход от исходных объясняющих переменных, связанных между собой достаточно тесной корреляционной зависимостью, к новым переменным, представляющим линейные комбинации исходных.

Например, пусть эмпирическое уравнение регрессии имеет вид:

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2,$$

причем X_1 и X_2 - коррелированные переменные.

В этой ситуации можно попытаться определять регрессионные зависимости относительных величин:

$$\frac{\hat{Y}}{X_1} = b_0 + b_1 \frac{X_2}{X_1},$$
$$\frac{\hat{Y}}{X_2} = b_0 + b_1 \frac{X_1}{X_2}.$$

5) и другие преобразования.

Рассмотрим метод включения - исключения переменных. Метод заключается в том, что выполняются следующие действия:

1. строится регрессионная модель методом наименьших квадратов;

2. оценивается значимость коэффициентов регрессии (по критерию Стьюдента);

3. выявляется наличие зависимости между факторными признаками путем анализа матрицы парных коэффициентов корреляции и матрицы частных коэффициентов корреляции.

4. строится новое уравнение регрессии с исключением незначимых и части взаимно коррелирующих переменных. При этом из числа коррелирующих переменных в модели оставляют те, которые более соответствуют экономическому содержанию модели, либо те которые имеют наибольшее значение частной корреляции с зависимой переменной.

5. повторяются 3-й и 4-й шаги до тех пор, пока мультиколлинеарность не будет исключена.

6. исследуется целесообразность исключения факторов из модели с помощью коэффициента детерминации и скорректированного коэффициента детерминации.

В эконометрическом исследовании важное значение, имеет отбор наиболее существенных объясняющих переменных в регрессионной модели.

Отбор факторов не самостоятельный процесс, он сопровождается построением модели. Принятие решения об исключении факторов производится на основе анализа значений специальных статистических характеристик и с учетом управляемости факторов на уровне предприятия.

Процедура отбора главных факторов обязательно включает следующие этапы:

1. Анализ факторов на мультиколлинеарность и её исключение.

Еще одним из возможных методов устранения или уменьшения мультиколлинеарности является использование пошаговых процедур отбора наиболее информативных переменных. Например, на первом шаге рассматривается лишь одна объясняющая переменная, имеющая с зависимой переменной наибольший коэффициент детерминации; на

втором шаге включается в регрессию новая объясняющая переменная, которая вместе с первоначально отобранной, образует пару объясняющих переменных, имеющую с зависимой переменной наиболее высокий (скорректированный) коэффициент детерминации; на третьем шаге вводится в регрессию еще одна объясняющая переменная, которая вместе с двумя первоначально отобранными образует тройку объясняющих переменных, имеющую с зависимой переменной наибольший (скорректированный) коэффициент детерминации и т.д.

Процедура введения новых переменных продолжается до тех пор, пока будет увеличиваться соответствующий (скорректированный) коэффициент детерминации.

Кроме рассмотренной пошаговой процедуры присоединения объясняющих переменных используется также пошаговые процедуры присоединения – удаления и процедура удаления объясняющих переменных. В большинстве случаев получаемые с помощью пошаговых процедур наборы переменных оказываются оптимальными или близкими к оптимальным.

2. Анализ тесноты взаимосвязи факторов (x) с зависимой переменной (y).

Для анализа тесноты взаимосвязи x и y используются значения коэффициента парной корреляции между фактором и функцией ($r_{x,y}$).

Факторы, для которых $r_{x,y} = 0$, т.е. не связанные с y , подлежат исключению в первую очередь. Факторы, имеющие наименьшее значение $r_{x,y}$, могут быть потенциально исключены из модели. Вопрос об их окончательном исключении решается в ходе анализа других статистических характеристик.

В модель регрессии включаются те факторы, которые более сильно связаны с зависимой переменной, но слабо связаны с другими факторами.

3. Анализ β - коэффициентов факторов, которые потенциально могут быть исключены.

Коэффициент β учитывает влияние анализируемых факторов на зависимую переменную y с учетом различия в уровне их колеблемости.

Коэффициент β показывает, на сколько σ (СКО) изменяется функция с изменением соответствующего аргумента на одну сигму при фиксированном значении остальных аргументов:

$$\beta_k = \beta_i \frac{\sigma_{x_k}}{\sigma_y}, \quad i = \overline{0, m}, \quad k = \overline{1, m},$$

где β_k - коэффициент β количественного фактора;

σ_{x_k} - СКО количественного фактора;

σ_y - СКО функции;

β_i - коэффициент регрессии при количественном факторе.

Из двух факторов x_i и x_j может быть исключен тот фактор, который имеет меньшее значение β .

Например, исключению подлежит один из мультиколлинеарных факторов x_i или x_j . Оба фактора управляемы на уровне предприятия, коэффициенты регрессии β_i и β_j статистически значимы. Фактор x_i более тесно связан с y , т.е. $r_{x_i y} > r_{x_j y}$, но при этом $\beta_{x_i} < \beta_{x_j}$. В этом случае обычно исключению подлежит фактор x_j .

4. Проверка коэффициентов регрессии на статистическую значимость (по критерию Стьюдента или по критерию Фишера).

5. Анализ факторов на управляемость.

В ходе логического анализа на основе экономических знаний исследователь должен сделать вывод: можно ли разработать организационно – технические мероприятия, направленные на улучшение (изменение) выбранных факторов на уровне предприятия. Если это воз-

можно, то данные факторы управляемы. Неуправляемые факторы на уровне предприятия могут быть исключены из модели.

6. Построение новой регрессионной модели без исключенных факторов.

Для этой модели определяется коэффициент множественной (скорректированный) детерминации \bar{R}^2 .

7. Исследование целесообразности исключения факторов из модели с помощью коэффициента детерминации.

Прежде, чем вынести решение об исключении переменных из анализа в силу их незначимости влияния на зависимую переменную, производят исследования с помощью коэффициента детерминации. Проверяется гипотеза об одновременном равенстве нулю части коэффициентов регрессии.

В первой регрессии содержится m объясняющих переменных, во второй – только часть из них, а именно m_1 объясняющих переменных. При этом $m = m_1 + m_2$, т.е. во вторую регрессию не включили m_2 объясняющих переменных.

Теперь следует проверить, вносят ли совместно эти m_2 переменных существенную долю в объяснение вариации переменной y . Для этого используется статистика:

$$F = \frac{(R_m^2 - R_{m_1}^2) \cdot (n - m - 1)}{(m - m_1) \cdot (1 - R_m^2)},$$

которая имеет F -распределение с $f_1 = m - m_1 = m_2$ и $f_2 = n - m - 1$ степенями свободы.

R_m^2 - коэффициент детерминации регрессии с m объясняющими переменными.

$R_{m_1}^2$ - коэффициент детерминации регрессии с m_1 - факторами.

Разность $(R_m^2 - R_{m_1}^2)$ является мерой дополнительного объяснения вариации переменной y за счет включения m_2 переменных.

Если $F \leq F_{f_1, f_{2\alpha}}$, то включение дополнительно объясняющих переменных совместно не оказывает значимого влияния на зависимую переменную y .

Если $F > F_{f_1, f_{2\alpha}}$, то m_2 объясняющих переменных совместно оказывают существенное влияние на вариацию переменной y , и, следовательно, в этом случае все m_2 переменные нельзя исключать из модели.

При реализации первой ситуации ($F \leq F_{f_1, f_{2\alpha}}$) факторы окончательно исключаются из модели.

2.4. Гетероскедастичность и автокорреляция остатков

При практическом проведении регрессионного анализа с помощью МНК следует обратить внимание на проблемы, связанные с выполнимостью свойств случайных отклонений моделей.

Свойства оценок коэффициентов регрессии напрямую зависят от свойств случайного члена в уравнении регрессии. Для получения качественных оценок необходимо следить за выполнимостью условий Гаусса-Маркова, т.к. при их нарушении МНК может давать оценки с плохими статистическими свойствами.

Одной из ключевых предпосылок МНК является условие постоянства дисперсий случайных отклонений: дисперсия случайных отклонений постоянна, т.е. $D(\varepsilon_i) = D(\varepsilon_j) = \sigma^2$ для любых наблюдений i и j . Выполнимость данной предпосылки называется гомоскедастичностью (постоянством дисперсии отклонений). Невыполнимость данной предпосылки называется гетероскедастичностью (непостоянством дисперсии отклонений). Иначе, условие независимости дисперсии ошибки от номера наблюдения называется

гомоскедастичностью; если это условие не выполняется то гетероскедастичностью [15,33].

Гетероскедастичность часто возникает, если анализируемые объекты, неоднородны. Например, если исследуется зависимость прибыли предприятия от размера основного фонда, то естественно ожидать, что для больших предприятий колебание прибыли будет выше, чем для малых.

На практике гетероскедастичность не так уж и редка. Зачастую есть основания считать, что вероятностные распределения случайных отклонений ϵ_t при различных наблюдениях будут различными. Это не означает, что случайные отклонения обязательно будут большими при определенных наблюдениях и малыми - при других, но это означает, что априорная вероятность этого велика. Поэтому важно помнить суть этого явления и его последствия.

Проблема гетероскедастичности характерна для перекрестных данных и довольно редко встречается при рассмотрении временных рядов. При перекрестных данных учитываются экономические субъекты, (потребители, фирмы, отрасли, страны и др.) имеющие различные доходы, размеры, потребности и т.п. В этом случае возможны проблемы, связанные с эффектом масштаба.

Во временных рядах обычно рассматриваются одни и те же показатели в различные моменты времени (ВВП, чистый экспорт, темпы инфляции и т.д. в определенном регионе за определенный период времени). Однако при увеличении (уменьшении) рассматриваемых показателей с течением времени может возникнуть проблема гетероскедастичности.

Последствия гетероскедастичности следующие.

При рассмотрении классической линейной регрессионной модели МНК дает наилучшие линейные несмещенные оценки лишь при выполнении ряда предпосылок, одной из которых является постоянство дисперсии отклонений (гомоскедастичность): $\sigma^2(\epsilon_t) = \sigma^2$ для всех наблюдений $t, t = 1, 2, \dots, n$.

При гетероскедастичности последствия применения МНК будут следующими:

1. Оценки коэффициентов по-прежнему останутся несмещенными и линейными;

2. Оценки не будут эффективными (т.е. они не будут иметь наименьшую дисперсию по сравнению с другими оценками данного параметра). Они не будут даже асимптотически эффективными. Увеличение дисперсии оценок снижает вероятность получения максимально точных оценок;

3. Дисперсии оценок будут рассчитываться со смещением. Смущенность появляется вследствие того, что не объясненная уравнением регрессии дисперсия $S^2 = \frac{\sum e_t^2}{n - m - 1}$ (m - число объясняющих переменных), которая используется при вычислении оценок дисперсий всех коэффициентов, не является более несмещенной;

4. Вследствие вышесказанного все выводы, получаемые на основе соответствующих t и F -статистик, а также интервальные оценки будут ненадежными. Следовательно, статистические выводы, получаемые при стандартных проверках качества оценок, могут быть ошибочными и приводить к неверным заключениям по построенной модели. Вполне вероятно, что стандартные ошибки коэффициентов будут занижены, а следовательно, t -статистики будут завышены. Это может привести к признанию статистически значимыми коэффициентов, таковыми на самом деле не являющимися.

Рассмотрим методы смягчения (коррекции) проблемы гетероскедастичности.

Гетероскедастичность приводит к неэффективности оценок, несмотря на их несмещенность. Это может обусловить необоснованные выводы по качеству модели. Поэтому при установлении гетероскедастичности возникает необходимость преобразования модели с целью устранения данного недостатка.

Вид преобразования зависит от того, известны или нет дисперсии σ_t^2 отклонений $\varepsilon_t, t = 1, 2, \dots, n$.

1. Метод взвешенных наименьших квадратов - (ВНК).

Этот метод применяется при известных для каждого наблюдения значениях σ_t^2 . В этом случае можно устранить гетероскедастичность, разделив каждое наблюдаемое значение на соответствующее ему значение дисперсии σ_t^2

Опишем метод ВНК на примере парной регрессии: $y_t = \beta_0 + \beta_1 x_t + E_t$ (1)

Разделим обе части равенства на известное $\sigma_t = \sqrt{\sigma_t^2}$:

$$\frac{y_t}{\sigma_t} = \frac{1}{\sigma_t} \beta_0 + \frac{x_t}{\sigma_t} \beta_1 + \frac{\varepsilon_t}{\sigma_t}$$

Обозначим:

$$\begin{aligned} \frac{y_t}{\sigma_t} &= y_t^*; & \frac{\varepsilon_t}{\sigma_t} &= V_t \\ \frac{x_t}{\sigma_t} &= x_t^*; & \frac{1}{\sigma_t} &= Z_t \end{aligned}$$

Получим уравнение регрессии без свободного члена, но с дополнительной объясняющей переменной Z и с «преобразованным» отклонением V : $y_t^* = \beta_0 Z_t + \beta_1 x_t^* + V_t$ (2)

При этом для V_t выполняются условие гомоскедастичности.

$$\text{Действительно, } \sigma_t^2(V_t) = M(V_t - M(V_t))^2 = M(V_t^2) + M^2(V_t)$$

Так как по предпосылке 1 МНК:

$$M(E_t) = 0, \text{ то } M(V_t) = M\left(\frac{\varepsilon_t}{\sigma_t}\right) = \frac{1}{\sigma_t} M(\varepsilon_t) = 0 \text{ и тогда:}$$

$$\sigma_t^2(V_t) = M(V_t^2) = M\left(\frac{\varepsilon_t^2}{\sigma_t^2}\right) = \frac{1}{\sigma_t^2} M(\varepsilon_t^2) = \frac{1}{\sigma_t^2} M(\varepsilon_t - M(\varepsilon_t))^2 = \frac{1}{\sigma_t^2} \sigma_t^2 = 1 - const$$

Следовательно, для преобразованной модели (2) выполняются предпосылки 1-3 МНК. В этом случае оценки полученные по МНК, будут наилучшими линейными несмещенными оценками.

Таким образом ВНК включает следующие этапы:

1. Значения каждой пары наблюдений (x_t, y_t) делят на известную величину σ_t . Тем самым наблюдениям с наименьшими дисперсиями придаются наибольшие «веса», а с максимальными дисперсиями - наименьшие «веса». Действительно, наблюдения с наименьшими дисперсиями отклонений будут более значимыми при оценке коэффициентов регрессии, чем наблюдения с большими дисперсиями. Учет этого факта увеличивает вероятность получения более точных оценок.

2. По МНК для преобразованных значений $\left(\frac{1}{\sigma_t}, \frac{x_t}{\sigma_t}, \frac{y_t}{\sigma_t}\right)$ строится уравнение регрессии без свободного члена с гарантированными качествами оценок.

1) Дисперсии отклонений неизвестны.

Для применения ВНК необходимо знать фактические значения дисперсий σ_t^2 отклонений. На практике такие значения известны крайне редко. Следовательно, чтобы применить ВНК, необходимо сделать реалистические предположения о значениях σ_t^2 .

Например, может оказаться целесообразным предположить, что дисперсии σ_t^2 отклонений ε_t пропорциональны значениям x_t или значениям x_t^2 .

а) Дисперсии σ_t^2 пропорциональны x_t , т.е. $\sigma_t^2 = \sigma_t^2 x_t$ (σ_t^2 - коэффициент пропорциональности). Тогда уравнение (1) преобразуется делением его левой и правой частей на $\sqrt{x_t}$:

$$\frac{y_t}{\sqrt{x_t}} = \beta_0 \frac{1}{\sqrt{x_t}} + \beta_1 \frac{x_t}{\sqrt{x_t}} + \frac{e_t}{\sqrt{x_t}} \Rightarrow \frac{y_t}{\sqrt{x_t}} = \beta_0 \frac{1}{\sqrt{x_t}} + \beta_1 \sqrt{x_t} + V_t \quad (3)$$

Можно показать, что для случайных отклонений $V_t = \frac{\varepsilon_t}{\sqrt{x_t}}$ выполняется условие гомоскедастичности. Следовательно, для регрессии (3) применим

обычный МНК. Действительно, в силу выполнимости предпосылки

$$\sigma_t^2 = \sigma^2(\varepsilon_t) = \sigma^2 x_t$$

$$V_t = \frac{\varepsilon_t}{x_t} \quad \text{имеем:} \quad \sigma_t^2(V_t) = \sigma^2 \left(\frac{\varepsilon_t}{\sqrt{x_t}} \right) = \frac{1}{x_t^*} \sigma^2(\varepsilon_t) = \frac{1}{x_t^*} \sigma^2 x_t = \sigma^2 = const$$

α

Таким образом, оценив для (3) по МНК- коэффициенты β_0 и β_1 , затем возвращаются к исходному уравнению регрессии (1).

б) Дисперсии σ_t^2 пропорциональны x_t^2 .

Если зависимость σ_t^2 от x_t выражается не линейной функцией а квадратичной то соответствующим преобразованием будет деление уравнения регрессии $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$ на x_t :

$$\frac{y_t}{x_t} = \beta_0 \frac{1}{x_t} + \beta_1 + \frac{\varepsilon_t}{x_t} \quad \Rightarrow \quad \frac{y_t}{x_t} = \beta_0 \frac{1}{x_t} + \beta_1 + V_t \quad (4), \quad \text{где } V_t = \frac{\varepsilon_t}{x_t}$$

Несложно показать, что для отклонений V_t будет выполняться условие гомоскедастичности. После определения по МНК - оценок коэффициентов β_0 и β_1 для уравнения (4) возвращаются к исходному уравнению $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$.

Отметим, что для применения описанных выше преобразований весьма значимы знания об исходных значениях дисперсий отклонений σ_t^2 , либо предположения, какими эти дисперсии могут быть. Во многих случаях дисперсии отклонений зависят не от включенных в уравнение регрессии объясняющих переменных, а от тех, которые не включены в модель, но играют существенную роль в исследуемой зависимости. В этом случаи они должны быть включены в модель. В ряде случаев для устранения гетероскедастичности необходимо изменить спецификацию модели (например, линейную на лог- линейную, мультипликативную на аддитивную и т.п.).

Наличие гетероскедастичности не позволяет получить эффективные оценки, что приводит к необоснованным выводам по их качеству.

Обнаружение гетероскедастичности является сложной задачей. Зная характер данных, появления проблемы гетероскедастичности можно предвидеть и попытаться устранить этот недостаток еще на этапе спецификации. Однако значительно чаще эту проблему приходится решать после построения уравнения регрессии.

Не существует однозначного метода обнаружения гетероскедастичности. Для определения гетероскедастичности разработано довольно большое число тестов и критериев для них:

1. Графический анализ остатков (отклонений).

В этом случае по оси абсцисс откладываются значения x_t обьясняющей переменной X , а по оси ординат либо отклонение e_t , либо их квадраты e_t^2 . Примеры таких графиков приведем на рис. 7:

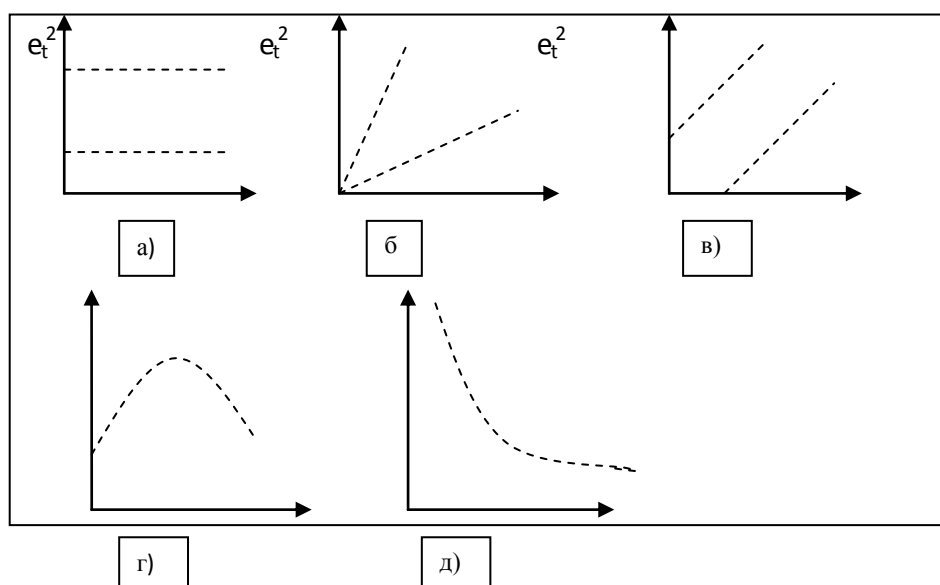


Рисунок 7 – Графический анализ остатков

На рис. а) все отклонения e_t^2 находятся внутри полуполосы постоянной ширины, параллельной оси абсцисс. Это говорит о независимости дисперсий e_t^2 от значений переменной X и их постоянстве, т.е. в этом случае выполняются условия гомоскедастичности. На рис. б-д) наблюдаются некоторые систематические изменения в соотношениях между значениями x_t , переменной

X и квадратами отклонений e_t^2 . Ситуации б-д) отражают большую вероятность наличия гетероскедастичности для рассматриваемых статистических данных.

