

УДК 330
ББК 65с631
Э 40

Авторы-составители: Л. П. Авдашкова, канд. физ.-мат. наук, доцент;
М. А. Грибовская, канд. физ.-мат. наук, доцент;
Т. Ф. Калмыкова, канд. техн. наук, доцент;
Л. Н. Марченко, канд. техн. наук, доцент;
Т. М. Моисеева, канд. техн. наук, доцент

Рецензенты: Л. В. Федосенко, канд. экон. наук, доцент кафедры финансов и кредита ГГУ им. Ф. Скорины;
В. В. Бондарева, канд. техн. наук, доцент кафедры информационно-вычислительных систем Белорусского торгово-экономического университета потребительской кооперации

Рекомендовано к изданию научно-методическим советом учреждения образования «Белорусский торгово-экономический университет потребительской кооперации». Протокол № 5 от 14 мая 2011 г.

Эконометрика и экономико-математические методы и модели :
Э 40 пособие для студентов экономических специальностей / авт.-сост. :
Л. П. Авдашкова [и др.]. – Гомель : учреждение образования «Белорусский торгово-экономический университет потребительской кооперации», 2012. – 116 с.
ISBN 978-985-461-946-0

УДК 330
ББК 65с631

В данном пособии излагаются технология построения и анализ эконометрических моделей с использованием MS Excel. Пособие предназначено для использования при аудиторной и самостоятельной работе студентами экономических специальностей и всеми, интересующимися эконометрикой.

ISBN 978-985-461-946-0

© Учреждение образования «Белорусский торгово-экономический университет потребительской кооперации», 2012

СОДЕРЖАНИЕ

Пояснительная записка.....	4
ВВЕДЕНИЕ В ЭКОНОМЕТРИКУ	6
Тема 1. ПАРНАЯ РЕГРЕССИЯ	9
Постановка задачи	9
Технология вычислений в MS Excel для построения и анализа парной регрессии.....	9
Эконометрический анализ построения модели парной регрессии	18
Вопросы для самоконтроля	39
Индивидуальные задания	41
Тема 2. МНОЖЕСТВЕННАЯ РЕГРЕССИЯ	41
Постановка задачи	41
Технология вычислений в MS Excel для построения и анализа линейной множественной регрессии	42
Эконометрический анализ построения модели множественной регрессии	51
Вопросы для самоконтроля	64
Индивидуальные задания	65
Тема 3. ВРЕМЕННЫЕ РЯДЫ	66
Постановка задачи	66
Технология вычислений в MS Excel при построении модели временного ряда ..	66
Эконометрический анализ построения модели временного ряда	70
Вопросы для самоконтроля	79
Индивидуальные задания	80
Тема 4. ЗАВИСИМОСТЬ ПЕРЕМЕННЫХ, ЗАДАННЫХ ВРЕМЕННЫМИ РЯДАМИ.....	80
Постановка задачи	81
Технология вычислений в MS Excel взаимосвязи двух временных рядов	81
Эконометрический анализ построения модели взаимосвязи двух временных рядов	85
Вопросы для самоконтроля	93
Индивидуальные задания	94
Тема 5. СИСТЕМЫ ВЗАИМОЗАВИСИМЫХ УРАВНЕНИЙ	94
Постановка задачи	94
Технология вычислений в MS Excel для модели системы одновременных уравнений.....	95
Эконометрический анализ построения модели системы одновременных уравнений	97
Вопросы для самоконтроля	102
Индивидуальные задания	103
Приложения.....	102
Список рекомендуемой литературы.....	117

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Постоянно усложняющиеся экономические процессы требуют повышения уровня образования современных специалистов по экономике и управлению. Изучение дисциплины «Эконометрика и экономико-математические методы и модели» позволяет использовать моделирование и количественный анализ в экономических исследованиях. Специалист экономического профиля для решения поставленных задач должен владеть научными основами исследования социально-экономических систем, анализа исходных данных, формализации, прогнозирования и принятия оптимальных управленческих решений, используя для этой цели современные технические средства.

Пособие «Эконометрика и экономико-математические методы и модели» предназначено для студентов экономических специальностей для организации аудиторной и самостоятельной работы.

В пособии рассмотрены основные эконометрические вопросы, связанные:

- с построением классических линейных регрессионных моделей (парных и множественных);
- с изучением эконометрических моделей, выраженных системой одновременных уравнений;
- с анализом структуры временных рядов и изучением взаимосвязей между ними;
- с рассмотрением основных ошибок, возникающих при нарушении классических модельных предположений, методикой их диагностики и устранения.

Эконометрическое моделирование предполагает получение модели и ее анализ на качество по определенным статистическим параметрам, которые необходимо найти, используя знания математической статистики. Поэтому вначале предлагается выполнить вычисления параметров с помощью приложения MS Excel, описанные в разделе «Технология вычислений в MS Excel ...», а затем провести дальнейший анализ построения и оценки модели с помощью полученных значений параметров. Такой подход при построении занятия позволяет, с одной стороны, вспомнить сведения из математической статистики и применить их для решения конкретной задачи эконометрического моделирования на этапе расчетов, с другой стороны, на этапе анализа акцентировать внимание на необходимость выполнения определенных статистических условий, не отвлекаясь на выполнение вычислений.

По каждой теме в пособии предлагается технология вычислений,

анализ по полученным параметрам соответствующей модели, вопросы для самоконтроля, индивидуальные задания. Так как этапы эконометрического моделирования справедливы для любой модели, то в темах 1–3 в разделах по технологии вычислений и анализу модели используется одинаковая нумерация этих этапов.

В разделе «Эконометрический анализ построения модели ...» для каждого этапа эконометрического моделирования приводится соответствующий теоретический материал, который позволяет организовать обсуждение проблемных ситуаций, улучшить организацию самостоятельной работы, ответить на вопросы для самоконтроля.

Индивидуальные задания позволяют организовать самостоятельную работу студентов.

В приложениях приводятся примеры оформления листов рабочих книг MS Excel, содержащих вычисления, которые описаны в разделе «Технология вычислений ...», а также образец оформления отчетов по лабораторным работам.

ВВЕДЕНИЕ В ЭКОНОМЕТРИКУ

Экономические исследования требуют от экономистов умений применять экономико-математические методы, создавать эконометрические модели, основываясь на знании экономической теории, экономической статистики, математического моделирования, теории вероятностей, математической статистики, для понимания количественных взаимосвязей между экономическими факторами в целях анализа и прогнозирования реальных экономических процессов.

Экономическая теория с помощью качественного анализа устанавливает совокупность факторов и показателей, влияющих на изучаемое экономическое явление, их роль и теоретические взаимосвязи. Экономическая статистика обеспечивает информационную базу экономических исследований, осуществляя первичную обработку эмпирических значений выбранных экономических показателей. Экономическая статистика, как правило, ограничивается простейшими качественными выводами. Математическая статистика обеспечивает инструментом работы со случайными величинами. Математическое моделирование формализует рассматриваемую экономическую задачу на языке математики. Эконометрика оценивает количественные взаимосвязи изучаемых факторов и использует эти оценки для прогнозирования экономических процессов.

Одной из основных задач эконометрики является построение и анализ эконометрической модели. При этом под *эконометрической моделью* понимается такая форма представления исследуемой экономической задачи с помощью математических терминов и соотношений на основе статистических данных, которая удобна для проведения количественного анализа.

Существуют различные классификации эконометрических моделей. Например, одна из них выделяет типы эконометрических моделей по фактору времени (*статические* и *динамические (модели временных рядов)*). Первые из них исследуют состояние системы в определенный момент времени, т. е. опираются на единовременный срез информации по изучаемым объектам. Вторые строятся по данным, характеризующим изучаемые объекты за ряд последовательных периодов времени, т. е. они используют не только текущие значения показателей, но и некоторые предыдущие по времени значения, а также и само время t .