Графический анализ отклонений является удобным и достаточно надежным в случае парной регрессии. При множественной регрессии графический анализ возможен для каждой из объясняющих переменных X_j ($j = 1, 2, \dots, m$) отдельно. При большом числе объясняющих переменных, вместо объясняющих переменных X_j ($j = 1, 2, \dots, m$) по оси абсцисс откладывают значения \hat{y}_t ($t = 1, 2, \dots, n$), получаемые из эмпирического уравнения регрессии; \hat{y}_t является линейной комбинацией X_{ij} ; график, отражающий зависимость e_t^2 и \hat{y}_t может указать на наличие гетероскедастичности аналогично ситуациям на рис. б-д).

2. Тест Уайта (White) - универсальный.

Если в модели присутствует гетероскедастичность, то очень часто это связано с тем, что дисперсии ошибок некоторым образом зависят от регрессоров, а гетероскедастичность должна как-то отражаться в остатках обычной регрессии исходной модели. Исходя из этого, Уайт предложил метод тестирования гипотезы $H_0 = \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2$, без каких-либо предположений относительно структуры гетероскедастичности. Сначала к исходной модели $y = X\beta + \varepsilon$ применяется обычный МНК и находятся остатки регрессии e_t ($t = 1, 2, \dots, n$). Затем осуществляется регрессия квадратов этих остатков e_t^2 на все регрессоры X, их квадраты, но парные произведения и константу, если ее не было в составе исходных регрессоров. Тогда при гипотезе H_0 величина nR^2 асимптотически имеет распределение $\chi^2(N - 1)$, где R^2 - коэффициент детерминации, а N - число регрессоров второй регрессии. Однако, если гипотеза H_0 отвергается, этот тест не дает указания на функциональную форму гетероскедастичности.

3. Тест Голдфелда - Куандта (Goldfeld - Quandt).

Этот тест применяется, когда есть предположение о прямой зависимости дисперсии ошибок от величины некоторой независимой переменной (или стандартное отклонение $\sigma_t^2 = \sigma(\varepsilon_t)$ пропорционально значению x_t переменной X в этом наблюдении, т.е. $\sigma_t^2 = \sigma^2 x_t^2$, $t = 1, 2, \dots, n$. Предполагается, что ε_t имеет нормальное распределение и отсутствует автокорреляция остатков).

Тест состоит в следующем:

1. Все n наблюдений упорядочиваются по величине X , или упорядочить данные по убыванию той независимой переменной, относительно которой есть подозрение на гетероскедастичность;

2. Вся упорядоченная выборка после этого разбивается на три подвыборки размерностей k , $(n - 2k)$, k соответственно.

3. Оцениваются отдельные регрессии для первой подвыборки (k первых наблюдений) и для третьей подвыборки (k последних наблюдений). Если предположение о пропорциональности дисперсий отклонений значениям X верно, то дисперсия регрессии по первой подвыборке (сумма квадратов отклонений $S_1 = \sum_{t=1}^k e_t^2$) будет существенно меньше дисперсии регрессии по

третьей выборке (суммы квадратов отклонений $S_{31} = \sum_{t=n-k+1}^n e_t^2$).

4. Для сравнения соответствующих дисперсий строится следующая F – статистика:

$$F = \frac{S_3 / (k - m - 1)}{S_1 / (k - m - 1)} = \frac{S_3}{S_1}$$

Здесь $(k - m - 1)$ – число степеней свободы соответствующих выборочных дисперсий (m – количество объясняющих переменных в уравнении регрессии).

При сделанных предположениях относительно случайных отклонений построенная F-статистика имеет распределение Фишера с числом степеней свободы $\nu_1 = \nu_2 = k - m - 1$

5. Если $F_{набл} = \frac{S_3}{S_1} > F_{кр} = F_{\alpha; \nu_1; \nu_2}$, то гипотеза об отсутствии гетероскедастичности отклоняется (α - выбранный уровень значимости).

Какими должны быть размеры подвыборок для принятия обоснованных решений?

Для парной регрессии Голдфелд и Куандт предлагают следующие пропорции: $n = 30, k = 11; n = 60, k = 22$.

Для множественной регрессии этот тест проводится для той объясняющей переменной, которая в наибольшей степени связана с σ_t . При этом k должно быть больше, чем $(m+1)$.

Этот тест может быть использован при предположении об обратной пропорциональности между σ_t и значениями объясняющей переменной. При этом статистика Фишера примет вид: $F = S_1/S_3$.

4. Тест Бреуша-Пагана (Breusch-Pagan).

Он применяется в тех случаях, когда априорно предполагается, что дисперсии σ_t^2 зависят от некоторых переменных: $\sigma_t^2 = \gamma_0 + Z_t' \gamma$, $t = \overline{1, n}$, где

$Z_t = (Z_{t1}, \dots, Z_{tp})'$ - вектор (наблюдаемых) независимых переменных;

$\gamma_0, \gamma = (\gamma_1, \dots, \gamma_p)'$ - неизвестные параметры.

Тест состоит в следующем:

1. Провести обычную регрессию $y = X\beta + \varepsilon$ и получить вектор остатков $e = (e_1, \dots, e_n)'$;

2. Построить оценку $\hat{\sigma}^2 = \frac{1}{n} \sum e_t^2$;

3. Провести регрессию $\frac{e_t^2}{\hat{\sigma}^2} = \gamma_0 + Z_t' \gamma + V_t$ и найти для нее объясненную часть вариации RSS;

4. Построить статистику $RSS/2$. Если верна гипотеза $H_0 = \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2$ (отсутствие гетероскедастичности), то величина $RSS/2$ асимптотически имеет распределение $\chi^2(p)$.

При выявлении гетероскедастичности с помощью этого теста можно попытаться осуществить коррекцию (преобразования, стабилизирующее дисперсию) с помощью метода ВНК, выбирая в качестве весов величины $(\gamma_0 + Z'_t \hat{\gamma})^{-\frac{1}{2}}$, где $\hat{\gamma}_0, \hat{\gamma}$ - оценки полученные в п. 3.

При этом может оказаться, что $\gamma_0 - Z'_t \hat{\gamma} < 0$ для некоторых t . Если число таких наблюдений невелико, то их можно просто выбросить. В противном случае, можно попытаться использовать мультипликативную форму гетероскедастичности: $\sigma_t^2 = e^{\gamma_0 + Z'_t \gamma}; t = \overline{1, n}$.

На практике имеет смысл применить несколько методов определения гетероскедастичности и способов ее корректировки

5. Тест ранговой корреляции Спирмина

Выдвигается нулевая гипотеза об отсутствии гетероскедастичности случайного члена. При выполнении теста ранговой корреляции Спирмина предполагается, что дисперсия случайного члена будет либо увеличиваться, либо уменьшаться по мере увеличения x , и поэтому в регрессии, оцениваемой с помощью МНК, абсолютные величины остатков $|e|$ и значения x будут коррелированы.

Данные по x и остатков $|e|$ ранжируются по переменной x , и определяются их ранги.

Ранг – это порядковый номер значений переменной в ранжированном ряду.

Коэффициент ранговой корреляции Спирмена определяется по формуле $r = 1 - \frac{6 \sum D_i^2}{n(n^2 - 1)}$, где D_i – разность между рангами x и $|e|$.

Если предположить, что коэффициент корреляции для генеральной совокупности равен нулю, то коэффициент ранговой корреляции имеет нормальное распределение с нулевым математическим ожиданием и дисперсией $1/(n-1)$ в больших выборках. Соответствующая тестовая статистика $r\sqrt{n-1}$ сравнивается с критическим значением $t_{кр}$ при заданном уровне значимости ($t_{кр}=1,96$ при $\alpha=5\%$, $t_{кр}=2,58$ при $\alpha=1\%$).

Если $r\sqrt{n-1} > t_{кр}$, то нулевая гипотеза об отсутствии гетероскедастичности будет отклонена.

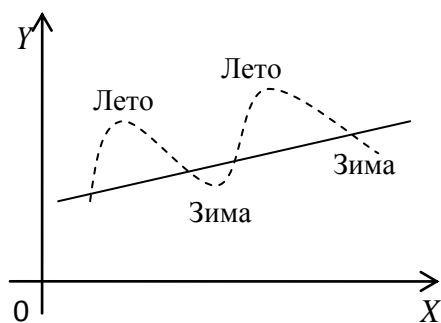
Если в модели регрессии имеется более одной объясняющей переменной, то проверка гипотезы может выполняться с использованием любой из них.

Важной предпосылкой построения качественной регрессионной модели по МНК является независимость значений случайных отклонений ε_t от значений отклонений во всех других наблюдениях. Отсутствие зависимости гарантирует отсутствие коррелированности между любыми отклонениями ($\sigma(\varepsilon_t, \varepsilon_j) = Cov(\varepsilon_t, \varepsilon_j) = 0$ при $t \neq j$) и, в частности, между соседними отклонениями ($\sigma(\varepsilon_{t-1}, \varepsilon_t) = 0, i = \overline{1, n}$).

Автокорреляция (последовательная корреляция) определяется как корреляция между наблюдаемыми показателями, упорядоченными во времени (временные ряды) или в пространстве (перекрестные данные).

Автокорреляция остатков (отклонений) обычно встречается в регрессивном анализе при использовании данных временных рядов. При использовании перекрестных данных, наличие автокорреляции (пространственной корреляции) крайне редко.

В экономических задачах значительно чаще встречается положительная автокорреляция ($\sigma(\varepsilon_{t-1}, \varepsilon_t) > 0$), чем отрицательная автокорреляция ($\sigma(\varepsilon_{t-1}, \varepsilon_t) < 0$).



В большинстве случаев положительная автокорреляция вызывается направленным постоянным воздействием некоторых не учтенных в модели факторов. Суть автокорреляции поясним следующим примером: Пусть

исследуется спрос Y на мороженое в зависимости от дохода X по ежемесячным данным. Трендовая зависимость, отражающая увеличение спроса с ростом дохода, может быть представлена линейной функцией $Y = \beta_0 + \beta_1 X$, изображенной на рис.

Фактические точки наблюдений обычно будут превышать трендовую линию в летние периоды и будут ниже ее в зимние.

Отрицательная автокорреляция фактически означает, что за положительным отклонением следует отрицательное и наоборот.

Среди основных причин, вызывающих появление автокорреляции, можно выделить ошибки спецификации, инерцию в изменении экономических показателей, эффект паутины, сглаживание данных.

Рассмотрим причины, вызывающие появление автокорреляции.

1. Ошибки спецификации. Не учет в модели какой-либо важной объясняющей переменной, либо неправильный выбор формы зависимости обычно приводит к системным отклонениям точек наблюдений от линии регрессии, что может обусловить автокорреляцию.

2. Инерция. Многие экономические показатели (инфляция, безработица, ВВП и т.п.) обладают определенной цикличностью, связанной с волнообразностью деловой активности. Действительно, экономический подъем приводит к росту занятости, сокращению инфляции увеличению ВВП и так далее. Этот рост продолжается до тех пор, пока изменения конъюнктуры рынка и ряда экономических характеристик не приведет к замедлению роста, затем остановке и

движению вспять рассматриваемых показателей. В любом случае эта трансформация происходит не мгновенно, а обладает определенной инертностью.

3. Эффект паутины. Во многих производственных и других сферах экономические показатели реагируют на изменение экономических условий с запаздыванием (временным лагом).

Пример: предложение с/х продукции реагирует на изменение цены с запаздыванием (равным периоду созревания урожая). Большая цена с/х продукции в прошедшем году вызовет ее перепроизводство в текущем году, а, следовательно, на нее снизится и так далее.

4. Сглаживание данных. Зачастую данные по некоторому продолжительному временному периоду получают усреднением данных по составляющим его подынтервалам. Это может привести к определенному сглаживанию колебаний, которые имелись внутри рассматриваемого периода, что в свою очередь может послужить причиной автокорреляции.

Рассмотрим последствия автокорреляции:

1. Оценки параметров, оставаясь линейными и несмещенными, перестают быть эффективными. Следовательно, они перестают обладать свойствами наилучших линейных несмещенных оценок (BLUE - оценок).

2. Дисперсии оценок являются смещенными. Часто дисперсии, вычисляемые по стандартным формулам, являются заниженными, что влечет за собой увеличение t -статистик. Это может привести к признанию статистически значимыми объясняющие переменные, которые в действительности таковыми могут и не являться.

3. Оценка дисперсии регрессии $S^2 = \sum_{t=1}^T \frac{e_t^2}{T-m-1}$ (T - объем выборки) является смещенной оценкой истинного значения σ^2 , во многих случаях занижая его.

4. В силу вышесказанного, выводы по t - и F -статистикам,

определяющим значимость коэффициентов регрессии и коэффициента детерминации, возможно, будут неверными. Вследствие этого ухудшаются прогнозные качества модели.

Рассмотрим методы определения автокорреляции:

1. Графический метод

На рис. 8 приведены последовательно - временные графики, увязывающие отклонение e_t с моментами t их получения.

По оси абсцисс обычно откладываются либо время получения статистических данных, либо порядковый номер наблюдения, а по оси ординат - отклонения ε_t , (либо оценки отклонений e_t).

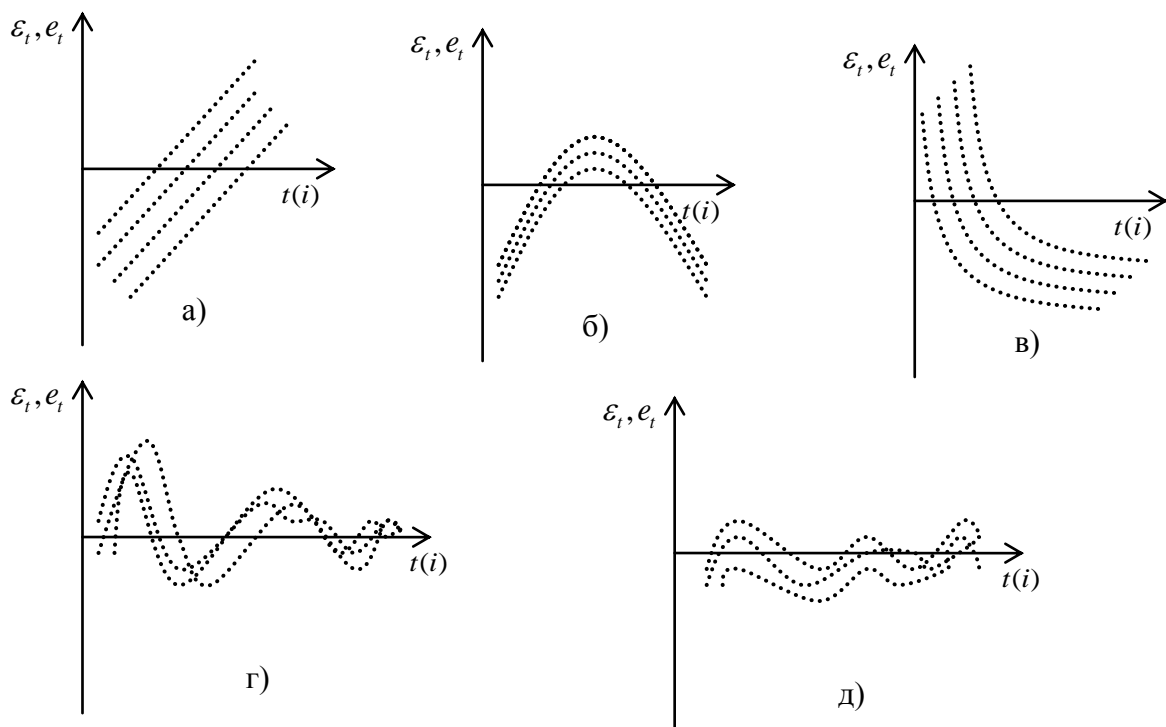


Рисунок 8 – Графический метод определения автокорреляции

На рис. а - г) имеются определенные связи между отклонениями, то есть автокорреляция имеет место. Отсутствие зависимости на рис. д) свидетельствует об отсутствии автокорреляции.

В современных компьютерных прикладных программах для решения задач по эконометрике аналитическое выражение регрессии дополняется

графическим представлением результатов. На график реальных колебаний зависимой переменной накладывается график колебаний переменной по уравнению регрессии. Сопоставив эти два графика, можно выдвинуть гипотезу о наличии автокорреляции остатков. Если эти графики пересекаются редко, то можно предположить наличие положительной автокорреляции остатков.

2. Критерий Дарбина-Уотсона (Durbin - Watson).

Большинство тестов на наличие корреляции по времени в ошибках системы $y = X\beta + \varepsilon$ используют идею: если корреляция есть у ошибок ε , то она присутствует и в остатках e , получаемых после применения к $y = X\beta + \varepsilon$ обычного МНК.

Пусть нулевая гипотеза состоит в отсутствии корреляции, то есть $H_0: \rho = 0$ (ρ - коэффициент корреляции между соседними ошибками).

В качестве альтернативной может выступать либо просто $H_1: \text{“не } H_0 \text{”}$, либо односторонняя гипотеза: $H_1: \rho > 0$.

Наиболее широко используется тест Дарбина-Уотсона. Суть его состоит в том, что на основе вычисленной статистики DW (важнейшая характеристика качества регрессионной модели) делается вывод об автокорреляции.

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}.$$

Эта статистика тесно связана с выборочным коэффициентом корреляции $r_{e_t e_{t-1}}$:

$$DW \approx 2(1 - r_{e_t e_{t-1}}),$$

$$\text{где } r \approx \frac{\sum_{t=2}^n e_t e_{t-1}}{\sqrt{\sum_{t=2}^n e_t^2 \sum_{t=1}^{n-1} e_t^2}} \approx \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2}, \quad 0 \leq DW \leq 4.$$

Содержательный смысл статистики DW : если между e_t и e_{t-1} имеется достаточно высокая положительная корреляция, то e_t и e_{t-1} близки друг к другу и величина статистики DW мала. Действительно, если $r_{e_t e_{t-1}} \approx 1$ (положительная автокорреляция), то $DW = 0$. Отсутствие корреляции означает, что $DW \approx 2$. Если $r_{e_t e_{t-1}} \approx -1$ (отрицательная автокорреляция), то $DW \approx 4$.

Для более точного определения, какое значение DW свидетельствует об отсутствии автокорреляции, а какое - об ее наличии, была построена таблица критических точек распределения Дарбина-Уотсона. По ней для заданного уровня значимости α , числа наблюдений n и количества объясняющих переменных m определяются два значения: d_l - нижняя граница (*l - low - нижняя*), d_u - верхняя граница (*u - upper*) (которые зависят лишь от α, n, m , а, следовательно, могут быть затабулированы) и обладают следующим свойством: если $DW > d_u$, то $DW > d^*$ и, значит, гипотеза H_0 не отвергается; если $DW < d_l$, то $DW < d^*$, и гипотеза H_0 отвергается в пользу H_1 ; если $d_l < DW < d_u$, то ситуация неопределенна, то есть нельзя высказаться в пользу той или иной гипотезы. Если альтернативной является гипотеза об отрицательной корреляции $H_1 : \rho < 0$, то соответствующими верхними и нижними границами будут $4 - d_l$ и $4 - d_u$.

Разработаны специальные таблицы критических точек статистики Дарбина-Уотсона, позволяющие при данном числе наблюдений n , количестве объясняющих переменных m и заданном уровне значимости α определять границы приемлемости (критические точки) наблюдаемой статистики DW .

Для проверки гипотезы об отсутствии автокорреляции остатков используется числовой отрезок, изображенный ниже:



Общая схема критерия Дарбина-Уотсона следующая:

1. По построенному эмпирическому уравнению регрессии:

$$\hat{y}_t = b_0 + b_1 x_t + \dots + b_m x_{tm}$$

определяются значения отклонений $e_t = y_t - \hat{y}_t$ для каждого наблюдения t , $t = 1, 2, \dots, n$.

2. По формуле рассчитывается статистика DW .