В зависимости от формы математического представления эконометрические модели подразделяют на *модели с одним уравнением* и *модели системы одновременных уравнений*.

В первом случае объясняемый фактор u выражается через объяс-

няющие переменные x_1, x_2, \dots, x_m с помощью одного уравнения, в котором каждому конкретному набору (x_1, x_2, \dots, x_m) объясняющих факторов x_1, x_2, \dots, x_m соответствует некоторое вероятностное значение зависимой переменной y , выраженное математическим ожиданием $M(y | x_1, x_2, \dots, x_m)$. Такая зависимость называется *корреляционной*. При этом модель называется *парной*, если уравнение связывает только две переменные x, y и описывается уравнением *парной регрессии* вида $M(y | x) = f(x)$. Если же речь идет о зависимости величины y от нескольких факторов x_1, x_2, \dots, x_m ($m \geq 2$), то модель с одним уравнением $M(y | x_1, x_2, \dots, x_m) = f(x_1, x_2, \dots, x_m)$ называется *множественной моделью*. В зависимости от вида функции эконометрические модели разделяются на *линейные* и *нелинейные*.

Модели системы эконометрических уравнений используются при изучении достаточно сложных экономических явлений взаимосвязи между исследуемыми показателями, которые описываются не одним, а несколькими уравнениями (например, модель равновесия спроса и предложения в рыночной экономике).

Выделяют следующие этапы решения эконометрической задачи:

1. *Постановочный этап*. Предполагает определение целей и задач исследования; выделение факторов и показателей, определяющих изучаемые экономические процессы; установление на базе экономической теории роли выбранных показателей.

2. *Этап спецификации*. Выбирается формула связи между переменными, обозначающими выделенные факторы. Эта формула имеет общий вид и содержит параметры (коэффициенты при переменных), требующие статистической оценки.

3. *Этап параметризации*. Решает задачу оценки значений параметров выбранной функции связи.

4. *Этап верификации*. Предполагает проверку адекватности модели, т. е. проверку соответствия модели реальному экономическому явлению или процессу. Кроме того, на данном этапе выясняется, насколько удачно решены проблемы спецификации и параметризации, совершенствуется форма модели, уточняется состав объясняющих переменных, устанавливается точность расчетов по данной модели, определяются общее качество уравнения, статистическая значимость найденных параметров, а также разрешаются многие другие вопросы, определяющие надежность выводов по модели.

Реальное значение зависимой переменной y не совпадает с услов-

ным математическим ожиданием $M(y|x_1, x_2, \dots, x_m)$ и отличается от него на некоторое значение ε , которое носит случайный характер, т. е. $y = M(y|x_1, x_2, \dots, x_m) + \varepsilon$.

Таким образом, величина y разбивается на две части: одна из них (объясняемая) задает ту часть y , которая объясняется факторами x_1, x_2, \dots, x_m , вторая часть ε является случайной величиной и определяет влияние на y неучтенных уравнением $M(y|x_1, x_2, \dots, x_m) = f(x_1, x_2, \dots, x_m)$ других факторов.

Уравнение $y = f(x_1, x_2, \dots, x_m) + \varepsilon$ называется *регрессионной моделью* (или *уравнением регрессионной модели*).

Общая задача эконометрического моделирования заключается в следующем: по имеющимся данным n наблюдений за изменением признака y в зависимости от наборов значений факторов x_1, x_2, \dots, x_m выбрать эконометрическую модель $y = f(x_1, x_2, \dots, x_m) + \varepsilon$, оценить ее параметры и статистически обосновать, что факторы x_1, x_2, \dots, x_m существенны, а построенная функция $f(x_1, x_2, \dots, x_m)$ такова, что наиболее точно соответствует данным наблюдений.