3. По таблице критических точек Дарбина-Уотсона определяются два числа d_l и d_u и осуществляются выводы:

Значение статистики DW	Вывод
$4 - d_l < DW < 4$	Гипотеза H_0 отвергается, есть отрицательная корреляция.
$4 - d_u < DW < 4 - d_l$	Неопределенность.
$d_u < DW < 4 - d_u$	Гипотеза H_0 не отвергается.
$d_l < DW < d_u$	Неопределенность.
$0 < DW < d_l$	Гипотеза H_0 отвергается, есть положительная корреляция.

При использовании критерия Дарбина-Уотсона необходимо учитывать следующие ограничения:

1. Критерий DW применяется лишь для тех моделей, которые содержат свободный член.

2. Предполагается, что случайные отклонения ε_t определяются по итерационной схеме: $\varepsilon_t = \rho \varepsilon_{t-1} + v_t$, называемой авторегрессионной схемой первого порядка $AR(1)$; v_t — случайный член.

3. Статистические данные должны иметь одинаковую

периодичность (то есть не должно быть пропусков в наблюдениях).

4. Критерий Дарбина-Уотсона не применим для регрессионных моделей, содержащих в составе объясняющих переменных зависимую переменную с временным лагом в один период, то есть для так называемых авторегрессионных моделей вида:

$$y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_m x_{tm} + \gamma y_{t-1} + \varepsilon_t.$$

Рассмотрим методы устранения автокорреляции.

Основной причиной наличия случайного члена в модели являются несовершенные знания о причинах и взаимосвязях, определяющих то или иное значение зависимой переменной. Поэтому свойства случайных отклонений, в том числе и автокорреляция, в первую очередь зависят от выбора формулы зависимости и состава объясняющих переменных. Так как автокорреляция чаще всего вызывается неправильной спецификацией модели, то необходимо, прежде всего, скорректировать саму модель. Возможно, автокорреляция вызвана отсутствием в модели некоторой важной объясняющей переменной. Также можно попробовать изменить формулу зависимости (линейную на логарифмическо-линейную).

Однако если все процедуры изменения спецификации исчерпаны, а автокорреляция имеет место, то можно предположить, что она обусловлена какими-то внутренними свойствами ряда $\{e_t\}$. В этом случае можно воспользоваться авторегрессионным преобразованием, в котором ошибки образуют так называемый авторегрессионный процесс (схему) первого порядка $AR(1)$.

Для изложения $AR(1)$ рассмотрим модель: $y = X\beta + \varepsilon$, где t -я компонента вектора y представляет значение зависимой переменной в момент времени t , $t = \overline{1, n}$. Будем считать, для определенности, что первым регрессором в X является константа.

Запишем уравнение для наблюдения в момент времени t :

$$y_t = \beta_1 + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t = x'_t \beta + \varepsilon_t,$$

где $x'_t = (1, x_{t2}, \dots, x_{tk})$ - t -я строка матрицы X .

Один из наиболее простых способов учета коррелированности ошибок (в разные моменты времени) состоит в предположении, что случайная последовательность $\{\varepsilon_t, t = \overline{1, n}\}$ образует авторегрессионный процесс первого порядка. Это означает, что все ошибки удовлетворяют рекуррентному соотношению $\varepsilon_t = \rho \varepsilon_{t-1} + v_t$, где $\{v_t, t = \overline{1, n}\}$ - последовательность независимых нормально распределенных случайных величин с нулевым средним и постоянной дисперсией σ_v^2 , а ρ - некоторый параметр, называемый коэффициентом авторегрессии ($|\rho| < 1$).

Однако на практике, когда значение коэффициента ρ известно, встречается крайне редко; обычно параметр авторегрессии ρ неизвестен и его необходимо оценивать. Существует несколько методов оценивания, как правило, они имеют итеративный характер.

Приведем наиболее употребляемые.

1. Оценивание ρ на основе статистики Дарбина-Уотсона (Д - У).

Статистика Д-У тесно связана с коэффициентом корреляции между соседними отклонениями через соотношение: $DW \approx 2(1 - r_{e_t e_{t-1}})$.

Тогда в качестве оценки коэффициента ρ может быть взят коэффициент $r = r_{e_t e_{t-1}}$, то есть $r \approx 1 - \frac{DW}{2}$.

Этот метод неплох при большом числе наблюдений; оценка r параметра ρ будет достаточно точной.

2. Метод (процедура) Кохрейна-Оркатта (Cochrane-Orcutt)

Опишем итерационный процесс этого метода на примере парной регрессии: $Y = \beta_0 + \beta_1 X + \varepsilon$ и авторегрессионной схемы первого порядка

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t.$$

1) Оценивается по МНК данная регрессия, и для нее определяются оценки e_t отклонений ε_t , $t = \overline{1, n}$.

2) С использованием схем $AR(1)$ оценивается регрессионная зависимость $e_t = \hat{\rho}e_{t-1} + v_t$, где $\hat{\rho}$ - оценка коэффициента ρ .

3) На основе данной оценки строится уравнение:

$$(y_t - \hat{\rho}y_{t-1}) = \alpha(1 - \hat{\rho}) + \beta(x_t - \hat{\rho}x_{t-1}) + (\varepsilon_t - \hat{\rho}\varepsilon_{t-1})$$

с помощью которого оцениваются коэффициенты α и β (в этом случае значение $\hat{\rho}$ известно).

4) Значения $\beta_0 = \alpha(1 - \hat{\rho})$ и $\beta_1 = \beta$ подставляются в $Y = \beta_0 + \beta_1X + \varepsilon$.

Вновь вычисляются оценки e_t отклонений и процесс возвращается к этапу 2.

Процесс обычно заканчивается, когда очередное приближение ρ мало отличается от предыдущего. Иногда просто фиксируется количество итераций. Процедура Кохрейна-Оркатта реализована в большинстве эконометрических компьютерных программ.

6. Процедура Хилдрета-Лу (Hildreth-Lu)

По данному методу регрессия $y_t - \rho y_{t-1} = \beta_0(1 - \rho) + \beta_1(x_t - \rho x_{t-1}) + (\varepsilon_t - \rho \varepsilon_{t-1})$ оценивается для каждого возможного значения ρ из отрезка $[-1; 1]$ с любым шагом (0,1; 0,0001; 0,05 и так далее). Величина $\hat{\rho}$, дающая наименьшую стандартную ошибку регрессии (сумма квадратов отклонений минимальна), принимается в качестве оценки коэффициента ρ . И значения β_0^* и β_1 оцениваются из этого же уравнения регрессии именно с данным значением $\hat{\rho}$. Затем в некоторой окрестности этого значения $\hat{\rho}$ устраивается более мелкая сетка и процесс повторяется. Итерация заканчивается, когда будет достигнута желаемая точность.

Этот итерационный метод широко используется в пакетах прикладных программ.

Таким образом, при установлении автокорреляции необходимо в первую очередь проанализировать правильность спецификации модели. Если после уточнения состава объясняющих переменных, либо изменения формы зависимости автокорреляции по-прежнему имеет место, то, возможно, это связано с внутренними свойствами ряда отклонений $\{\varepsilon_t\}$. В этом случае возможны определенные преобразования, устраняющие автокорреляцию. Среди них выделяется $AR(1)$, которая может быть обобщена в $AR(k)$, $k = 2, 3, \dots$.

Для применения схем необходимо оценить коэффициент корреляции между отношениями. Это может быть сделано различными методами: на основе статистики Дарбина-Уотсона, Кохрейна-Оркатта, Хилдрета-Лу и другие. В случае наличия среди объясняющих переменных лаговой зависимости переменной наличие автокорреляции устанавливается с помощью h статистики Дарбина. А для ее устранения в этом случае предпочтителен метод Хилдрета-Лу.

2.5. Линейные регрессионные модели с переменной структурой

В регрессионных моделях в качестве объясняющих переменных часто приходится использовать не только количественные (определяемые численно), но и качественные переменные.

Например, спрос на некоторое благо может определяться ценой, доходом потребителей, наличие у работника высшего образования и т.д. (эти показатели определяются количественно). Но спрос может также зависеть от вкусов потребителей, их ожиданий, национальных и религиозных особенностей и т.д. [21]. Эти показатели представить в численном виде нельзя.

Возникает проблема отражения в модели влияния таких переменных на исследуемую величину. Это сложная задача.

Качественные признаки могут существенно влиять на структуру линейных связей между переменными и приводить к скачкообразному изменению параметров регрессионной модели. В этом случае говорят об исследовании регрессионных моделей с переменной структурой или построении регрессионных моделей по неоднородным данным.

Обычно в моделях влияние качественного фактора выражается в виде фиктивной (искусственной) переменной, которая отражает два противоположных состояния качественного фактора (манекенов). Например, «фактор действует» - «фактор не действует», «курс валюты фиксированный» - «курс валюты плавающий», «сезон летний» - «сезон зимний», пол «мужской - женский» и т.д. В этом случае фиктивная переменная может выражаться в двоичной форме:

$$D = \begin{cases} 0, & \text{фактор не действует,} \\ 1, & \text{фактор действует.} \end{cases}$$

Например, $D=0$, если потребитель не имеет высшего образования, $D=1$, если – имеет.

Переменная D называется фиктивной (искусственной, двоичной) переменной (индикатором).

Таким образом, кроме моделей, содержащих только количественные объясняющие переменные (X_i), в регрессионном анализе рассматриваются также модели, содержащие лишь качественные переменные (D_i), либо те и другие одновременно.

Все процедуры регрессионного анализа (оценка параметров модели, проверка значимости её коэффициентов и т.п.) проводятся при включении фиктивных так же, как и обычных, количественных объясняющих переменных.

Коэффициент регрессии при фиктивной переменной интерпретируется как среднее изменение зависимой переменной при переходе от одной категории (женский пол) к другой (мужской пол) при неизменных значениях остальных параметров. На основе t – критерия Стьюдента делается вывод о значимости влияния фиктивной переменной, существенности расхождения между категориями.

Регрессионные модели, содержащие лишь качественные объясняющие переменные, называются ANOVA – моделями (моделями дисперсионного анализа).

Например, пусть Y - начальная заработная плата.

$$D = \begin{cases} 0, & \text{если претендент не имеет высшего образования,} \\ 1, & \text{если претендент имеет высшее образование.} \end{cases}$$

Какова ожидаемая средняя заработная плата претендента?

Зависимость между Y и D можно выразить моделью парной регрессии:

$$Y = \beta_0 + \gamma D + \varepsilon.$$

Очевидно $M(Y | D = 0) = \beta_0 + \gamma \cdot 0 = \beta_0$; $M(Y | D = 1) = \beta_0 + \gamma \cdot 1 = \beta_0 + \gamma$.

При этом β_0 – определяет среднюю начальную заработную плату при отсутствии высшего образования.

Коэффициент γ указывает, на какую величину отличаются средние начальные заработные платы при наличии и при отсутствии высшего образования у претендента.

Проверяя статистическую значимость коэффициента γ с помощью t -статистики либо значимость коэффициента детерминации R^2 с помощью F –статистики, можно определить, влияет или нет наличие высшего образования на начальную заработную плату.

Модели ANOVA – представляют собой кусочно-постоянные функции. Однако такие модели в экономике крайне редки. Чаще

встречаются модели, содержащие как качественные, так и количественные переменные.

Модели, в которых объясняющие переменные носят как количественный, так и качественный характер, называются ANCOVA – моделями (моделями ковариационного анализа).

1. ANCOVA – модель при наличии у фиктивной переменной двух альтернатив.

Рассмотрим простейшую ANCOVA – модель с одной количественной и одной качественной переменной, имеющей два альтернативных состояния: $Y = \beta_0 + \beta_1 X + \gamma D + \varepsilon$.

Пусть, например, Y – заработная плата сотрудника фирмы, X – стаж сотрудника, D – пол сотрудника, т.е.

$$D = \begin{cases} 0, & \text{если сотрудник - женщина,} \\ 1, & \text{если сотрудник - мужчина.} \end{cases}$$

Тогда ожидаемое значение зарплаты сотрудников при x годах трудового стажа будет:

$$M(Y | x, D = 0) = \beta_0 + \beta_1 x \text{ для женщин } (\bullet),$$

$$M(Y | x, D = 1) = \beta_0 + \beta_1 x + \gamma = (\beta_0 + \gamma) + \beta_1 x \text{ для мужчин } (\bullet\bullet).$$

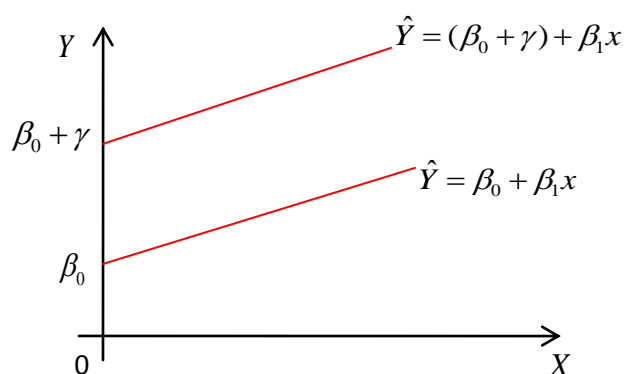


Рисунок 9 - ANCOVA – модель с одной количественной и одной качественной переменной

Заработная плата в данном случае является линейной функцией от стажа работы (рис. 9).

Причём для мужчин, и для женщин зарплата меняется с одним и тем же коэффициентом пропорциональности β_1 . А свободные члены в моделях (•) и (••) отличаются на величину γ .

Проверив с помощью t –статистики статистические зависимости коэффициентов β_0 и $(\beta_0 + \gamma)$, можно определить, имеет ли место в фирме дискриминация по половому признаку. Если эти коэффициенты окажутся статистически значимыми, то, дискриминация есть, при этом, при $\gamma > 0$ она будет в пользу мужчин, при $\gamma < 0$ - в пользу женщин.

В данном случае пол сотрудников имеет два альтернативных значения, и в модели это отражается одной фиктивной переменной.

Общее правило: Если качественная переменная имеет k альтернативных значений, то при моделировании используются только $(k - 1)$ фиктивных переменных.

Если не следовать правилу, то при моделировании исследователь попадает в ситуацию совершенной мультиколлинеарности или так называемую ловушку фиктивной переменной.

Значения фиктивной переменной можно изменять на противоположные. Суть модели от этого не изменится. Например, в модели ANCOVA можно положить, что:

$$D = \begin{cases} 0, & \text{если сотрудник - мужчина,} \\ 1, & \text{если сотрудник - женщина.} \end{cases}$$

Однако при этом знак коэффициента γ изменится на противоположный.

Значение качественной переменной, для которого применяется $D = 0$, называется базовым или сравнительным.

Выбор базового значения обычно диктуется целями исследования, но может быть и произвольным.

Коэффициент γ в модели называется дифференциальным коэффициентом свободного члена, так как он показывает, на какую величину отличается свободный член модели при значении фиктивной переменной, равном единице, от свободного члена модели при базовом значении фиктивной переменной.

2. Модели ANCOVA при наличии у качественных переменных более двух альтернатив.

Пусть рассматривается модель с двумя объясняющими переменными, одна из которых количественная, а другая – качественная. Причем качественная переменная имеет три альтернативы.

Например, расходы на содержание ребенка могут быть связаны с доходами домохозяйств и возрастом ребенка: дошкольный, младший школьный и старший школьный.

Так как качественная переменная связана с тремя альтернативами, то по общему правилу моделирования необходимо использовать две фиктивные переменные.

Таким образом, модель может быть представлена в виде:

$$Y = \beta_0 + \beta_1 X + \gamma_1 D_1 + \gamma_2 D_2 + \varepsilon,$$

где Y – расходы, X – доходы домохозяйств,

$$D_1 = \begin{cases} 0, & \text{если дошкольник,} \\ 1, & \text{в противном случае,} \end{cases} \quad D_2 = \begin{cases} 0, & \text{если дошкольник или младший школьник,} \\ 1, & \text{в противном случае,} \end{cases}$$

Образуются следующие зависимости:

Средний расход на дошкольника: $M(Y | D_1 = 0, D_2 = 0) = \beta_0 + \beta_1 X$. (✓)

Средний расход на младшего школьника:
 $M(Y | D_1 = 1, D_2 = 0) = (\beta_0 + \gamma_1) + \beta_1 X$. (✓✓)

Средний расход на старшего школьника:
 $M(Y | D_1 = 1, D_2 = 1) = (\beta_0 + \gamma_1 + \gamma_2) + \beta_1 X$. (✓✓✓)

Здесь γ_1, γ_2 - дифференциальные свободные члены. Базовым значением качественной переменной является значение «дошкольник».

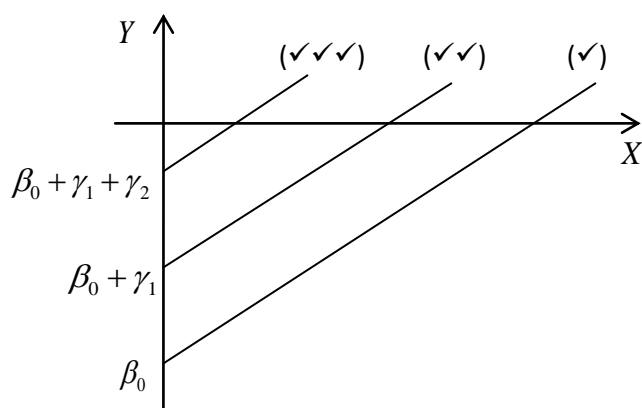


Рисунок 10 - Модели ANCOVA при наличии у качественных переменных более двух альтернатив.

Таким образом, получаются три регрессионные прямые, параллельные друг другу (рис. 10).

После вычисления коэффициентов уравнений регрессий (✓), (✓✓), (✓✓✓) определяется статистическая значимость коэффициентов γ_1 и γ_2 на основе t -статистики.

Если коэффициенты γ_1 и γ_2 оказываются статистически незначимыми, то можно сделать вывод, что возраст ребенка не оказывает существенного влияния на расходы по его содержанию.

3. Регрессия с одной количественной и двумя качественными переменными.

Пусть Y – заработная плата сотрудников фирмы, X – стаж работы, D_1 – наличие высшего образования, D_2 – пол сотрудника:

$$D_1 = \begin{cases} 0, & \text{если нет высшего образования,} \\ 1, & \text{в противном случае,} \end{cases}$$

$$D_2 = \begin{cases} 0, & \text{если сотрудник - женщина,} \\ 1, & \text{если сотрудник - мужчина.} \end{cases}$$

Таким образом, получили следующую модель:

$$Y = \beta_0 + \beta_1 X + \gamma_1 D_1 + \gamma_2 D_2 + \varepsilon$$

Из этой модели выводятся следующие регрессионные зависимости.

Средняя заработная плата женщины без высшего образования:

$$M(Y | D_1 = 0, D_2 = 0) = \beta_0 + \beta_1 X .$$

Средняя заработная плата женщины с высшим образованием:

$$M(Y | D_1 = 1, D_2 = 0) = (\beta_0 + \gamma_1) + \beta_1 X .$$

Средняя заработная плата мужчины без высшего образования:

$$M(Y | D_1 = 0, D_2 = 1) = (\beta_0 + \gamma_2) + \beta_1 X .$$

Средняя заработная плата мужчины с высшим образованием:

$$M(Y | D_1 = 1, D_2 = 1) = (\beta_0 + \gamma_1 + \gamma_2) + \beta_1 X .$$

Все эти регрессии отличаются лишь свободными членами.

Дальнейшее определение статистической значимости коэффициентов γ_1 и γ_2 позволяет убедиться, влияют ли образование и пол сотрудника на его заработную плату.

Предложенные схемы могут быть распространены на ситуации с произвольным числом количественных и качественных факторов.

Многие экономические показатели напрямую связаны с сезонными колебаниями. Например, спрос на туристические путевки, охлажденную воду и мороженое существенно выше летом, чем зимой. Спрос на обогреватели, шубы выше зимой. Некоторые показатели имеют существенные квартальные колебания и т.д.

Обычно сезонные колебания характерны для временных рядов. Устранение ли нейтрализация сезонного фактора в таких моделях

позволяет сконцентрироваться на других важных количественных и качественных характеристиках модели (например, тренде).

Такое устранение сезонного фактора называется сезонной корректировкой. Существует несколько методов сезонной корректировки, одним из них является метод фиктивных переменных.

Пусть переменная Y определяется количественной переменной X , причем эта зависимость существенно различается по кварталам.