Тема 1. ПАРНАЯ РЕГРЕССИЯ

Постановка задачи

Исследовать зависимость объема экспорта (y , усл. ед.) от валового внутреннего продукта (ВВП) (x , усл. ед.), используя данные 20 наблюдений, приведенные в таблице 1. Построить регрессионную модель $y = f(x) + \varepsilon$. Спрогнозировать объем экспорта при ВВП, равном 2 500 ед.

Таблица 1 – Исходные данные для построения модели

ВВП	1 000	1 090	1 150	1 230	1 300	1 360	1 400	1 470	1 500	1 580
Экспорт	190	220	240	240	260	250	280	290	310	350

ВВП	1 600	1 630	1 700	1 780	1 800	1 850	1 910	1 990	2 010	2 100
Экспорт	340	360	380	400	420	400	400	440	450	470

Технология вычислений в MS Excel для построения и анализа парной регрессии

1. Постановочный этап: *определение целей и задач исследования; выделение факторов и показателей, определяющих изучаемые экономические процессы; установление роли выбранных показателей; подготовка данных для расчетов*

Из экономической теории известно, что экспорт зависит от многих факторов, например, от ВВП, ценовой политики, сертификации и стандартизации продукции, политики других государств, таможенных пошлин и т. д.

Выделим один фактор – ВВП, который является наиболее существенным. Он называется объясняющим фактором (x) для результативного (объясняемого) фактора (y) – экспорта. Поэтому возникает задача количественного описания зависимости указанных экономических показателей уравнением парной регрессии.

Подготовьте данные для расчетов. В ячейку A1 введите название «ВВП», в ячейку B1 – название «Экспорт». В ячейки A2,..., A21 введите значения ВВП, в ячейки B2,..., B21 – соответствующие значения экспорта.

Переименуйте лист 1 в «Исходные данные», щелкнув правой кнопкой мыши на ярлыке «Лист 1» и выбрав команду *Переименование* или

в режиме редактирования после двойного щелчка левой кнопкой мыши на ярлыке «Лист 1». Сохраните открытый файл «Книга 1» под названием «Фамилия-регрессия», например, в папке «Мои документы» (*Файл* → *Сохранить как* → *Папка Мои документы* → *Имя файла* → *Фамилия-регрессия* → *Сохранить*).

2. Спецификация: выбор в общем виде формулы связи между переменными, обозначающими выделенные факторы

Вид и сила функциональной зависимости (линейная или нелинейная) определяются по ковариации и коэффициенту корреляции.

В меню *Сервис* выберите команду *Анализ данных*.

Примечание – Если в меню отсутствует эта команда, то следует выбрать *Сервис* → *Настройка* → установить флажок *Пакет анализа*. При необходимости перезагрузить компьютер.

Выделите анализ *Ковариация*, нажмите кнопку *ОК*. Установите значения параметров в появившемся диалоговом окне следующим образом:

- *Входной интервал* – введите ссылки на диапазон ячеек A1:B21 (курсор установите в поле *Входной интервал*, указатель мыши поместите в ячейку A1, удерживая нажатой левую кнопку, протяните указатель мыши до ячейки B21).

- *Группирование* – флажок *по столбцам* устанавливается автоматически.

- *Метки в первой строке* – установите флажок щелчком левой кнопки мыши.

- *Параметры вывода* – установите флажок на *Выходной интервал*, поставьте курсор в поле напротив, выделите ячейку D2. Нажмите кнопку *ОК*.

В ячейке D1 введите слово *Ковариация* – название таблицы.

Примечание – В приложении А приведены примеры оформления вычислений на листах MS Excel.

В меню *Сервис* выберите команду *Анализ данных* и выделите анализ *Корреляция*. Установите следующим образом значения параметров:

- *Входной интервал* – введите ссылки на ячейки, содержащие исходные данные A1:B21 (курсор установите в поле *Входной интервал*, указатель мыши поместите в ячейку A1, удерживая нажатой левую кнопку, протяните указатель мыши до ячейки B21).

- *Группирование* – флажок по столбцам устанавливается автоматически.

- *Метки в первой строке* – установите флажок.

- *Параметры вывода* – установите флажок на *Выходной интервал*, поставьте курсор в поле напротив, выделите ячейку D7. Нажмите кнопку *ОК*.

В ячейке D6 введите слово *Корреляция* – название таблицы.

Для проверки гипотезы о значимости коэффициента корреляции

сравниваются наблюдаемое $t = \frac{r_{xy} \sqrt{n-2}}{\sqrt{1-r_{xy}^2}}$ и критическое $t_{\alpha, v} = t_{кр}$ зна-

чения статистики Стьюдента, для нахождения которых выполните следующие действия:

- В ячейку D11 введите обозначение tнабл.

- В ячейку E11 введите формулу = E9*КОРЕНЬ(20 – 2)/КОРЕНЬ(1 – E9^2) для вычисления tнабл, где 20 – число наблюдений, 2 – число факторов.

- В ячейку D12 введите tкр.

- В ячейке E12 вычислите критическое значение tкр следующим образом:

- нажмите на f_x (вставка функций);

- в поле *Категория* окна *Мастер функций* выберите *статистические*, из предложенных ниже функций выделите СТЬЮДРАСПОБР и нажмите кнопку *ОК*. Откроется окно *Аргументы функции*. Заполните поля:

- ♦ *Вероятность* – наберите значение альфа, равное 0,05;

- ♦ *Степени свободы* – введите 20–2, где 20 – число наблюдений, 2 – число факторов (x, y). Нажмите кнопку *ОК*.

Примечания:

1. Выводы о существовании зависимости и выборе вида функции связи подробно описаны в разделе «Эконометрический анализ построения модели парной регрессии».

2. Предполагаем, что между ВВП (x) и экспортом (y) существует линейная зависимость. Далее находится уравнение парной линейной регрессии. Если доказана нелинейность зависимости, то проводится процедура линеаризации [6], [9].

3. Параметризация модели: нахождение оценок значений параметров выбранной функции связи

Найдите МНК-оценки неизвестных параметров b_0, b_1 парной линейной регрессионной модели $y = b_0 + b_1x + \varepsilon$, где ε – случайная пе-

ременная, которая включает в себя суммарное влияние всех неучтенных в модели факторов, выполнив действия, представленные ниже.

В меню *Сервис* выберите команду *Анализ данных*, укажите анализ *Регрессия*. Нажмите кнопку *ОК*. Установите в появившемся окне значения параметров следующим образом:

- *Входной интервал Y* – введите ссылки на ячейки B1:B21.
- *Входной интервал X* – введите ссылки на ячейки A1:A21.
- *Метки* – установите флажок.
- *Уровень надежности* – установите флажок.
- *Константа-ноль* – не активизируйте.
- *Параметры вывода* – установите переключатель на *Новый рабочий лист* и в поле напротив введите слово *Регрессия*.
- *Остатки* – установите флажок.
- *Стандартизированные остатки* – оставьте пустым.
- *График остатков* – установите флажок.
- *График подбора* – установите флажок.
- *График нормальной вероятности* – оставьте пустым.

Нажмите кнопку *ОК*. Расположите диаграммы рядом (на поле диаграммы нажмите левую кнопку мыши, затем поместите курсор на белое поле и при нажатой левой кнопке передвигайте диаграмму вниз) и растяните (на поле диаграммы нажмите левую кнопку мыши, нижнюю линию границы диаграммы при нажатой левой кнопке протяните вниз).

Примечание – Выводы о значениях оценок параметров уравнения регрессии подробно описаны в разделе «Эконометрический анализ построения модели парной регрессии».