Такую общую модель в этой ситуации представим в виде:

$$Y_t = \beta_0 + \beta_1 X_t + \gamma_1 D_{1t} + \gamma_2 D_{2t} + \gamma_3 D_{3t} + \varepsilon_t, (**)$$

где

$$D_{1t} = \begin{cases} 1, & \text{если рассматривается II квартал,} \\ 0, & \text{в противном случае,} \end{cases}$$

$$D_{2t} = \begin{cases} 1, & \text{если рассматривается III квартал,} \\ 0, & \text{в противном случае,} \end{cases}$$

$$D_{3t} = \begin{cases} 1, & \text{если рассматривается IV квартал,} \\ 0, & \text{в противном случае.} \end{cases}$$

Заметим, что число кварталов равно четырем, а, следовательно, число фиктивных переменных должно быть равно трем. В данном примере в качестве базы выбран I квартал.

Если значения Y существенно различаются по кварталам (сезонам), то в уравнении (**) коэффициенты при фиктивных переменных окажутся статистически значимыми. Тогда ожидаемое значение Y по кварталам определяется следующими соотношениями:

$$M(Y | D_{1t} = 0, D_{2t} = 0, D_{3t} = 0) = \beta_0 + \beta_1 X \quad \text{- для I квартала,}$$

$$M(Y | D_{1t} = 1, D_{2t} = 0, D_{3t} = 0) = (\beta_0 + \gamma_1) + \beta_1 X \quad \text{- для II квартала,}$$

$$M(Y | D_{1t} = 0, D_{2t} = 1, D_{3t} = 0) = (\beta_0 + \gamma_2) + \beta_1 X \quad \text{- для III квартала,}$$

$$M(Y | D_{1t} = 0, D_{2t} = 0, D_{3t} = 1) = (\beta_0 + \gamma_3) + \beta_1 X \quad \text{- для IV квартала.}$$

В данной модели (**) рассматриваются ситуации, при которых квартальные различия отражаются лишь в различии свободных членов моделей.

Если же различия затрагивают и изменения коэффициента пропорциональности, то этот факт может быть отражен в модели:

$$Y_t = \beta_0 + \beta_1 X_t + \gamma_1 D_{1t} + \gamma_2 D_{2t} + \gamma_3 D_{3t} + \gamma_4 D_{1t} X_t + \gamma_5 D_{2t} X_t + \gamma_6 D_{3t} X_t + \varepsilon_t. \quad (***)$$

В данной ситуации, выбор правильной формы модели регрессии является серьезной проблемой, т.к. вполне вероятны ошибки спецификации.

Наиболее рациональной практической стратегией выбора модели является следующая схема: вначале рассматривается модель (***). Определяется статистическая значимость коэффициентов. Если дифференциальные угловые коэффициенты оказываются статистически незначительными, то переходят к модели (**). Если в этой модели дифференциальные свободные члены оказываются статистически незначимыми, то делают вывод, что квартальные (сезонные) изменения несущественны для рассматриваемой зависимости.

Во многих экономических задачах встречаются лагированные (взятые в предыдущий момент времени) переменные. Например, y_t – выпуск предприятия за год t – может зависеть не только от инвестиций I_t в этот год, но и от инвестиций в предыдущие годы.

Эконометрическая модель, содержащая в качестве факторов не только текущие переменные, но и лаговые их значения, называется динамической.

Выделим два основных типа динамических эконометрических моделей:

- 1) модели с распределенными лагами;
- 2) модели авторегрессии.

Моделями с распределенным лагом называются модели, содержащие в качестве факторов лаговые значения факторных переменных, например модель вида:

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t.$$

Моделями авторегрессии называются модели, содержащие в качестве факторов лаговые значения зависимой переменной, например модель вида:

$$y_t = \alpha + \beta_0 x_t + \beta_1 y_{t-1} + \varepsilon_t.$$

Обе модели включают в себя лаговые значения переменных, но существенно различаются с точки зрения статистического оценивания параметров.

Модель с распределенным лагом в предположении, что максимальная величина лага конечна, имеет вид:

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \dots + \beta_p x_{t-p} + \varepsilon_t.$$

В этой модели влияние x на y сохраняется в течение времени p .

В краткосрочном (текущем) периоде влияние x на y отражается величиной β_0 , называемой краткосрочным мультипликатором.

В долгосрочном периоде (через p моментов времени) суммарное влияние x на y отражается величиной $\beta = \beta_0 + \beta_1 + \dots + \beta_p$, называемой долгосрочным мультипликатором.

В моделях с распределенным лагом объясняющие переменные некоррелированы со случайным членом, поэтому модель можно оценить с помощью МНК. Однако, на практике, оценка параметров модели затруднительна из-за высокой мультиколлинеарности факторов.

Для уменьшения числа объясняющих переменных и уменьшения мультиколлинеарности разработан ряд подходов, например модель геометрических лагов и модель полиномиальных лагов.

Предположим, что в модели с бесконечным лагом коэффициенты при лаговых значениях объясняющих переменных убывают в геометрической прогрессии. Модель имеет вид:

$$y_t = \alpha + \beta_0 x_t + \beta_0 \delta x_{t-1} + \beta_0 \delta^2 x_{t-2} + \beta_0 \delta^3 x_{t-3} + \dots + \varepsilon_t$$

где $\delta \in (0;1)$.

В этой модели влияние x на y продолжается бесконечно.

В краткосрочном (текущем) периоде влияние x на y отражается коэффициентом β_0 .

В долгосрочном периоде суммарное влияние x на y равно

$$\sum_{k=0}^{\infty} \beta_0 \delta^k = \frac{\beta_0}{1 - \delta}.$$

Модель содержит только три параметра (α , β_0 , δ) и является нелинейной.

Процедура оценивания нелинейной модели заключается в следующем:

- 1) перебирается с некоторым шагом значение δ из интервала $(0;1)$;
- 2) для каждого δ рассчитывается $z_t = x_t + \delta x_{t-1} + \delta^2 x_{t-2} + \delta^3 x_{t-3} + \dots + \delta^p x_{t-p}$ с таким значением p , при котором дальнейшие лаговые значения x не оказывают существенного воздействия на z ;

3) оценивается уравнение регрессии: $y_t = \alpha + \beta_0 z_t + \varepsilon$;

4) выбирается такое значение δ , которое обеспечивает наибольший коэффициент детерминации R^2 при оценке уравнения. Выбранному δ соответствуют вычисленные значения α , β_0 этого уравнения.

Использование этого метода при оценке параметров позволяет избежать проблему мультиколлинеарности объясняющих переменных.

В модели полиномиальных лагов предполагается, что зависимость коэффициентов при лаговых значениях объясняющей переменной от величины лага описывается полиномом m -й степени. Модель имеет вид:

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \dots + \beta_p x_{t-p} + \varepsilon_t.$$

где

$$\beta_s = \gamma_0 + \gamma_1 S + \gamma_2 S^2 + \dots + \gamma_m S^m, \quad m \leq p$$

Предположим, что величина лага p известна. Кроме того, необходимо установить степень полинома m . Обычно на практике ограничиваются рассмотрением полиномов второй и третьей степени.

Пусть, например, $p = 3$, $m = 2$, тогда исходная модель есть

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + \varepsilon_t. \quad \text{где}$$

$$\beta_0 = \gamma_0, \quad \beta_1 = \gamma_0 + \gamma_1 + \gamma_2, \quad \beta_2 = \gamma_0 + 2\gamma_1 + 4\gamma_2, \quad \beta_3 = \gamma_0 + 3\gamma_1 + 9\gamma_2.$$

Преобразованная модель имеет вид:

$$y_t = \alpha + \gamma_0 z_0 + \gamma_1 z_1 + \gamma_2 z_2$$

где

$$z_0 = x_t + x_{t-1} + x_{t-2} + x_{t-3}, \quad z_1 = x_{t-1} + 2x_{t-2} + 3x_{t-3}, \quad z_2 = x_{t-1} + 4x_{t-2} + 9x_{t-3}.$$

Используя МНК, оцениваем параметры преобразованной модели и затем рассчитываем параметры исходной модели с распределенным лагом.

Построим модель с распределенным лагом для $p = 3$ в предположении, что структура лага описывается полиномом второй степени. Имеются данные об объеме валового внутреннего продукта y некоторой страны в зависимости от инвестиций x в ее экономику за 25 лет.

Общий вид исходной модели:

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + \varepsilon_t.$$

Исходные (y_t , x_t) и преобразованные (z_0 , z_1 , z_2) данные (усл.ед.) представлены ниже:

t	y _t	x _t	z ₀	z ₁	z ₂
---	----------------	----------------	----------------	----------------	----------------

1	193	30	—	—	—
2	197	29	—	—	—
3	202	29	—	—	—
4	213	32	120	177	415
5	222	34	124	177	409
6	234	37	132	185	423
7	247	41	144	201	461
8	262	44	156	217	495
9	269	42	164	237	541
10	280	44	171	253	587
11	287	46	176	260	608
12	287	43	175	260	600
13	296	48	181	267	623
14	310	53	190	272	634
15	326	59	203	278	632
16	325	54	214	309	703
17	322	44	210	331	767
18	338	52	209	329	791
19	353	60	210	302	714
20	370	66	222	296	664
21	380	67	245	342	774
22	377	54	247	379	871
23	384	63	250	386	916
24	376	54	238	372	882
25	390	60	231	342	792

Оцененная исходная модель имеет вид:

$$\hat{y}_t = 41,73 + 2,42x_t + 0,71x_{t-1} + 0,99x_{t-2} + 1,47x_{t-3} \quad R^2 = 0,981$$

(9,4) (0,33) (0,39) (0,41) (0,34)

в которой коэффициент 0,71 при переменной x_{t-1} незначим (в скобках указаны стандартные ошибки).

Оцененная преобразованная модель имеет вид:

$$\hat{y}_t = 41,6 + 2,32z_0 - 1,92z_1 + 0,56z_2 \quad R^2 = 0,981$$

(9,27) (0,30) (0,66) (0,21)

и все коэффициенты при переменных значимы.

Получили следующие оценки параметров преобразованной модели:

$$\gamma_0 = 2,32, \quad \gamma_1 = -1,92, \quad \gamma_2 = 0,56$$

$$\beta_0 = 2,32 \quad \beta_1 = 2,32 - 1,92 + 0,56 = 0,96$$

$$\beta_2 = 2,32 - 2 \cdot 1,92 + 4 \cdot 0,56 = 0,72$$

$$\beta_3 = 2,32 - 3 \cdot 1,92 + 9 \cdot 0,56 = 1,6.$$

Таким образом, модель с распределенным лагом имеет вид:

$$\hat{y}_t = 41,6 + 2,32x_t + 0,96x_{t-1} + 0,72x_{t-2} + 1,6x_{t-3}$$

Краткосрочный мультипликатор равен 2,32, а долгосрочный мультипликатор равен $2,32 + 0,96 + 0,72 + 1,6 = 5,6$. Это означает, что увеличение инвестиций в экономику страны на 1 усл.ед. приведет к росту валового внутреннего продукта в среднем на 2,32 усл.ед. в текущем периоде и на 5,6 усл.ед. через 3 года.

Пусть имеется модель авторегрессии вида:

$$y_t = \alpha + \beta_0 x_t + \beta_1 y_{t-1} + \varepsilon_t.$$

Для интерпретации коэффициентов модели авторегрессии сделаем предположение о наличии бесконечного лага в воздействии текущего значения переменной на ее последующие значения и о выполнении неравенства $|\beta_0| < 1$ (условие устойчивости).

В краткосрочном (текущем) периоде влияние x на y отражается величиной β_0 (краткосрочный мультипликатор).

В долгосрочной перспективе суммарное влияние x на y отражается величиной β (долгосрочный мультипликатор):

$$\beta = \beta_0 + \beta_0 \beta_1 + \beta_0 \beta_1^2 + \beta_0 \beta_1^3 + \dots = \frac{\beta_0}{1 - \beta_1}.$$

При построении модели авторегрессии возникают трудности из-за того, что из объясняющих переменных (y_{t-1}) является стохастической (случайной) и зависит от случайного члена.

Если y_{t-1} и ε_t зависимы, но одновременно некоррелированы, то МНК-оценки будут смещенными, но состоятельными.

Существенным здесь является отсутствие автокорреляции случайного члена. Это можно установить по h -тесту Дарбина.

При наличии автокорреляции в авторегрессионной модели МНК-оценки будут смещенными и несостоятельными. В этом случае одним из возможных методов оценки параметров уравнения авторегрессии является метод инструментальных переменных.

В качестве инструментальной переменной можно взять переменную x_{t-1} , которая коррелирована с y_{t-1} и некоррелирована с ε_t . Практически в качестве инструментальной переменной можно взять оценку:

$$\hat{y}_{t-1} = \gamma_0 + \gamma_1 x_{t-1},$$

полученную из предполагаемой линейной зависимости y_{t-1} и x_{t-1} .

Тогда оценку параметров модели автокорреляции можно найти обычным МНК из соотношения:

$$y_t = \alpha + \beta_0 x_t + \beta_1 \hat{y}_{t-1} + \varepsilon_t,$$

где x_t, y_t – исходные, а \hat{y}_{t-1} – расчетные данные.

Практическая реализация метода инструментальных переменных осложняется появлением мультиколлинеарности факторов x_t и \hat{y}_{t-1} в модели.

Построим модель авторегрессии о среднедушевом доходе x и среднедушевых расходах на конечное потребление y . Исходные и расчетные данные (усл.ед.) представлены ниже:

t	y_t	x_t	y_{t-1}	ε_t
1	70	73	–	–
2	73	76	70	-0,401
3	78	83	73	0,416
4	83	89	78	0,580
5	86	95	83	-1,256
6	89	100	86	-1,685
7	96	107	89	1,132
8	96	108	96	-2,852
9	103	113	96	2,264
10	109	119	103	2,397
11	112	121	109	1,550
12	114	122	112	1,627
13	115	131	114	-1,791
14	118	135	115	-0,812
15	122	139	118	0,134
16	123	140	122	-1,304

Применение обычного МНК для расчета параметров этой модели приводит к следующим результатам:

$$\hat{y} = 8,7 + 0,38x_t + 0,51y_{t-1}, \quad R^2 = 0,990$$

(2,7) (0,15) (0,17)

Коэффициент автокорреляции остатков первого порядка составляет $r = -0,071$, следовательно $DW = 2(1-r) = 2.142$.

Значение h-критерия Дарбина, определяемое выражением:

$$h = \left(1 - \frac{DW}{2}\right) \sqrt{\frac{n}{1 - n \text{var}(b)}} = 0,071 \sqrt{\frac{16}{1 - 16 \cdot 0,17^2}} = 0,4,$$

указывает на отсутствие автокорреляции остатков.

Краткосрочный мультипликатор представляет собой предельную склонность к потреблению в текущем периоде и равен 0,38.

Долгосрочный мультипликатор представляет собой предельную склонность к потреблению в перспективе и равен $0,38/(1-0,51)=0,775$.

Следовательно, увеличение дохода на 1 усл.ед. приведет к росту объема потребления в том же году на 0.38 усл.ед., а в перспективе – на 0.775 усл.ед.

В модели частичной корректировки предполагается, что поведенческое уравнение определяет не фактическое значение зависимой переменной y_t , а ее *желаемый* (целевой) уровень:

$$y_t^* = \alpha + \beta x_t + \varepsilon_t, \quad \varepsilon_t - N(0; \sigma^2).$$

Предполагается также, что фактическое значение зависимой переменной не выходит мгновенно на желаемый уровень, а изменяется только на долю λ в нужном направлении:

$$y_t - y_{t-1} = \lambda(y_t^* - y_{t-1}) \quad (0 \leq \lambda \leq 1)$$

Это выражение можно переписать следующим образом:

$$y_t = \lambda y_t^* + (1 - \lambda)y_{t-1},$$

откуда видно, что y_t получается как взвешенное среднее желаемого уровня и фактического значения этой переменной в предыдущем периоде.

Параметр λ называется корректирующим коэффициентом. Чем больше λ , тем быстрее происходит процесс корректировки. Если $\lambda=1$, то $y_t = y_t^*$ и полная корректировка происходит за один период. Если $\lambda=0$, то корректировка y_t не происходит совсем.

Подставляя y_t^{**} в выражение для y_t , получим:

$$y_t = \alpha\lambda + \beta\lambda x_t + (1-\lambda)y_{t-1} + \lambda\varepsilon_t$$

Полученное уравнение включает только фактические значения переменных.

Поскольку случайные члены некоррелированы, состоятельные оценки параметров можно получить, применяя МНК к оцениванию составных параметров $\alpha\lambda$, $\beta\lambda$ и $(1-\lambda)$ в уравнении.

Производственные компании распределяют прибыль Π , оставшуюся после уплаты налогов: одну часть на выплату доходов акционерам в форме дивидендов D , другую – на финансирование инвестиций.

Известны данные (усл.ед.) о деятельности производственных компаний за ряд предыдущих лет:

t	D	Π	t	D	Π
1	100	400	6	800	1100
2	300	600	7	900	1300
3	450	700	8	1000	1400
4	550	800	9	1100	1500
5	700	1000	10	1200	1700

Предположим, что у фирмы имеется целевая долгосрочная доля выплат γ и что желаемый объем дивидендов D_t^* соотносится с текущей

прибылью Π_t как $D_t^* = \gamma \Pi_t + \varepsilon_t$. Однако реальный объем дивидендов подвержен процессу частичной корректировки:

$$D_t - D_{t-1} = \lambda(D_t^* - D_{t-1}), \text{ или } D_t = \gamma \lambda \Pi_t + (1 - \lambda)D_{t-1} + \lambda \varepsilon_t \quad (0 \leq \lambda \leq 1).$$

На основе данных о деятельности производственных компаний за ряд лет построено уравнение регрессии $D_t = 68 + 0,29\Pi_t + 0,58D_{t-1}$, где все коэффициенты значимы.

Из соотношения $1 - \lambda = 0,58$ определяется корректирующий коэффициент $\lambda = 0,42$, а из соотношения $\gamma \lambda = 0,29$ – оценка доли выплат $\gamma = 0,69$.

Предположим, что зависимая переменная y_t связана с ожидаемым значением объясняющей переменной x в $(t+1)$ -м периоде соотношением:

$$y_t = \alpha + \beta x_t^* + \varepsilon_t,$$

где x_t^* – ожидаемое значение ненаблюдаемой объясняющей переменной, которую необходимо заменить наблюдаемыми переменными.

Такая модель возникает, например, в следующем случае: фирма принимает решение об объеме производимой в период t продукции y_t до того, как станет известной цена x_{t+1} , по которой эта продукция может быть продана в следующем периоде. Поскольку цена x_{t+1} неизвестна в период t , то решение принимается на основе ожидаемого значения x_t^* .

Процесс формирования ожиданий таков:

$$x_{t+1}^* - x_t^* = \lambda(x_t - x_t^*), \text{ или } x_{t+1}^* = \lambda x_t + (1 - \lambda)x_t^* \quad (0 \leq \lambda \leq 1),$$

т.е. ожидаемое значение переменной x_t^* в следующем периоде является взвешенным средним ее фактического и ожидаемого значений в текущем периоде.

Параметр λ называется коэффициентом ожидания. Величину x_t^* можно выразить через x_{t-1}^* и т.д. Повторяя эту процедуру бесконечное число раз, получим:

$$x_{t+1}^* = \lambda[x_t + (1-\lambda)x_{t-1} + (1-\lambda)^2 x_{t-2} + \dots].$$

В итоге получаем модель адаптивных ожиданий, в которой ожидаемое значение переменной является взвешенным средним ее прошлых значений с геометрически убывающим весом.

Подставим выражение для x_{t+1}^* в исходную модель и заменим $(1-\lambda)$ на δ :

$$y_t = \alpha + \beta\lambda(x_t + \delta x_{t-1} + \delta^2 x_{t-2} + \dots) + \varepsilon_t,$$

т.е. получим модель геометрических лагов. Параметры уравнения можно оценить методом нелинейного оценивания.

ГЛАВА 3. ИСПОЛЬЗОВАНИЕ ЭКОНОМЕТРИЧЕСКИХ МОДЕЛЕЙ В ПРОГНОЗИРОВАНИИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ

3.1. Применение многофакторных моделей прогнозирования для анализа и прогнозирования демографической ситуации

Одной из важнейших актуальных проблем экономики России является проблема безработицы. Снижение количества безработных является главной проблемой современности, и для этого необходимо выявление факторов, лучшим образом воздействующих на снижение количества безработных [16,27].

На основе статистических данных (табл.11) по 27 муниципальным районам Челябинской области покажем эконометрическое исследование зависимости количества безработных от социально-экономических факторов. Обработка исходной статистической информации проведена на ЭВМ, оснащенной пакетом стандартной программы Statistica [10].

Обозначим, результирующий фактор: Y – количество безработных (чел.); количественные факторы: X_1 – число малых и средних предприятий (ед.); X_2 – объем инвестиций в основной капитал (млн. руб.); X_3 – напряженность на рынке труда (чел./на 1 вак.); X_4 – заявленная организациями потребность в рабочих (чел.); X_5 – численность населения (тыс. чел.); X_6 – уровень занятости (%); X_7 – среднемесячная заработная плата работников (тыс.руб.); X_8 – среднегодовая численность экономически активного населения (тыс. чел.).

Качественные факторы: Z_1 – пол безработного (1 - большинство безработных мужчины, 0 - большинство безработных женщины); Z_2 – высшее образование (1 - большинство безработных имеют высшее образование, 0 - большинство безработных не имеют высшего образования).

Таблица 11 - Статистические данные по 27 муниципальным районам Челябинской области

	У	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
1	539	110	54,4	90	12	34,5	43,0	10,2	13,4
2	383	436	126,7	78	24	41,0	62,0	8,9	25,6
3	1100	183	47,9	272	21	63,8	76,0	12,4	32,2
4	591	87	87,9	102	19	27,8	89,0	8,4	8,9
5	243	210	101,8	90	14	26,7	67,0	12,1	8,6
6	468	102	104,1	68	43	34,3	49,0	16,9	12,3
7	460	241	208,5	22	65	53,0	35,0	17,3	32,1
8	213	320	103,7	81	17	30,7	56,7	10,6	10,3
9	978	220	94,5	56	32	48,7	45,8	11,7	23,4
10	275	190	105,8	89	21	34,0	40,5	9,3	17,0
11	961	297	101,5	45	24	25,6	40,1	9,2	10,6
12	513	120	217,3	77	45	25,4	39,8	8,0	9,5
13	632	267	206,3	54	56	63,8	48,3	12,0	21,5
14	410	156	261,2	67	33	42,4	50,3	12,3	17,9
15	133	187	126,1	44	22	151,1	48,9	10,9	54,2
16	653	221	144,9	46	13	28,9	63,2	9,2	9,0
17	375	97	94,7	70	39	20,4	41,9	14,1	7,8
18	650	266	128,3	55	21	17,8	48,9	9,3	6,7
19	701	96	109,3	81	12	20,7	64,1	7,1	7,9
20	227	174	312,9	67	35	13,8	56,9	8,1	7,9
21	1173	287	178,9	88	15	85,2	65,9	10,8	45,2
22	685	92	125,4	63	14	63,3	61,9	11,9	33,3
23	273	324	222,9	66	14	27,8	58,1	7,9	10,9
24	144	120	175,5	51	18	111,2	69,9	9,3	11,8
25	498	87	138,9	43	16	25,6	69,0	8,0	9,4
26	405	100	198,7	68	12	29,6	50,1	7,6	14,4
27	637	102	157,0	78	13	19,9	50,7	9,9	9,3

Не все переменные следует включать в уравнение регрессии, т.к. есть переменные, которые зависимы между собой, которые существенно не влияют на результирующий фактор У и их нецелесообразно включать в выборочное уравнение регрессии [16]. Для решения этого вопроса проводится анализ факторов на значимость, на мультиколлинеарность с помощью прикладного статистического пакета Statistika, с использованием матрицы парных коэффициентов корреляции (табл.12).

Таблица 12 - Матрица парных коэффициентов корреляции

Пере- менн ые	Переменные								
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	Y
X ₁	1,0000	0,0215	0,8087	0,2108	0,0532	0,5467	0,1689	0,2572	0,6466
X ₂	0,0215	1,0000	0,4361	0,0505	-0,1612	0,1376	0,0643	- 0,1405	0,1915
X ₃	0,8087	0,4361	1,0000	0,2783	-0,0786	0,6905	0,9767	0,1071	0,7980
X ₄	0,2108	0,0505	0,2783	1,0000	0,2823	0,3381	- 0,0807	0,2244	-0,0083
X ₅	0,0532	-0,1612	-0,0786	0,2823	1,0000	0,1084	0,0060	0,4765	0,1085
X ₆	0,5467	0,1376	0,6905	0,3381	0,1084	1,0000	- 0,0987	0,3912	0,6220
X ₇	0,1689	0,0643	0,9767	-0,0807	0,0060	-0,0987	1,0000	- 0,2684	0,1680
X ₈	0,2572	-0,1405	0,1071	0,2244	0,4765	0,3912	- 0,2684	1,0000	0,0621
Y	0,6466	0,1915	0,7980	-0,0083	0,1085	0,6220	0,1680	0,0621	1,0000

Анализируя матрицу парных коэффициентов корреляции, очевидна корреляционная связь между переменными X₃ с X₇ ($r_{X_3 X_7} = 0,97$) примерно равную связи Y с X₇.

Проведем оценку параметров по МНК и дадим оценку их значимости. Сравнивая (табл.3) значения коэффициентов в столбце *p-values* с уровнем значимости 0,05, получаем, что коэффициенты при регрессорах X₁, X₂, X₃, X₄, X₅, X₆, X₈ значимы, так как *p-value* < 0,05, X₇ – незначим, так как *p-value* > 0,05. Поэтому для улучшения модели фактор X₇ - среднемесячная заработная плата работников может быть исключен как недостаточно статистически надежный (в табл.3 столбец *b* - столбец значений параметров при переменных X₁, ..., X₈ соответственно, *StdErrorfb* - стандартная ошибка коэффициента *b*, *t(18)* – значение *t* - статистики Стьюдента для каждого коэффициента *b*, *p-value* - значение равное вероятности того, что случайная величина с данным распределением примет значение, не меньшее, чем фактическое значение тестовой статистики; это наименьшая величина уровня значимости).

Были рассмотрены различные варианты сочетания факторов в модели и оценены параметры (табл.13) моделей с помощью прикладного

статистического пакета Statistika: модель (1) - с количественными факторами ($X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8$), модель (2) - с исключенным фактором X_7 -среднемесячная заработная плата рабочих ($X_1, X_2, X_3, X_4, X_5, X_6, X_8$), модель (3) - с фиктивной переменной - пол безработного ($X_1, X_2, X_3, X_4, X_5, X_6, X_8, Z_1$), модель (4) - с фиктивной переменной высшее образование ($X_1, X_2, X_3, X_4, X_5, X_6, X_8, Z_2$), модель (5) - с объединенными количественными и качественными факторами ($X_1, X_2, X_3, X_4, X_5, X_6, X_8, Z_1, Z_2$).

Таблица13 - Оценка значимости коэффициентов множественной регрессии

Regression Summary for Dependent Variable: Y (Spreadsheet2)						
R= ,94295404 R ² = ,88916232 Adjusted R ² = ,83990113						
F(8,18)=18,050 p<,00000 Std. Error of estimate: 6,0782						
N=27	b*	Std.Err. of b*	b	Std.Err. of b	t(18)	p-value
Intercept			6,103725	5,727552	1,06568	0,300656
X1	-0,591240	0,189389	-0,072046	0,023078	-3,12183	0,005891
X2	-0,408927	0,114952	-0,000003	0,000001	-3,55738	0,002251
X3	1,410665	0,206191	0,191297	0,027961	6,84156	0,000002
X4	-0,427520	0,089733	-0,892169	0,187259	-4,76435	0,000155
X5	0,375969	0,096195	0,000177	0,000045	3,90840	0,001030
X6	0,327694	0,136484	0,001280	0,000533	2,40096	0,027373
X7	-0,060287	0,091149	-0,039535	0,059774	-0,66141	0,516727
X8	-0,222019	0,103127	-0,000313	0,000146	-2,15286	0,045141

Результаты корреляционно - регрессионного анализа представлены в табл. 14. [16]. Очевидно, среди всех рассмотренных моделей - адекватные и точные:

- линейная модель (2) с исключенным фактором X_7 :

$$\hat{Y}_t = 3,915621 + 0,073412 \cdot X_1 - 0,000003 \cdot X_2 + 0,186752 \cdot X_3 - 0,866916 \cdot X_4 + 0,000173 \cdot X_5 + 0,001383 X_6 - 0,000312 \cdot X_8$$

- линейная модель (5) с объединенными количественными и качественными переменными $X_1, X_2, X_3, X_4, X_5, X_6, X_8, Z_1, Z_2$:

$$\hat{Y}_t = 0,168285 - 0,059906 \cdot X_1 - 0,000001 \cdot X_2 - 0,000003 \cdot X_3 + 0,183856 \cdot X_4 + 0,000165 \cdot X_5 + 0,001179 X_6 - 0,000319 \cdot X_8 + 6,453327 \cdot Z_1 + 3,124144 \cdot Z_2$$

Таблица 14 - Результаты корреляционно - регрессионного анализа

Проверяемое условие	Линейная модель с исключенным фактором (2)	Линейная модель с включением одной фиктивной переменной (3)	Линейная модель с включением одной фиктивной переменной (4)	Линейная модель с количественными и качественными переменными (5)
Качество модели				
1. Статистическая значимость коэффициентов регрессии	Prob. < 0,05 все коэффициенты значимы	Prob. < 0,05 все коэффициенты значимы	Prob. < 0,05 все коэффициенты значимы	Prob. < 0,05 все коэффициенты значимы
2. Коэффициент детерминации	$R^2 = 0,88$	$R^2 = 0,91$	$R^2 = 0,89$	$R^2 = 0,92$
3. Проверка значимости модели множественной регрессии (F-тест)	$F_{фак} = 22,09 > F_{кр} = 2,51$ уравнение регрессии значимо	$F_{фак} = 25,21 > F_{кр} = 2,5$ уравнение регрессии значимо	$F_{фак} = 18,60 > F_{кр} = 2,51$ уравнение регрессии значимо	$F_{фак} = 14,4 > F_{кр} = 2,51$ уравнение регрессии значимо
Точность модели				
1. Средняя относительная ошибка аппроксимации	$\bar{E} = 2,58\% < 10\%$	$\bar{E} = 12,23\% > 10\%$	$\bar{E} = 14,89\% > 10\%$	$\bar{E} = 4,95\% < 10\%$
Адекватность модели				
1. Проверка на гетероскедастичность (тест ранговой корреляции Спирмена)	$t_{наб.} < t_{крит.}$ Гетероскедастичность отсутствует	$t_{наб.} < t_{крит.}$ Гетероскедастичность отсутствует	$t_{наб.} < t_{крит.}$ Гетероскедастичность отсутствует	$t_{наб.} < t_{крит.}$ Гетероскедастичность отсутствует
2. Проверка наличия автокорреляции (критерий Дарбина-Уотсона).	$DW = 1,88 \in (1,86; 2)$ гипотеза о наличии автокорреляции отвергается	$DW = 1,92 \in (0,90; 2)$ гипотеза о наличии автокорреляции отвергается	$DW = 1,93 \in (1,86; 2)$ гипотеза о наличии автокорреляции отвергается	$DW = 1,97 \in (1,89; 2)$ гипотеза о наличии автокорреляции отвергается
3. Равенство нулю математического ожидания остаточного ряда	$M(\bar{E}) = 0,02 < 0,5$ выполняется	$M(\bar{E}) = 0,03 < 0,5$ выполняется	$M(\bar{E}) = 0,04 < 0,5$ выполняется	$M(\bar{E}) = 0,006 < 0,5$ выполняется
4. Соответствие нормальному закону распределения	выполняется	выполняется	выполняется	выполняется
Вывод	выборочная модель (2) адекватная	выборочная модель (3) адекватная	выборочная модель (4) адекватная	выборочная модель (5) адекватная

Адекватные и точные модели (2 и 5) были использованы для прогнозирования: предположим: в следующем году число малых и средних предприятий в районе увеличилось на 3% (за счет расширения индивидуального предпринимательства) и стало равным 334; объём инвестиций в основной капитал увеличился на 5% (возрастает объём промышленного производства) и стал равным 232,1 млн.руб., т.е. объясняющие переменные приняли значения вектора $X_0' = (1; 334; 232,1; 61; 16; 28,2; 60,2; 11,5; 1; 1)$.

Точечный прогноз по модели (2):

$$\hat{Y}_0^{(2)} = 3,915621 + 0,073412 \cdot 334 - 0,000003 \cdot 232,1 + 0,186752 \cdot 61 - 0,866916 \cdot 16 + 0,000173 \cdot 28,2 + 0,001383 \cdot 60,2 - 0,000312 \cdot 11,5 = 267$$

Среднее значение количества безработных составит 267 человек.

Точечный прогноз по модели(5):

$$\hat{Y}_0^{(5)} = 0,168285 - 0,059906 \cdot 334 - 0,000001 \cdot 232,1 - 0,000003 \cdot 61 + 0,183856 \cdot 16 + 0,000165 \cdot 28,2 + 0,001179 \cdot 60,2 - 0,000319 \cdot 11,5 + 6,453327 \cdot 1 + 3,124144 \cdot 1 = 281$$

Среднее значение количества безработных составит 281 человек.

Точечный прогноз уточняется указанием интервала значений, в котором с 95% долей уверенности можно ожидать появления прогнозируемой величины: по модели (2) - с надежностью 95% можно утверждать, что количество безработных в муниципальном районе в следующем году составит значение от 246 человек до 288 человек, а по модели (5) - от 257 человек до 305 человек.

Исследование показало, что, с помощью построенных моделей можно вести учет количества безработных, делать прогноз на будущее, определять факторы, которые позволят уменьшить количество безработных и принимать эффективные управленческие решения, воздействуя на уровень безработицы [16].

3.2. Применение многофакторных моделей прогнозирования в социальных процессах

Область эконометрических исследований в настоящее время значительно расширилась, т.к. в наполнении эмпирическим содержанием нуждаются априорные в разных областях рассуждения. К таким областям относятся исследования и в области медицины [17, 28].

За последние десятилетия Россия пережила несколько этапов падения и роста младенческой смертности, вызванные неблагоприятной ситуацией на экономическом рынке. Поэтому возникла проблема выявления факторов, воздействующих на уровень младенческой смертности и факторов, приводящих к ее снижению. Принятие обоснованных решений в этом случае должно опираться на тщательный анализ информации, позволяющий понять закономерности, взаимосвязи, зависимости между различными показателями. Исследования на основе регрессионной модели позволяют решить эту проблему.

В соответствие с основными причинами, приводящими к младенческой смертности, были собраны статистические данные за каждый квартал (табл.15) в течение 10 лет [28].

Были введены переменные:

Y - младенческая смертность (чел.);

X_1 - количество родившихся детей (чел.);

X_2 - коэффициент врачей на 10 тыс.детей;

X_3 - количество детей, рожденных с врожденными пороками развития (чел.);

X_4 - количество детей, родившихся при преждевременных родах (чел.);

X_5 - число аборт (шт.);

X_6 - количество врачей, имеющих сертификат (чел.);

X_7 - мертворожденные дети (чел.).

Покажем результаты исследования (с использованием компьютерной программы EcometricViews) зависимости уровня младенческой смертности от названных факторов на основе множественной регрессионной модели [38].

Таблица 15 - Статистические данные

№ п/п	Период	Младенческая смертность, Y	Родилось детей, X ₁	Коэффициент врачей на 10 тыс.детей, X ₂	Врожденные пороки развития, X ₃	Преждевременные роды, X ₄	Число абортгов, X ₅	Врачи, имеющие сертификат, X ₆	Мертворожденные дети, X ₇
1	1квартал 1г.	29	133	1,9	14	11	286	2	5
2	2квартал 1г.	27	124	1,8	13	10	266	3	5
3	3квартал 1г.	35	161	2,3	16	13	345	3	4
4	4квартал 1г.	28	129	1,8	13	10	276	3	5
5	1квартал 2г.	33	196	2,4	25	11	348	3	3
6	2квартал 2г.	29	172	2,1	22	10	305	3	4
7	3квартал 2г.	35	207	2,6	27	12	368	3	3
8	4квартал 2г.	24	142	1,8	18	8	252	3	3
9	1квартал 3г.	38	218	3,2	27	12	499	3	3
10	2квартал 3г.	36	206	3,1	26	12	472	3	3
11	3квартал 3г.	36	206	3,1	26	12	472	3	2
12	4квартал 3г.	36	206	3,1	26	12	472	3	2
13	1квартал 4г.	50	236	2,9	26	15	533	3	2
14	2квартал 4г.	42	198	2,5	22	12	447	3	0
15	3квартал 4г.	45	212	2,6	24	13	479	3	1
16	4квартал 4г.	43	203	2,5	23	13	458	3	2
17	1квартал 5г.	43	214	2,8	18	15	484	3	3
18	2квартал 5г.	45	224	2,9	18	15	507	3	1
19	3квартал 5г.	39	194	2,6	16	13	439	3	0
20	4квартал 5г.	44	219	2,9	18	15	495	4	0
21	1квартал 6г.	43	226	2,9	21	12	506	4	3
22	2квартал 6г.	39	205	2,6	19	11	459	4	2
23	3квартал 6г.	48	253	3,2	24	14	564	4	4
24	4квартал 6г.	39	205	2,6	19	11	459	4	0
25	1квартал 7г.	37	243	2,6	29	14	542	4	4

Продолжение таблицы									
№ п/п	Период	Младенческая смертность, Y	Родилось детей, X ₁	Коэффициент врачей на 10 тыс.детей, X ₂	Врожденные пороки развития, X ₃	Преждевременные роды, X ₄	Число аборт, X ₅	Врачи, имеющие сертификат, X ₆	Мертворожденные дети, X ₇
26	2квартал 7г.	34	223	2,4	26	13	498	4	2
27	3квартал 7г.	35	230	2,5	27	13	513	4	2
28	4квартал 7г.	34	223	2,4	26	13	498	4	3
29	1квартал 8г.	35	240	2,9	26	15	542	4	3
30	2квартал 8г.	32	219	2,7	24	14	495	4	3
31	3квартал 8г.	36	247	3,0	27	15	557	4	3
32	4квартал 8г.	29	199	2,4	22	12	449	5	3
33	1квартал 9г.	49	234	3,2	18	12	464	5	0
34	2квартал 9г.	45	215	2,9	17	11	426	5	2
35	3квартал 9г.	51	243	3,3	19	12	483	5	3
36	4квартал 9г.	49	234	3,2	18	12	464	5	2
37	1квартал 10г.	29	254	4,0	34	14	419	5	3
38	2квартал 10г.	25	219	3,5	29	12	361	5	3
39	3квартал 10г.	26	228	3,6	30	12	376	5	2
40	4квартал 10г.	27	236	3,7	31	13	390	5	2

Обработка статистической информации проводится с помощью методов корреляционно-регрессионного анализа, и компьютерных программ для их реализации: EconometricViews, STATISTICA .

В ходе исследования особое внимание уделено анализу факторов на мультиколлинеарность, анализу тесноты взаимосвязи факторов с зависимой переменной, вопросам спецификации, оценке параметров модели методом наименьших квадратов, оценке статистической значимости коэффициентов регрессии и самого уравнения регрессии, проблеме гетероскедастичности и автокорреляции остатков.

Решая проблему спецификации модели, были рассмотрены модели: с количественными переменными; с количественными переменными и одной качественной переменной. Из них пять моделей – линейные, две –

нелинейные. Для каждой модели был проведен анализ остатков. В результате чего из семи построенных моделей была выбрана точная и адекватная регрессионная модель (*), которая использовалась для прогнозирования:

$$\hat{Y}_i = 24,78 + 0,40 \cdot (X_1 + X_7) - 1,48 \cdot (X_3 + X_4) - 5,46 \cdot X_6. \quad (*)$$

где $X_1 + X_7$ - количество детей, рожденных вообще, в том числе и мертворожденных,

$X_3 + X_4$ - количество детей, рожденных при преждевременных родах и с врожденными пороками развития,

X_6 - количество сертифицированных врачей,

\hat{Y}_i - групповая средняя.

Коэффициенты полученной регрессии, показывают: если увеличить количество детей, родившихся здоровыми и мертворожденными на 1 человека, то величина младенческой смертности увеличится в среднем на 0,40 человека; если увеличить количество детей, родившихся при преждевременных родах и с врожденными пороками развития на 1 человека, то величина младенческой смертности уменьшится в среднем на 1,48 человека; если увеличить количество сертифицированных врачей на 1 человека, то величина младенческой смертности уменьшится в среднем на 5,46 человека [20,28].