4. Верификация модели: проверка адекватности модели

4.1. Общее качество уравнения: проверка значимости коэффициента детерминации

Для проверки гипотезы о значимости коэффициента детерминации сравниваются наблюдаемое значение статистики Фишера, найденное с помощью анализа *Регрессия*, и критическое значение $F_{\alpha, v_1, v_2} = F_{кр}$, которое вычислите на листе «Регрессия» в ячейке E15 следующим образом:

- нажмите на f_x (вставка функций);
- в поле *Категория* окна *Мастер функций* выберите *статистические*, из предложенных ниже функций выделите *FPАСПОБР* и нажмите кнопку *ОК*. Откроется окно *Аргументы функции*. Заполните поля:

- *Вероятность* – наберите значение 0,05;
- *Степени свободы 1* – установите курсор в поле и выделите ячейку B12 столбца *df* таблицы «Дисперсионный анализ»;
- *Степени свободы 2* – установите курсор в поле и выделите ячейку B13 столбца *df* таблицы «Дисперсионный анализ». Нажмите кнопку *OK*. В ячейку D15 введите обозначение $F_{кр}$.

Примечание – Выводы о качестве уравнения подробно описаны в разделе «Эконометрический анализ построения модели парной регрессии».

4.2. Нормальность распределения остатков: устанавливается для возможности использования статистики Стьюдента при проверке гипотез (визуально по гистограмме, по асимметрии и эксцессу, с помощью проверки параметрической гипотезы)

В меню *Сервис* выберите команду *Анализ данных*, укажите инструмент анализа *Гистограмма*. Нажмите кнопку *OK*. Значения параметров в появившемся диалоговом окне установите следующим образом:

- *Входной интервал* – введите ссылки на ячейки C24:C44 (ячейки со значениями остатков и названием «Остатки»).
- *Интервал карманов* – не заполняйте.
- *Метки* – установите флажок.
- *Параметры вывода* – установите переключатель на *Новый рабочий лист* и в поле напротив введите его название «Условие 1-нормальность».
- *Парето* – оставьте пустым.
- *Интегральный процент* – оставьте пустым.
- *Вывод графика* – установите флажок. Нажмите кнопку *OK*. Растяните гистограмму вниз. При необходимости ее можно перемещать.

Для нахождения асимметрии и эксцесса, проверки гипотезы о нормальности распределения остатков найдите числовые характеристики остатков.

Скопируйте ячейки C24:C44 листа «Регрессия» со значениями остатков и названием «Остатки» на лист «Условие 1-нормальность» в ячейки C12:C32.

Выберите в меню *Сервис* → *Анализ данных* → *Описательная статистика*, нажмите кнопку *OK*. Значения параметров в диалоговом окне установите следующим образом:

- *Входной интервал* – введите ссылки на ячейки C12:C32.
- *Группирование* – установите флажок *по столбцам*.
- *Метки в первой строке* – установите флажок.
- *Параметры вывода* – установите переключатель на *Выходной*

интервал, курсор установите в поле напротив, выделите ячейку A12.

• Установите флажки *Итоговая статистика*, *Уровень надежности* (95%). Нажмите кнопку *ОК*.

Для проверки гипотезы о нормальности распределения остатков с помощью критерия хи-квадрат Пирсона сравниваются наблюдаемое и критическое значения статистики хи-квадрат.

Рассчитайте наблюдаемое значение хи-квадрат статистики

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}, \text{ выполнив действия, приведенные ниже.}$$

На листе «Условие 1-нормальность» удалите слово *Еще* в ячейке A6 и в этой же ячейке введите формулу = B24*3, нажмите клавишу *Enter* (т. е. значение *максимума остатков* увеличили в три раза).

В ячейку C1 введите значение 0 – значение функции распределения на минус бесконечности.

В ячейки C2:C6 введите **формулу массива**:

- 1) выделите ячейки C2:C6;
- 2) нажмите функциональную F2;
- 3) введите формулу = НОРМРАСП(A2:A6;B14;B18;ИСТИНА);
- 4) нажмите комбинацию клавиш *Ctrl+Shift+Enter*.