В ходе исследования была построена оценка зависимой переменной [14] модели (*), в предположении, что объясняющие переменные $X_1 + X_7, X_3 + X_4, X_6$ приняли значения, задаваемые вектором $X'_0 = (1 \quad x_{10} \quad x_{20} \quad x_{30})$, где x_{10}, x_{20}, x_{30} - средние значения факторов $X_1 + X_7, X_3 + X_4, X_6$ за 3 года (с 1-3г.г.) и (с 8-10г.г.)

Средние значения факторов первых трех лет составляют:

$\overline{X_1 + X_7} = 178,5; \overline{X_3 + X_4} = 32,17; \overline{X_6} = 2,92$. Средние значения факторов последних трех лет составляют: $\overline{X_1 + X_7} = 233,08; \overline{X_3 + X_4} = 37,42; \overline{X_6} = 4,75$.

Предсказанным значениям Y является оценка \hat{Y}_0 (точечный прогноз):

$$\hat{Y}_0 = 24,78 + 0,40 \cdot (\overline{X_1 + X_7}) - 1,48 \cdot (\overline{X_3 + X_4}) - 5,46 \cdot \overline{X_6}.$$

Для вектора $X'_0 = (1 \ 178,5 \ 32,17 \ 2,92)$ точечный прогноз математического ожидания Y_0 составил:

$$\hat{Y}_0 = 24,78 + 0,40 \cdot 178,5 - 1,48 \cdot 32,17 - 5,46 \cdot 2,917 = 31,34 \quad \text{для вектора}$$

$X'_0 = (1 \ 233,08 \ 37,42 \ 4,75)$ точечный прогноз математического ожидания Y_0 составил: $\hat{Y}_0 = 24,78 + 0,40 \cdot 233,08 - 1,48 \cdot 37,42 - 5,46 \cdot 4,75 = 36,34$.

Для оценки точности определения зависимой переменной был построен доверительный интервал для индивидуальных значений зависимой переменной [14]:

$$\hat{Y}_0 - t_{1-\alpha, n-k-1} \cdot S_{\hat{Y}_0} \leq Y_0^* \leq \hat{Y}_0 + t_{1-\alpha, n-k-1} \cdot S_{\hat{Y}_0},$$

где $S_{\hat{Y}_0} = S \sqrt{1 + X'_0 (X'X)^{-1} X_0}$ - стандартная ошибка \hat{Y}_0 ;

\hat{Y}_0 - точечный прогноз.

$$S_{\hat{Y}_0} = 1,75 \sqrt{1 + \begin{pmatrix} 1 \\ 233,08 \\ 37,42 \\ 4,75 \end{pmatrix} \cdot \begin{pmatrix} 1,21 & 0,00 & -0,01 & -0,04 \\ 0,00 & 0,00 & 0,00 & 0,00 \\ -0,01 & 0,00 & 0,00 & 0,00 \\ -0,04 & 0,00 & 0,00 & 0,06 \end{pmatrix} \cdot (1 \ 233,08 \ 37,42 \ 4,75)} = 1,81$$

С учетом $\hat{Y}_0 = 36,34$, доверительный интервал для индивидуальных значений зависимой переменной Y_0^* трех последних лет составил:

$$30,58 \leq Y_0^* \leq 42,10.$$

С надежностью 0,95 индивидуальное значение уровня младенческой смертности за последние три года, вычисленное при условии, что число детей, рожденных вообще составит 233 человека, количество детей, рожденных при преждевременных родах и с врожденными пороками развития, составит 37 человек и количество сертифицированных врачей составит 5 человек, находится в пределах от 31 до 42 человек.

Доверительный интервал для функции регрессии:

$$\hat{Y} - t_{1-\alpha, n-k-1} \cdot S_{\hat{Y}} \leq Y \leq \hat{Y} + t_{1-\alpha, n-k-1} \cdot S_{\hat{Y}},$$

где $S_{\hat{Y}} = S \sqrt{X_0'(X'X)^{-1}X_0}$ - стандартная ошибка \hat{Y} ;

\hat{Y} - групповая средняя.

$$S_{\hat{Y}} = 1.75 \sqrt{\begin{pmatrix} 1 \\ 233.08 \\ 37.42 \\ 4.75 \end{pmatrix} \begin{pmatrix} 1.21 & 0.00 & -0.01 & -0.04 \\ 0.00 & 0.00 & 0.00 & 0.00 \\ -0.01 & 0.00 & 0.00 & 0.00 \\ -0.04 & 0.00 & 0.00 & 0.06 \end{pmatrix} \begin{pmatrix} 1 & 233.08 & 37.42 & 4.75 \end{pmatrix}} = 0.44$$

С учетом $\hat{Y} = 36,34$, доверительный интервал для функции регрессии зависимой переменной Y трех последних лет составил:

$$34,93 \leq Y \leq 37,75.$$

С надежностью 0,95 средний уровень младенческой смертности за последние три года, при условии, что число детей, рожденных вообще, составит 233 человека, количество детей, рожденных при преждевременных родах и с врожденными пороками развития, составит 37 человек и количество сертифицированных врачей составит 5 человек, находится в пределах от 35 до 38 человек.

Так как в настоящее время количество рожденных детей увеличивается, то предположим, что с 8 – 10гг. родится среднее число – 250 младенцев ($\overline{X_1 + X_7} = 250$). В связи, с улучшающейся экономической обстановкой, государством выделяются денежные средства на различное медицинское оборудование ранней диагностики патологий развития беременности, а также больше выделяется денежных средств на обучение и повышение квалификации врачей-гинекологов и педиатров, поэтому предположим, что за период с 8г. по 10г., количество детей, рожденных при преждевременных родах и с врожденными пороками развития, сократится до 32 человек в среднем ($\overline{X_3 + X_4} = 32$), а количество сертифицированных врачей увеличится до 8 человек ($\overline{X_6} = 8$).

Получим на основе эконометрической модели (*) точечный прогноз математического ожидания Y_0 для «условных» экзогенных переменных: $X'_0 = (1 \ 250 \ 32 \ 8)$: $\hat{Y}_0 = 24,78 + 0,40 \cdot 250 - 1,48 \cdot 32 - 5,46 \cdot 8 = 33,74$, то есть по сравнению с предыдущими тремя годами среднее значение - уровня смертности детей - уменьшится на 2 человека.

Прогнозный интервал математического ожидания Y составит:

$$33,74 - 3,2 \cdot 1,57 \leq Y \leq 33,74 + 3,2 \cdot 1,57$$

$$28,71 \leq Y \leq 38,76$$

$$\text{где } S_{\hat{Y}} = 1,75 \sqrt{\begin{pmatrix} 1 \\ 250 \\ 32 \\ 8 \end{pmatrix} \cdot \begin{pmatrix} 1,21 & 0,00 & -0,01 & -0,04 \\ 0,00 & 0,00 & 0,00 & 0,00 \\ -0,01 & 0,00 & 0,00 & 0,00 \\ -0,04 & 0,00 & 0,00 & 0,06 \end{pmatrix} \cdot (1 \ 250 \ 32 \ 8)} = 1,57$$

С надежностью 0,95 средний уровень младенческой смертности на 8-10 гг., при условии, что число детей, рожденных вообще составит 255 человек, количество детей, рожденных при преждевременных родах и с врожденными пороками развития, составит 32 человек и количество сертифицированных врачей составит 8 человек, составит от 29 до 39 человек, т.е. снизится по сравнению с предыдущими тремя годами.

Таким образом, чтобы снизить уровень младенческой смертности, необходимо:

- развивать и совершенствовать профилактическую и лечебно-диагностическую помощь беременным женщинам, женщинам, желающим завести ребенка, и детям младенческого возраста;
- финансировать здравоохранение на достаточном уровне;
- повышать квалификацию врачей - гинекологов и педиатров.

Выводы, содержащиеся в исследовании, могут быть использованы администрацией и сотрудникам больницы в принятии эффективных управленческих решений по разработке мероприятий по снижению роста младенческой смертности.

Модель призвана помочь администрации и сотрудникам больницы рационально использовать имеющиеся статистические данные, делать краткосрочные прогнозы результативного показателя.

На основе спрогнозированной величины уровня младенческой смертности, с использованием полученной регрессионной модели можно воздействовать на ее уровень, можно определять факторы, которые позволят уменьшить младенческую смертность и принимать наиболее эффективные управленческие решения, тем самым, контролируя ход событий по этой проблеме. Что позволит принимать эффективные управленческие решения по разработке мероприятий по снижению роста младенческой смертности в г. Троицке и Троицком районе.

3.3. Применение многофакторных моделей прогнозирования в сфере информатизации муниципальных образовательных учреждений

За прошедшие тысячелетия школа как социальный институт пережила различные изменения. И каждый раз находились пути и средства, чтобы совершенствовать и обновлять школу. Сегодня в условиях происходящих перемен в российской школе и перехода ее к компьютерным технологиям возникают новые задачи. Необходимо детально рассмотреть происходящие в школе изменения и дать им соответствующую оценку. Поиски решения данной проблемы привели к разработке эконометрической модели, которая позволяет рассматривать информатизацию школы как интерактивный процесс в работе муниципальных образовательных учреждений [9].

Информатизация образования — широкомасштабный процесс трансформации содержания, методов и организационных форм учебной работы, обеспечивающий эффективную социализацию школьников к жизни в условиях информационного общества [15].

Решение задач, связанных с оценкой различных форм учебной работы в условиях информатизации общества и школы, эконометрическими методами, методами прикладной статистики, стало возможным благодаря электронно-вычислительным машинам, оснащенным компьютерными эконометрическими пакетами по реализации математико-статистического анализа.

Сегодня доступными средствами для анализа данных являются различные статистические программные продукты (СПП). Мировая практика исследовательской работы в банковской сфере, в маркетинговых и социологических исследованиях, других социально-экономических областях свидетельствует о широком применении компьютерных систем статистического анализа и обработки данных. За последнее десятилетие вырос спрос на статистические программные продукты и в Российской Федерации. .

В настоящее время существуют множество эконометрических пакетов, в том числе и разработанные в России. Они позволяют наиболее эффективно решать задачи корреляционно-регрессионного анализа и прогнозирования, оказывая существенную помощь исследователю.

Econometric Views – специальный пакет для решения эконометрических задач, разработанный компанией Quantitative Micro Software. В работе с пакетом EViews можно рассматривать методы обработки и анализа информации, а также способы табличного и графического представления полученных результатов. Статистические программные продукты позволяют выполнять на компьютере наиболее трудоемкую работу по расчету различных статистик, выявлению наличия статистической зависимости, оценки параметров, построению графиков и прогноза изучаемых показателей [23].

Область эконометрических исследований в настоящее время значительно расширилась, т.к. в наполнении эмпирическим содержанием

нуждаются априорные в разных областях рассуждения.

Покажем результаты эконометрического исследование уровня информатизации общеобразовательных школ Троицкого района Челябинской области на основе имеющихся пространственных данных [34].

Алгоритм решения поставленной проблемы следующий: необходимо определить круг факторов и показателей, участвующих в модели, провести их анализ; осуществить подготовку исходной информации; построить математическую модель и оценить ее параметры; провести численное решение задачи исследования с использованием программных продуктов; провести анализ полученных результатов и их интерпретацию.

Английские ученые разработали Матрицу ВЕСТА для определения состояния информатизации школы [15]:

1. NotYetdecided - процесс использование информационных компьютерных технологий (ИКТ) только начался и никак не регламентируется.

2. Localized - процесс информатизации в целом не координируется. ИКТ используют по мере необходимости.

3. Coordinated - процесс информатизации и использования ИКТ координируется в рамках всей школы.

4. Transformative - процесс информатизации и средства ИКТ используются регулярно.

5. Embedded - процесс информатизации вышел на новый качественный уровень. Использование ИКТ стало нормой, повседневным элементом образовательной работы всех педагогов и учащихся.

6. Innovative - использование средств ИКТ обновило работу школы по качеству подготовки учащихся.

Для оценки уровня информатизации школ была разработана следующая шкала [14, 30]:

1 уровень: 0 – 20 баллов;
2 уровень: 21 – 40 баллов;
3 уровень: 41 – 55 баллов;

4 уровень: 56 – 70 баллов;
5 уровень: 71 – 90 баллов;
6 уровень: 91 – 100 баллов.

В Троицком районе в настоящее время 28 общеобразовательных школ.

В соответствии с матрицей ВЕСТА для определения состояния информатизации школы и разработанной шкалой была проведена детальная оценка уровня информатизации каждой школы (табл. 16).

Таблица 16- Оценка уровня информатизации школ

№ п/п	Код школы	Уровень информатизации (бал.)
1	74320_s_001	42
2	74320_s_002	46
3	74320_s_003	41
4	74320_s_004	47
5	74320_s_005	42
6	74320_s_006	35
7	74320_s_007	49
...
27	74320_s_027	33
28	74320_s_028	29

Метод анкетирования был использован в процессе сбора данных. Сбор данных проводился в течение месяца с помощью Анкет об информатизации школ. Анкеты заполнялись представителями школ (учителями информатики, членами школьных команд, работниками администрации) на специально организованных семинарах.

Анкетные данные были разделены на две группы: количественные показатели, которые были представлены числом, и качественные, значения которых определяли уровень выполнения некоторого требования .

В соответствии с анкетными результатами были выбраны восемь факторов, для которых были введены следующие обозначения: X_1 - «доступность аппаратных средств»; X_2 - «уровень организации информационно-технологической среды школы»; X_3 - «использование Интернета в учебном процессе»; X_4 - «использование ЦОР в учебном

процессе»; X_5 - «ИКТ-подготовка учителей»; X_6 - «использование ИКТ как инструмента учебной работы школьников»; X_7 - «использование ИКТ для решения административных задач»; X_8 - «современность аппаратных средств» [12]. Обозначим: Y – уровень информатизации. Количественное выражение этих факторов приведено в табл.17.

Таблица 17 - Значения выбранных факторов

№ п/п	Код школы	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	74320_s_001	0,40	18,00	4,80	11,90	40,30	9,40	9,90	1,03
2	74320_s_002	1,80	18,30	0,00	10,20	18,40	15,90	8,30	0,10
3	74320_s_003	0,76	18,60	3,00	7,80	35,70	9,60	4,80	0,08
4	74320_s_004	0,76	15,40	7,00	11,80	38,50	14,80	6,30	0,06
5	74320_s_005	0,79	11,50	9,80	9,80	18,20	6,40	7,60	1,08
6	74320_s_006	0,78	11,60	0,00	10,30	21,20	14,30	5,30	3,08
...
24	74320_s_024	0,51	6,80	0,00	5,40	15,20	14,20	4,30	0,09
25	74320_s_025	1,05	6,60	0,00	7,30	9,00	14,30	6,70	3,12
26	74320_s_026	0,53	14,50	0,00	8,90	19,00	8,70	5,50	0,10
27	74320_s_027	0,66	9,90	0,00	7,00	9,90	10,00	3,50	0,13
28	74320_s_028	1,14	6,00	0,00	10,60	15,30	10,20	2,90	0,19

Таким образом, была получена пространственная выборка, представленная в виде матрицы размерности 28×8 .

В качестве экзогенных (объясняющих) переменных выступают отобранные выше факторы ($X_1 - X_8$), эндогенная (объясняемая) переменная – уровень информатизации (Y) школ Троицкого района.

Проблеме спецификации модели уделено большое внимание. Предварительный анализ данных показал, что использование логарифмической и полулогарифмической моделей затруднено в связи с малыми значениями отдельных факторов (X_1, X_8), поэтому остановились на выборе линейной модели множественной регрессии [6,17] вида:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \varepsilon_i, \quad i=1,2,\dots,n \quad (1)$$

С помощью эконометрического пакета EViews легко просчитать числовые характеристики факторов, участвующих в исследовании.

Для этого выбирается в рабочем файле View/DescriptiveStats/IndividualSamples, в результате появляется окно, представленное на рис. 10.

	Y	X1	X2	X3	X4	X5	X6	X7	X8
Mean	38.17857	0.908248	13.80357	1.878571	9.596429	21.73929	12.90357	6.350000	0.634500
Median	38.00000	0.845692	13.30000	0.000000	10.00000	18.65000	12.60000	6.200000	0.124144
Maximum	48.00000	1.828233	20.80000	9.800000	12.80000	40.30000	21.30000	10.90000	3.116279
Minimum	27.00000	0.380952	6.000000	0.000000	5.400000	9.000000	6.400000	2.900000	0.037736
Std. Dev.	5.702849	0.349640	4.457244	2.907635	1.928247	8.898204	3.805501	2.004347	0.997398
Skewness	0.022456	1.013412	-0.090813	1.309856	-0.541704	0.604117	0.406994	0.278818	1.708221
Kurtosis	1.999546	4.163603	1.969342	3.442068	2.546749	2.197526	2.485260	2.465478	4.489613
Jarque-Bera	1.170079	6.372321	1.277784	8.234701	1.609078	2.454427	1.082123	0.696118	16.20620
Probability	0.557084	0.041330	0.527877	0.016288	0.447294	0.293108	0.582130	0.706057	0.000303
Observations	28	28	28	28	28	28	28	28	28

Рисунок 10 - Описательные статистики отмеченных переменных

Для оценки параметров множественной регрессии выделяем группу независимых переменных и зависимую переменную, затем открываем окно EquationSpecification, где выбираем метод оценивания - метод наименьших квадратов. В верхнем поле отображен список переменных (см. рис. 10). Оценка параметров модели показана на рис.11:

Equation Specification

Equation Specification:
 Dependent variable followed by list of regressors including ARMA and PDL terms. OR an explicit equation like $Y=c(1)+c(2)*X$.

Y X1 X2 X3 X5 C

Estimation Settings:
 Method: LS - Least Squares (NLS and ARMA)
 Sample: 1 28

OK
 Cancel
 Options

Рисунок 11 - Окно Equation Specification. Оценка параметров

Equation: UNTITLED Workfile: UNTITLED				
View Procs Objects Print Name Freeze Estimate Forecast Stats Resids				
Dependent Variable: Y				
Method: Least Squares				
Date: 06/12/07 Time: 10:06				
Sample: 1 28				
Included observations: 28				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
X1	5.381741	1.627385	3.306986	0.0031
X2	0.478051	0.151718	3.150918	0.0045
X3	0.872834	0.213800	4.082477	0.0005
X5	0.181479	0.076813	2.362624	0.0270
C	21.10690	2.545771	8.290964	0.0000
R-squared	0.777744	Mean dependent var	38.17857	
Adjusted R-squared	0.739091	S.D. dependent var	5.702849	
S.E. of regression	2.912971	Akaike info criterion	5.136657	
Sum squared resid	195.1642	Schwarz criterion	5.374551	
Log likelihood	-66.91320	F-statistic	20.12111	
Durbin-Watson stat	2.057781	Prob(F-statistic)	0.000000	

Рисунок 12 - Объект Equation для модели (*)

В результате лучше других аппроксимирует исходные данные линейная модель с четырьмя количественными переменными (рис.12):

$$Y = C(1)*X1 + C(2)*X2 + C(3)*X3 + C(5)*X5 + C(9)$$

$$\text{или } Y = 21,11 + 5,38X_1 + 0,48X_2 + 0,87X_3 + 0,18X_5, \quad (*)$$

где X_1 – доступность аппаратных средств;
 X_2 – уровень организации информационно-технологической среды школы;
 X_3 – использование Интернета в учебном процессе;
 X_5 – ИКТ-подготовка учителей;

Так как t -статистики всех коэффициентов регрессии больше критического значения ($t_{кр} = 2,069$), то с 95% вероятностью можно утверждать, что все коэффициенты уравнения статистически значимы.

Как видно из рис. 12, все значения в столбце Prob. меньше 0,05., т.е. значение Prob (F-statistic) = 0,000000 < 0,05, то уравнение регрессии статистически значимос вероятностью 95%.

Проведенный Reset – тест модели на ошибку спецификации (рис. 13), подтвердил линейное описание процесса информатизации школ: линейная зависимость включенных факторов наиболее точно описывает реальный процесс, т.к. значение Probability = 0,68 > 0,05.

Ramsey RESET Test:			
F-statistic	0.404990	Probability	0.802703
Log likelihood ratio	2.290978	Probability	0.682412

Test Equation:
 Dependent Variable: Y
 Method: Least Squares
 Date: 06/12/07 Time: 10:09
 Sample: 1 28
 Included observations: 28

Рисунок 13 -Reset – тест модели

Тест Уайта на гетероскедастичность показал (см. рис. 14): значение Probability = 0,62 > 0,05 это значит что дисперсия возмущения зависимой переменной постоянна для любого наблюдения.

Также, в ходе исследования (рис. 14) было установлено, что уровни остаточного ряда независимы, так как автокорреляция отсутствует: d-значение статистики Дарбина-Уотсона попадает в область принятия гипотезы (2;4- d_U) (значение d_U равно 1,747)., поэтому гипотеза об отсутствии автокорреляции принимается.

Equation: UNTITLED Workfile: UNTITLED				
View Procs Objects Print Name Freeze Estimate Forecast Stats Resids				
White Heteroskedasticity Test:				
F-statistic	1.457385	Probability	0.237094	
Obs*R-squared	10.64788	Probability	0.222460	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 06/12/07 Time: 10:16				
Sample: 1 28				
Included observations: 28				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6.663374	24.37374	-0.273383	0.7875
X1	11.01496	24.46679	0.450200	0.6577
X1^2	-3.323991	10.82226	-0.307144	0.7621
X2	-2.497294	2.862544	-0.872403	0.3939
X2^2	0.052109	0.101680	0.512485	0.6142
X3	5.076433	2.790027	1.819492	0.0846
X3^2	-0.652076	0.330240	-1.974552	0.0630
X5	2.652239	1.421616	1.865651	0.0776
X5^2	-0.053319	0.028316	-1.883016	0.0751
R-squared	0.380282	Mean dependent var	6.970152	
Adjusted R-squared	0.119347	S.D. dependent var	9.909167	
S.E. of regression	9.299069	Akaike info criterion	7.552797	
Sum squared resid	1642.981	Schwarz criterion	7.981006	
Log likelihood	-96.73916	F-statistic	1.457385	
Durbin-Watson stat	1.900823	Prob(F-statistic)	0.237094	

Рисунок 14 - Тест Уайта на гетероскедастичность

Оценку качества построенной модели дает коэффициент детерминации.

Найденный (рис.14) коэффициент детерминации: $R^2 = 0,78$, показал, что модель качественная и вариация зависимой переменной на 78% объясняется вариацией независимых переменных включенных в модель.

На основе F – статистики Фишера (рис.3) определяем, что модель в целом значима: $F = 20,12 > F_{кр}(0,05;4;23) = 2,80$.

Средняя относительная ошибка аппроксимации: $\bar{E} = 5,6\% < 10\%$, т.е. в среднем расчетные значения для линейной регрессии отличаются от фактических значений на 5,6%. Следовательно, построенная модель точная.

В ходе исследования установлено, что предпосылки множественного регрессионного анализа выполняются для построенной модели, и построенная выборочная модель множественной регрессии является адекватной, и, следовательно, ее можно использовать для прогнозирования.

Проанализируем, какую информацию дает оцененное уравнение регрессии: константа уравнения регрессии равна 21,11, она дает прогнозируемое значение Y , если входящие в модель переменные будут равны нулю. Это значение соответствует первому уровню информатизации школы по матрице ВЕСТА (NotYetdecided).

Сравним средние уровни информатизации основных и полных (средних) общеобразовательных школ Троицкого района. Для этого вычислим средние значения факторов X_1, X_2, X_3, X_5 для обеих групп школ.

Средние значения показателей информатизации средних (полных) общеобразовательных школ составляют: $\bar{X}_{1cp} = 0,99$, $\bar{X}_{2cp} = 15,93$, $\bar{X}_{3cp} = 3,29$, $\bar{X}_{5cp} = 25,97$.

Средние значения показателей информатизации основных общеобразовательных школ составляют: $\bar{X}_{1осн} = 0,79$, $\bar{X}_{2осн} = 10,61$, $\bar{X}_{3осн} = 0,01$, $\bar{X}_{5осн} = 14,92$.

Подставим средние значения показателей в модель. Тогда для средних школ оценка зависимой переменной составит:

$$\hat{Y} = 21,11 + 5,38 \cdot 0,99 + 0,48 \cdot 15,93 + 0,87 \cdot 3,29 + 0,18 \cdot 25,97 = 41,63;$$

для основных школ:

$$\hat{Y} = 21,11 + 5,38 \cdot 0,79 + 0,48 \cdot 10,61 + 0,87 \cdot 0,01 + 0,18 \cdot 14,92 = 33,12.$$

Очевидно, что уровни информатизации средних и основных школ существенно различаются. Первые начинают выходить на третий уровень (Coordinated), тогда как вторые уверенно находятся на втором (Localized) [15, 30].

В настоящее время активно ведется подключение школ к сети Интернет, не прекращаются поставки в школы новой компьютерной техники. Однако практически не совсем верно определять возможный уровень информатизации как отдельно взятой школы, так и среднее

значение их совокупности, увеличивая (уменьшая) значения только какого-либо фактора.

Все школы района должны иметь свободный доступ в Интернет по выделенной линии или через спутник. Предположим, что ресурсы сети будут активно использоваться во всех школах, как учителями, так и учениками, тогда среднее значение фактора X_3 должно быть примерно равно 7.

В настоящее время среднее значение уровня информатизации общеобразовательных школ находится на уровне 38,2.

Оценим уровень информатизации средних и общих общеобразовательных школ Троицкого района в предположении, что среднее значение фактора X_3 равно 7, а средние значения остальных факторов останутся неизменными, т.е. $\bar{x}_{1cp} = 0,99$, $\bar{x}_{2cp} = 15,94$, $\bar{x}_{3cp} = 7$, $\bar{x}_{5cp} = 25,98$. (для полных школ) и $\bar{x}_{1осн} = 0,79$, $\bar{x}_{2осн} = 10,60$, $\bar{x}_{3осн} = 7,00$, $\bar{x}_{5осн} = 14,90$ (для основных школ).

Подставим эти средние значения в модель. Получили:

для полных школ:

$$\hat{Y} = 21,11 + 5,38 \cdot 0,99 + 0,48 \cdot 15,94 + 0,87 \cdot 7,00 + 0,18 \cdot 25,98 = 45,38;$$

для основных школ:

$$\hat{Y} = 21,11 + 5,38 \cdot 0,79 + 0,48 \cdot 10,60 + 0,87 \cdot 7,00 + 0,18 \cdot 14,90 = 39,21.$$

Таким образом, разница между уровнями информатизации средних и основных школ существенно сократилась (45,38 балла - 39,21 балла) при активном использовании сети Интернет.

Оценим уровень информатизации общеобразовательных школ Троицкого района в предположении, что средние значения факторов увеличатся в два раза, а значение фактора X_3 будет равно 7, т.е. $\bar{x}_1 = 1,82$, $\bar{x}_2 = 27,60$, $\bar{x}_3 = 7,00$, $\bar{x}_5 = 43,48$. Подставляя в модель, получим:

$$\hat{Y} = 21,11 + 5,38 \cdot 1,82 + 0,48 \cdot 27,60 + 0,87 \cdot 7,00 + 0,18 \cdot 43,48 = 58,06.$$

Таким образом, для выхода школ Троицкого района в среднем на 4 уровень информатизации необходимо увеличить все показатели уровня информатизации в среднем в 2 раза и активно использовать Интернет в образовательной деятельности [14]. Уровень информатизации общеобразовательных школ Троицкого района в этом случае составит 58,06 балла.

Исследование показало, что можно учесть нежелательные тенденции, выявленные при прогнозировании и принять необходимые меры для их предупреждения. Применение эконометрической модели имеет значимую роль для принятия эффективных управленческих решений на основе оценки уровня информатизации муниципальных общеобразовательных учреждений [15].

3.4. Применение многофакторных моделей прогнозирования в сельском хозяйстве

Область эконометрических исследований в настоящее время значительно расширилась, т.к. в наполнении эмпирическим содержанием нуждаются априорные в разных областях рассуждения. К таким областям относятся исследования и в области сельскохозяйственной деятельности.

Всемирное увеличение производства зерна является главной задачей отраслей сельского хозяйства, зерноперерабатывающей промышленности, а, следовательно, и социально – экономической и продовольственной стабильности в стране. Для этого требуется внедрение наиболее продуктивных сортов, эффективное использование удобрений, улучшения технологии возделывания зерновых культур, улучшение качества почвы, предсказание точных метеорологических прогнозов и т.д. Поэтому проблема выявления факторов лучшим образом воздействующих на

уровень урожайности зерновых культур, проблема прогнозирования урожайности, актуальна [19,26,38].

На основе статистических данных по 74 региону РФ было проведено статистическое исследование (с помощью статистического пакета EcometricViews), зависимости урожайности зерновых культур (ц/га) от факторов сельскохозяйственного производства [19].

В результате была выбрана лучшая модель множественной регрессии:

$$\hat{Y}_t = 5.374345 + 0,000038 \cdot X_1 - 0.149164 \cdot X_2 - 0.0000074 \cdot X_3 + \\ + 0.002303 \cdot X_4 + 0,000406 \cdot X_6 - 1.122243 \cdot X_7 - \\ - 0.191838 \cdot X_8 + 0.910512 \cdot X_9 - 5.271535 \cdot D_1 + 2.637433 D_2$$

где Y – урожайность зерновых культур (ц/га);

X_1 – количество удобрений (кг/га);

X_2 – число орудий поверхностной обработки почвы (на 100 га);

X_3 – средняя температура воздуха (градусах);

X_4 – число зерноуборочных комбайнов (приведённой мощности на 100 га);

X_6 – число колесных тракторов (приведённой мощности на 100 га);

X_7 – количество осадков в период вегетации (мм);

X_8 – численность работников (чел.);

X_9 – количество химических средств защиты растений (т/га);

D_1 – сорт семян (1- твердый сорт, 0- мягкий сорт);

D_2 – культивирование земли (1- применялась, 0 – не применялась).

На основе построенной модели зависимости урожайности зерновых культур от факторов сельскохозяйственного производства можно делать прогноз на будущее и определять значения факторов, которые можно изменять, чтобы увеличить урожайность зерновых культур [19]. Для выполнения этих задач была создана конфигурация в среде 1С для работы с данными факторами [35]. Она позволяет получить отчет, в котором база данных дает среднее значение урожайности зерновых культур за выбранный период времени по любому району Челябинской области,

графическое представление изменения уровня урожайности зерновых культур и средние значения факторов, влияющих на уровень урожайности зерновых культур.

Была создана информационная система "Урожайность зерновых культур", которая выполняет следующие функции: хранение всех данных о факторах сельскохозяйственного производства; формирование отчетов и графиков [19].

Предусмотрено добавление, редактирование, удаление данных из базы данных. При запросе отчета за выбранный период времени в диапазоне имеющейся информации, база данных дает среднее значение урожайности за этот период в выбранном районе, графическое представление изменения уровня урожайности зерновых культур и средние значения факторов, влияющих на уровень урожайности зерновых культур.

Для созданной конфигурации [24] были спроектированы таблицы 17, 18:

Таблица 17 - Список районов Челябинской области

Название	Семантика	Тип данных	Ключевое	Пример
Код	Код	Число	*	00000001
Наименование	Наименование	Строка		Троицкий район

Таблица 18 - Статистические данные

Название	Семантика	Тип данных	Ключевое	Пример
Период	Период	Дата	*	01.01.2015
Район	Район	Строка	*	Кизильский район
Количество удобрений	Количество Удобрений	Число		30.5
Число орудий поверхностной обработки почвы	Число орудий поверхностной обработки почвы	Число		0.9
Средняя температура воздуха	Средняя температура воздуха	Число		22.5
Число зерноуборочных комбайнов	Число зерноуборочных комбайнов	Число		2.6
Количество минеральных удобрений	Количество минеральных удобрений	Число		20.4
Число колесных тракторов	Число колесных тракторов	Число		1.6
Количество осадков в период вегетации	Количество осадков в период вегетации	Число		57.9
Численность работников	Численность работников	Число		167
Количество химических средств защиты растений	Количество химических средств защиты растений	Число		0.6
Качество почвы	Качество почвы	Число		30.5
Сорт семян	Сорт семян	Число		0
Культивирование земли	Культивирование Земли	Число		1

Для реализации информационной системы были созданы следующие объекты конфигурации (табл. 19).

Таблица 19 - Объекты конфигурации

Наименование	Назначение
Справочники	«Список районов Челябинской области» - хранение списка областей
Регистры	«Статистические данные» - хранение данных о факторах сельскохозяйственного производства
Отчеты	«Отчет» - выдает среднее значение за выбранный период в выбранном районе, графическое представление изменения уровня урожайности зерновых культур и средние значения факторов, влияющих на уровень урожайности культур.

После запуска программы пользователь получает доступ к главному окну программы (рис. 15).

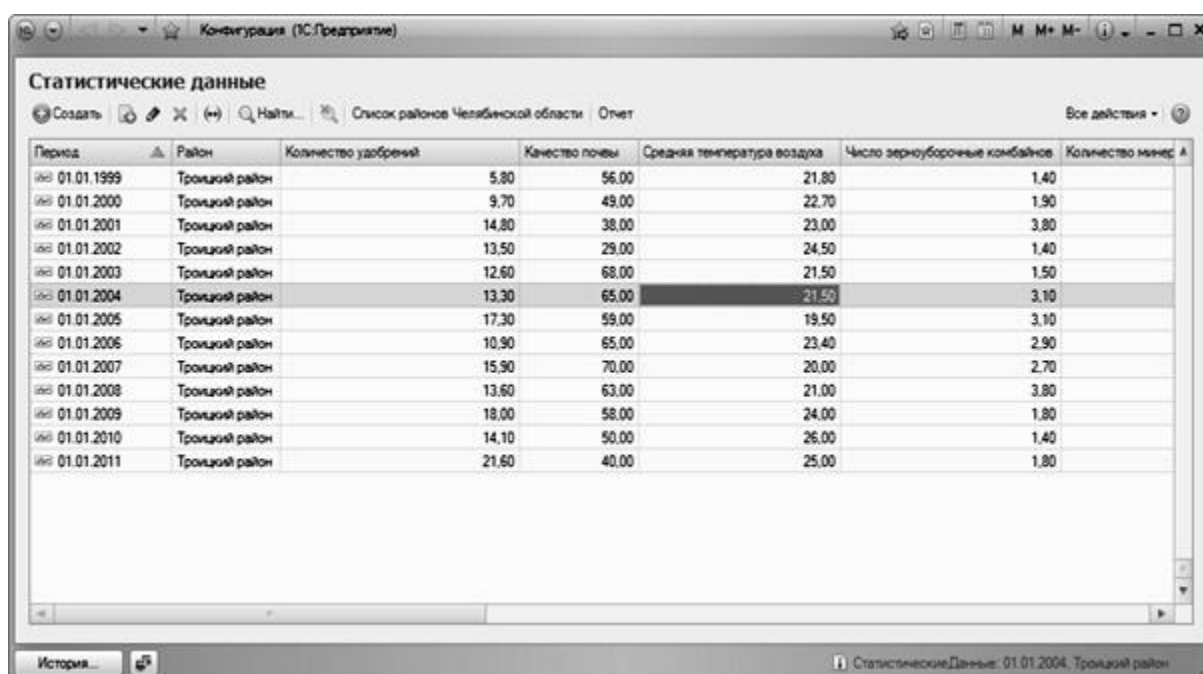


Рисунок 15 - Главная форма диалогового окна программы

На основной панели вкладок расположены основные вкладки доступа к подсистемам конфигурации (рис. 16).



Рисунок 16 - Основная панель

На основной панели при выборе пункта список районов Челябинской области появляется окно для редактирования списка районов, их добавление, удаление (рис. 17).

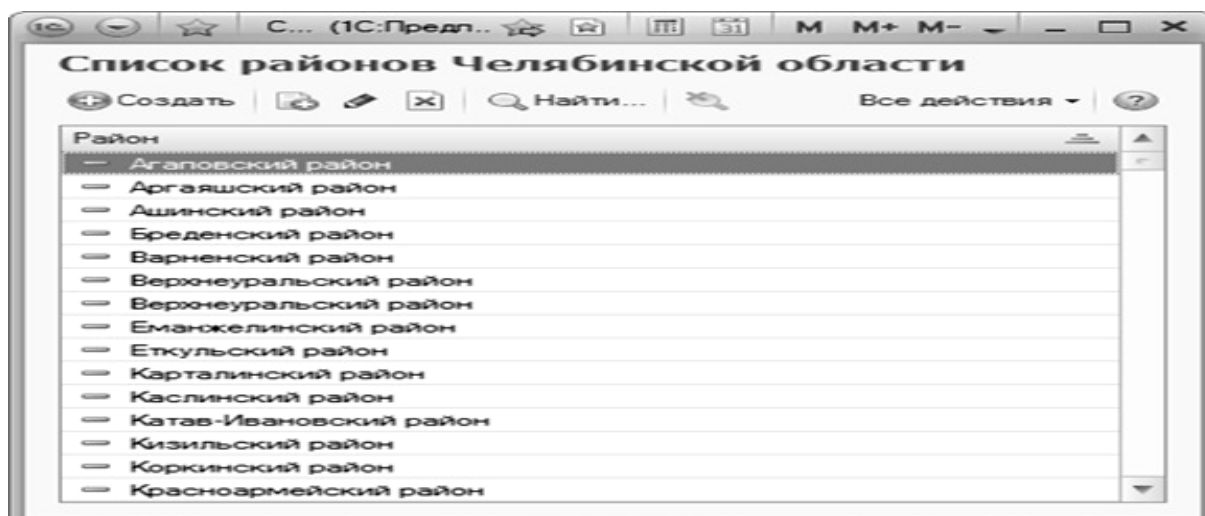


Рисунок 17- Вкладка «Список районов Челябинской области»

При первом запуске пользователь заполняет основные данные о факторах сельскохозяйственного производства в выбранном районе за выбранный год с командной панели, нажав на которую, пользователь сможет изменить, внести или удалить эти данные на вкладке «Статистические данные» (рис. 18):

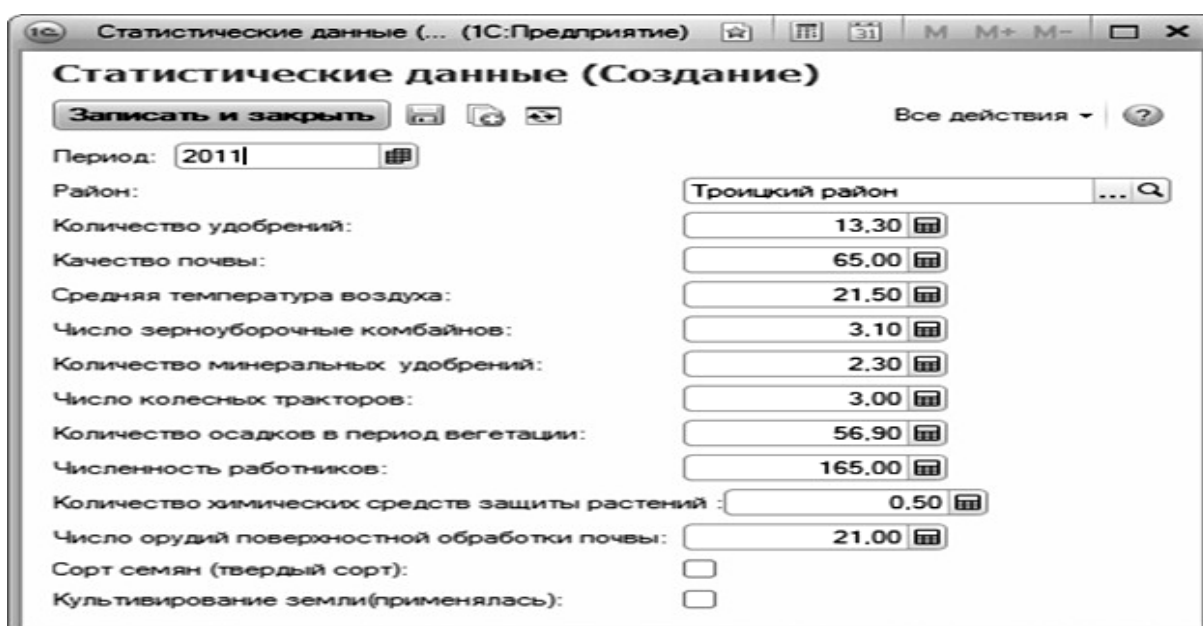


Рисунок18 - Вкладка «Статистические данные»

Для задания отчета на основной панели выбирается вкладка «Отчет» в появившемся окне выбирается район, по которому необходимо сделать отчет и период времени (рис. 19).