Если появилось только одно значение, то нажать F2 и снова *Ctrl+Shift+Enter*. В строке формул при активизации любой ячейки диапазона C2:C6 появится формула в фигурных скобках:

$$\{= \text{НОРМРАСП}(A2:A6;B14;B18;ИСТИНА)\}.$$

Примечание – В дальнейшем фраза «введите формулу массива» предполагает выполнение четырех действий: 1) выделить заполняемый диапазон ячеек; 2) нажать на клавиатуре кнопку F2; 3) ввести формулу; 4) нажать комбинацию клавиш *Ctrl+Shift+Enter*.

В ячейки D2:D6 введите формулу массива {= C2:C6 – C1:C5} (обратите внимание: во втором массиве ячейку C6 не включать).

В ячейки E2:E6 введите формулу массива {= B26*D2:D6}.

В ячейки F2:F6 введите формулу массива {= (B2:В6 – E2:E6)^2/E2:E6}.

В ячейку E7 введите обозначение хи-квадрат набл.

В ячейку F7 введите формулу = СУММ(F2:F6) для вычисления хи-квадрат набл.

Найдите критическое значение статистики Пирсона.

В ячейку E8 введите обозначение хи-квадрат кр.

В ячейку F8 введите формулу для вычисления хи-квадрат кр (используя *Мастер функций – статистические*) =ХИ2ОБР(0,05; 6 – 2 – 1), где 6 – число интервалов (6 получено как сумма числа значений в

кармане и 1).

Примечание – Выводы о нормальности распределения остатков подробно описаны в разделе «Анализ эконометрической модели парной регрессии».

4.3. Значимость коэффициентов регрессии: проверка соответствующих гипотез

Для проверки гипотез о значимости коэффициентов регрессии сравниваются наблюдаемые значения t -статистики, найденные с помощью анализа *Регрессия*, и критическое значение, которое уже найдено на листе «Исходные данные». Для удобства дальнейшего анализа скопируйте содержимое ячеек D12:E12 листа «Исходные данные» в ячейки C19:D19 листа «Регрессия».

Примечание – Выводы о значимости коэффициентов регрессии подробно описаны в разделе «Эконометрический анализ построения модели парной регрессии».

4.4. Проверка статистических свойств остатков (качества оценок коэффициентов регрессии)

4.4.1. Центрированность остатков

Для проверки гипотезы о значимости математического ожидания случайной переменной ε , выборочными оценками которой являются

остатки, сравниваются наблюдаемое $t = \frac{(\varepsilon - 0)\sqrt{n}}{S}$ и критическое

$t_{\alpha, v} = t_{кр}$ значения t -статистики.

На листе «Условие 1-нормальность» в ячейку D14 введите обозначение $t_{набл}$. В ячейку E14 введите формулу $= (B14 - 0) * \text{КОРЕНЬ}(20) / B18$ для подсчета наблюдаемого значения статистики $t_{набл}$. В ячейку D15 введите $t_{кр}$. В ячейку E15 введите формулу $= \text{СТЮДРАСПОБР}(0,05; 20 - 1)$ для подсчета критической точки распределения Стьюдента $t_{кр}$.

4.4.2. Гомоскедастичность (гетероскедастичность) остатков

Для проверки гипотезы о гомоскедастичности случайной переменной сравниваются наблюдаемое и критическое значения t -статистики.

Найдите наблюдаемое значение $t = r\sqrt{n-1}$ t -статистики, где

$r = 1 - \frac{6 \sum D_i^2}{n(n^2 - 1)}$ – коэффициент ранговой корреляции Спирмена; D_i –

разность между рангом x_i и рангом модуля остатка e_i .