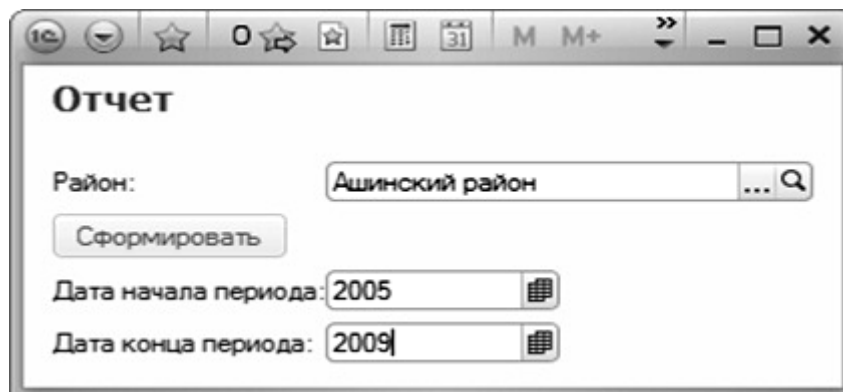


Рисунок 19 - Окно «Выбор района и периода времени для отчета»

При нажатии кнопки «Сформировать» появляется окно (рис. 20), которое содержит следующую информацию, необходимую для анализа, интерпретации полученных данных:

- среднее значение урожайности зерновых культур в выбранном районе, рассчитанное по выборочному регрессионному уравнению:

$$\hat{Y}_t = 5.374345 + 0,000038 \cdot X_1 - 0.149164 \cdot X_2 - 0.0000074 \cdot X_3 + \\ + 0.002303 \cdot X_4 + 0,000406 \cdot X_6 - 1.122243 \cdot X_7 - \\ - 0.191838 \cdot X_8 + 0.910512 \cdot X_9 - 5.271535 \cdot D_1 + 2.637433 D_2$$

- диаграмму изменения урожайности зерновых культур за выбранный период

- среднестатистические данные по всем факторам, имеющимся в исходных данных.

Данная база данных и программа позволяет руководителю на основе построенного выборочного регрессионного уравнения делать прогноз на будущее и определять значения факторов, которые можно изменять, имитируя различные сценарии экономического развития анализируемого показателя [19].

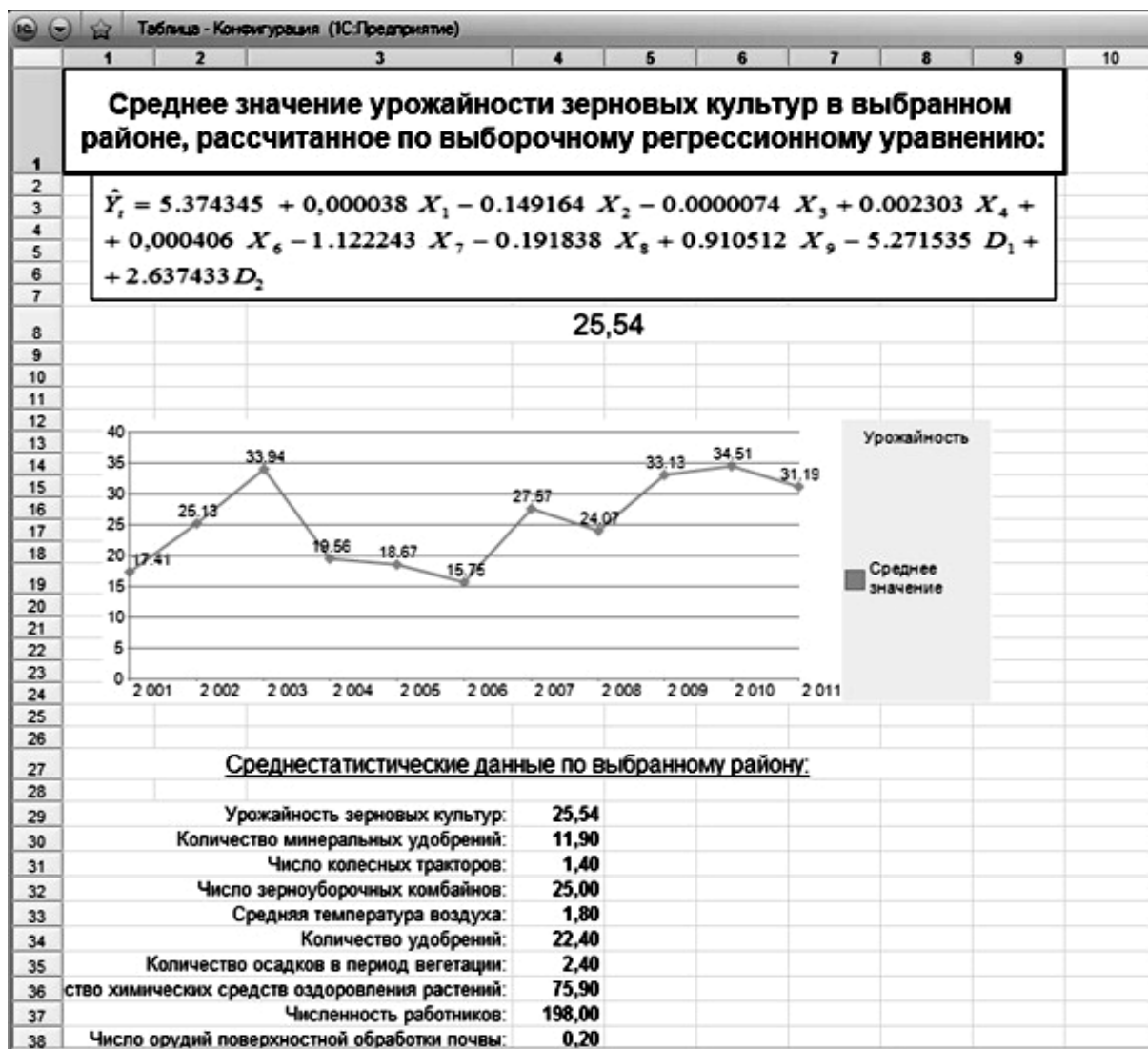


Рисунок 20 - Окно «Отчет по выбранному району и периоду»

Тем самым, выявляя возможные альтернативы развития в перспективе, для обоснованного их выбора и тем самым помогая принимать наиболее эффективные управленческие решения по разработке мероприятий для увеличения и прогнозирования урожайности зерновых культур.

ЗАКЛЮЧЕНИЕ

Исследование показало, что эконометрическое исследование на основе модели множественной регрессии состоит в модельном описании конкретных количественных взаимосвязей, обусловленных общими качественными закономерностями, изученными в экономической теории; в формировании знаний и навыков эконометрического анализа, необходимых в исследовании экономических процессов; в получении знаний по эконометрическим методам, необходимым для проверки предлагаемых и выявленных новых эмпирических зависимостей; в построении надежного прогноза в результате научно – исследовательских, технологических и др. работ; в овладении практическими навыками в построении эконометрических моделей при изучении экономических явлений и процессов, с использованием компьютерных технологий.

Рыночная экономика требует улучшения использования статистической и экономической информации, проведения эконометрических расчетов будущих результатов хозяйственной деятельности и обоснования путей их достижения

В монографии показано использование эконометрических моделей в прогнозировании социально-экономических процессов в обществе. Область эконометрических исследований в настоящее время значительно расширилась, т.к. в наполнении эмпирическим содержанием нуждаются априорные в разных областях рассуждения.

Приведены эконометрические исследования на протяжении нескольких лет. Представлена методология применения многофакторных моделей прогнозирования для анализа и прогнозирования демографической ситуации. Рассмотрено применение многофакторных моделей в сельском хозяйстве, в социальных процессах, в сфере информатизации муниципальных образовательных учреждений,

Выводы, содержащиеся в исследовании, могут быть использованы руководителями предприятий в принятии эффективных управленческих решений. На основе построенного выборочного регрессионного уравнения они могут делать прогноз на будущее и определять значения факторов, которые можно изменять, имитируя различные сценарии экономического развития анализируемого показателя. Исследование показало, что можно учесть нежелательные тенденции, выявленные при прогнозировании и принять необходимые меры для их предупреждения

Все перечисленные методы наиболее эффективны при активном применении статистических пакетов прикладных программ.

Стандартные статистические методы обработки данных включены в состав электронных таблиц, таких как Excel, Lotus 1-2-3, QuattroPro, и в математические пакеты общего назначения, например Mathcad. Но гораздо большими возможностями обладают специализированные статистические пакеты, позволяющие применять самые современные методы математической статистики для обработки данных.

Среди программных средств можно выделить узкоспециализированные пакеты, в первую очередь статистические - STATISTICA, SPSS, STADIA, STATGRAPHICS, EconometricViews. которые имеют большой набор статистических функций: факторный анализ, регрессионный анализ, кластерный анализ, многомерный анализ, критерии согласия и т. д. Данные программные продукты обычно содержат и средства для визуальной интерпретации полученных результатов: различные графики, диаграммы, представление данных на географической карте.

При анализе данных пользователю статистического программного пакета приходится выполнять вычисления широкого спектра статистик, передавать и преобразовывать данные для их анализа, а также представлять полученные результаты в наглядном виде. Поэтому при

выборе того или иного статистического пакета, для сравнения пакетов, необходимо прежде всего обращать внимание на такие характеристики, как:

- удобство управления данными (экспорт/импорт данных, их реструктуризация);
- статистическое разнообразие (количество статистических модулей);
- графические возможности (наличие встроенного графического редактора);
- возможность показа отдельных элементов графика, возможности экспорта графиков.

Кроме того, большое значение имеет удобство работы с пакетом, легкость его освоения (наличие встроенной системы помощи, руководства пользователя, степень удобства управления данными, результатами вычислений, таблицами и графиками), а также скорость произведения вычислений.

Существуют также нестатистические пакеты, решающие задачи классификации (PolyAnalyst, ДА-система, АРГОНАВТ, ЛОРЕГ, пакет ОТЭКС и разнообразные нейросетевые пакеты).

Для реализации рассмотренных методов многомерной классификации в настоящем исследовании была выбрана достаточно популярная в нашей стране система STATISTICA и EconometricViews.

EconometricViews – эконометрическое программное приложение, позволяющее вводить и обрабатывать статистические данные, позволяет выполнить регрессионный анализ, проводить графический анализ, строить эконометрические модели, строить прогнозы в *Windows*-ориентированной компьютерной среде.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Айвазян, С.А. Прикладная статистика и основы эконометрики / С.А. Айвазян, В.С. Мхитарян. - М. : ЮНИТИ, 2010. - 656 с.
2. Айвазян, С.А. Программное обеспечение персональных ЭВМ по статистическому анализу данных // Компьютер и экономика: экономические проблемы компьютеризации общества. - М., 2005. - С. 91-107.
3. Белько, И.В. Эконометрика : практикум: учебное пособие / И.В. Белько, Е.А. Криштапович. - М. : Издательство Гревцова, 2011. - 224 с.
4. Бережная, Е.В. Математические методы моделирования экономических систем: учебное пособие / Е. В. Бережная, В.И. Бережной. - М. : Финансы и статистика, 2006. - 432 с.
5. Борисенко, Я. М. Механизм управления и оценки ресурсов на агропредприятиях молочной специализации / Я.М. Борисенко // Экономика и предпринимательство. - 2017. - № 6 (83). - С.750 - 756.
6. Бородачев, С. Эконометрика / С.Бородачев. - М. : PalmariumAcademicPublishing, 2012. - 84 с.
7. Бородич, С. А. Эконометрика : практикум : учебное пособие : учебное пособие / С.А. Бородич. - М. : Инфра-М, 2016. - 336 с.
8. Булдык, Г.М. Статистическое моделирование и прогнозирование: учебник / Г.М. Булдык. - Минск : НО ООО «БИП - С», - 2013.- 562 с.
9. Водопьян, Г.М. О построении модели процесса информатизации школы / Г.М. Водопьян, А.Ю. Уваров. - М. : РФРДПО, 2006. - 424 с.
10. Вуколов, Э.А. Основы статистического анализа : практикум по статистическим методам и исследованию операций с использованием

пакетов Statistica и Excel / Э.А. Вуколов. - М. : ИНФРА - М, 2004. - 464 с.

11. Гладилин, А.В. Практикум по эконометрике : учебное пособие / А.В. Гладилин, А.Н. Герасимов, Е.И. Громов. - Ростов-на-Дону : Феникс, 2011. - 336 с.

12. Гордеева, Д.С. Экономика образования : учебное пособие для студентов / Н.А. Дегтярева, Д.А. Гордеева. - Челябинск : «Цицеро», 2017. - 101 с.

13. Дайитбегов, Д. М. Компьютерные технологии анализа данных в эконометрике / Д.М. Дайитбегов, М. : ИНФРА , 2008. - 578 с.

14. Дегтярева, Н. А. Информатизация общеобразовательных школ // Функциональная и прикладная наука: сборник научных статей по итогам научно-исследовательской работы за 2014 г. / под науч. ред. М.В. Потаповой. - Челябинск : Изд-во Челяб.гос.пед.ун-та, 2015. - С. 91-94.

15. Дегтярева, Н.А. Использование информационных технологий в управлении / Н.А. Дегтярева, И.Д. Колмакова // Вестник факультета управления Челябинского государственного университета. Серия: «Управление». - 2016. - № 1. - С. 99 - 101.

16. Дегтярева, Н.А. Исследование зависимости количества безработных от социально-экономических факторов на основе модели множественной регрессии // Фундаментальная и прикладная наука. – Челябинск : Из-во Челяб. гос. пед. ун-та. - 2016. - № 2. - С. 13-16.

17. Дегтярева, Н.А. Модели принятия управленческих решений в сфере информатизации муниципальных образовательных учреждений / Н.А. Дегтярева, Е.М. Колмакова, И.Д. Колмакова, // Вестник Челябинского государственного университета. Серия: «Экономические науки ». - № 5 (401) 2017. - вып. 57. - С. 15 - 23.

18. Дегтярева, Н.А. Повышение экономической эффективности функционирования крестьянских (фермерских) хозяйств в условиях рынка: диссертация на соискание ученой степени кандидата экономических наук /

Дегтярева Нина Адамовна; Челябинская государственная агроинженерная академия. - Челябинск, 2000. - 218 с.

19. Дегтярева, Н.А. Применение статистических методов исследования в сельском хозяйстве / Н.А. Дегтярева, Н.А. Берг // Известия высших учебных заведений. Уральский регион. - 2017. - № 1. - С.42 - 47.

20. Дегтярева, Н.А. Сборник задач по статистике : учебное пособие для студентов / Н.А. Дегтярева. - Челябинск : Цицеро, 2017. - 90 с.

21. Дегтярева, Н.А. Сборник задач по экономико-математическим методам и моделям : учебное пособие для студентов / Н.А. Дегтярева. - Челябинск : Цицеро, 2017. - 77 с.

22. Дубина, И.Н. Математико-статистические методы в эмпирических социально-экономических исследованиях / И.Н. Дубина. - М. : Финансы и статистика, Инфра, 2010. - 416 с.

23. Дубров, А.М. Последовательный анализ в статистической обработке информации / А.М. Дубров. - М. : Статистика, 2004. - 160 с.

24. Дубянский, В.М. 1С Предприятие. Конфигурирование и администрирование для начинающих: эспресс-курс / В.М. Дубянский. - М. : БХВ- Петербург, 2005. - 176 с.

25. Елисеева, И. И. 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, 2009. - 576 с.

26. Елисеева, И. И. Статистика : учебник для бакалавров / И. И. Елисеева и др. - Москва : Юрайт, 2011. - 565 с.

27. Колмакова, Е.М. Роль государственных и муниципальных органов власти в развитии социально-трудовой сферы / Е.М. Колмакова, И.Д. Колмакова, Н.А. Дегтярева // Вестник Челябинского государственного университета. Серия: «Экономические науки». - № 5. - С.15 - 23.

28. Колмакова, И.Д. К вопросу о применении эконометрических методов исследования в социальных процессах / И.Д. Колмакова, Н.А. Дегтярева, Е.М. Колмакова // Вестник Челябинского государственного университета. Серия: «Экономика». - 2016. - вып. 55. - С. 51 - 57.
29. Колпаков, В. Компьютерный практикум по эконометрике : учебное пособие / В. Колпаков. - М. : PalmariumAcademicPublishing, - 2013. - 192 с.
30. Концепция информатизации образования // Информатика и образование. - 1988. - № 6. - С. 3 - 29.
31. Костюнин, В.И. Эконометрика. Учебник и практикум : учебное пособие / В.И. Костюнин. - М. : Юрайт, 2014. - 286 с.
32. Кремер, Н.Ш. Эконометрика : учебник для вузов / Н.Ш. Кремер, Б.А. Путко, под ред. Н.Ш.Кремера. - М. : ЮНИТИ - ДАНА, 2011. - 311 с.
33. Латышева, Е. Лабораторный практикум по математической статистике и эконометрике : учебное пособие / Е. Латышева, С. Строганов. - М. : LAP LambertAcademicPublishing, 2014.- 92 с.
34. Мардас, А.Н. Эконометрика : учебник и практикум : учебное пособие / А.Н. Мардас. - М. : Юрайт, 2016. -182 с.
35. Митичкин, С.А. Практика программирования в среде 1С: учебник / С.А. Митичкин. - М. : КомБук, 2004. - 272 с.
36. Наследов, А.Д. SPSS 15: профессиональный статистический анализ данных. - СПб. : Питер, 2008. - 416 с.
37. Орлов, А.И. Эконометрика : учебник для вузов / А.И. Орлов. - 3-е изд., перераб. и доп.- М. : Издательство «Экзамен», 2004. - 576 с.
38. Практикум по эконометрике: учебное пособие / И.И. Елисеева, С.В. Курышева, Н.М. Гордеенко и др.; под ред. И.И. Елисеевой. - 2-е изд., перераб. и доп. - М. : Финансы и статистика, 2007. - 344 с.
39. Социально-экономическая статистика : учебник для бакалавров

/ под ред. М.Р. Ефимовой. 2-е изд., перераб. и доп., - М. : Издательство Юрайт; ИД Юрайт, 2012.

40. Теория статистики с элементами эконометрики : практикум : учебное пособие. - М. : Юрайт, 2015. - 386 с.

41. Тимофеев, В. С. Эконометрика : учебник для академического бакалавриата для студентов вузов, обучающихся по экономическим направлениям и специальностям / В. С. Тимофеев, А. В. Фадеенков, В. Ю. Щеколдин; Новосиб. гос. техн. ун-т. - 2-е изд., перераб. и доп. - М. : Юрайт, 2015. - 328 с.

42. Тинберген, Я. О методе статистического исследования делового цикла. Ответ Дж. М. Кейнсу (рус.) // Вопросы экономики. - 2007. - № 4.

43. Тихомиров, И.П. Методы эконометрики и многомерного статистического анализа / И.П. Тихомиров, Т.М. Тихомиров, О.С. Ушмаев. - М. : Экономика, 2011. - 640 с.

44. Тихомиров, Н.П. Эконометрика / Н.П. Тихомиров, Е.Ю. Дорохина. - М. : Изд-во «Экзамен», 2003. - 512 с.

45. Халафян, А.А. Современные статистические методы медицинских исследований. - М. : Ленанд, 2014. - 320 с.

46. Эконометрика : учебник / под. ред. В.С. Мхитаряна. - М. : Проспект, 2010. - 384 с.

47. Эконометрика : учебник / В. С. Мхитарян, М. Ю. Архипова, В. А. Балаш, О. С. Балаш, Т. А. Дуброва, В. С. Сиротин. - М. : Проспект, 2010. - 384 с.

48. Эконометрика : учебник для вузов / А.И. Орлов. - М. : Издательство «Экзамен», 2004. - 576 с.

49. Эконометрика : учебно-методическое пособие / Шалабанов А.К., Роганов Д.А. - Казань : ТИСБИ, 2005. - 56 с.

50. Эконометрика : практикум : учебное пособие / В.А. Валентинов. - М. : Дашков и Ко, 2010. - 436 с.

51. Экономико-математические методы и прикладные модели : учебное пособие / В.В. Федосеев и др. - М. : ЮНИТИ - ДАНА, 2012. - 304 с. - Режим доступа: <http://www.iprbookshop.ru/15500>.- ЭБС «IPRbooks».

Научное издание

Нина Адамовна Дегтярева

**ЭКОНОМЕТРИЧЕСКИЕ МОДЕЛИ
АНАЛИЗА И ПРОГНОЗИРОВАНИЯ**
Монография

Издательство «Цицеро»

454080, г. Челябинск, Свердловский пр., 60

Подписано к печати 30.11.2017

Формат 60x84 1/16 Объем 10,6 уч-изд.л.

Заказ № 537 Тираж 500 экз

Отпечатано на ризографе в типографии ФГБОУ ВО ЮУрГГПУ

454080, г. Челябинск, пр. Ленина, 69