На листе «Исходные данные» содержимое ячеек B1:B21 скопируйте в ячейку A1 нового листа и назовите лист «Условие 2». В ячейку B1 скопируйте из листа «Регрессия» столбец «Остатки» вместе с названием. В ячейку C1 введите название «Модуль ост». В ячейки C2:C21 введите формулу массива $\{=ABS(B2:B21)\}$.

Выберите в меню *Сервис* → *Анализ данных* → *Ранг и перцентиль* → *OK* и заполните диалоговое окно следующим образом: *Входной интервал* – введите ссылки на ячейки A1:A21; *Метки в первой строке* – установите флажок; *Выходной интервал* – установите курсор в поле и выделите ячейку D1. Нажмите кнопку *OK*.

Выберите в меню *Сервис* → *Анализ данных* → *Ранг и перцентиль* → *OK* и заполните диалоговое окно следующим образом: *Входной интервал* – введите ссылки на ячейки C1:C21; *Метки в первой строке* – установите флажок; *Выходной интервал* – установите курсор в поле и выделите ячейку H1. Нажмите кнопку *OK*.

Выделите ячейки D2:G21 и нажмите кнопку *Сортировка по возрастанию* на панели инструментов. Выделите ячейки H2:K21 и нажмите кнопку *Сортировка по возрастанию*. Выделите ячейки L2:L21 и введите формулу массива $\{=(F2:F21 - J2:J21)^2\}$. В ячейку L1 введите название «Квадрат разности рангов».

В ячейку K22 введите название «Коэффициент ранговой корреляции».

В ячейку L22 введите формулу

$$= 1 - 6 * \text{СУММ}(L2: L21) / (20 * (20^2 - 1)).$$

В ячейку K23 введите обозначение $t_{\text{наб}}$.

В ячейку L23 введите формулу для вычисления $t_{\text{наб}}$

$$= L22 * \text{КОРЕНЬ}(20 - 1).$$

Найдите критическое значение t -статистики.

В ячейку K24 введите обозначение $t_{\text{кр}}$.

В ячейку L24 введите формулу для вычисления $t_{\text{кр}}$

$$= \text{СТЮДРАСПОБР}(0,05; 20 - 2).$$

4.4.3. Автокорреляция остатков

Для проверки гипотезы об отсутствии автокорреляции случайной переменной сравниваются наблюдаемое и критические значения статистики Дарбина–Уотсона.

Найдите наблюдаемое значение статистики $d = \frac{\sum_{t=2}^n (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^n \varepsilon_t^2}$, ис-

пользуя в качестве оценок значений случайной переменной соответствующие значения остатков, выполнив действия, приведенные ниже.

На листе «Регрессия» объедините ячейки D23 и E23 и введите название «Условие 3».

В ячейку D24 введите название «Числитель днабл».

В ячейки D25:D43 введите формулу массива
 $\{=(C25:C43 - C26:C44)^2\}$.

В ячейку C45 введите слово *Суммы*.

В ячейку D45 введите формулу = СУММ(D25:D43).

В ячейку E24 введите название «Знаменатель днабл».

В ячейки E25:E44 введите формулу массива $\{=(C25:C44)^2\}$.

В ячейку E45 введите формулу = СУММ(E25:E44).

В ячейку D47 введите название «днабл».

В ячейку E47 введите формулу для вычисления днабл = D45/E45.

Примечание – Выводы о выполнении условий Гаусса–Маркова подробно описаны в разделе «Эконометрический анализ построения модели парной регрессии».

4.5. Анализ свойств модели: средний коэффициент эластичности

Средний коэффициент эластичности для линейной регрессии рассчитывается по формуле $\bar{\varepsilon}_{yx} = b_1 \frac{\bar{x}}{y}$.

На листе «Исходные данные» в ячейку A23 введите название «Среднее ВВП», ячейку B23 – «Среднее Экспорт». Введите формулы:

- в ячейку A24 = СРЗНАЧ(A2:A21);
- в ячейку B24 = СРЗНАЧ(B2:B21).

Из листа «Регрессия» скопируйте ячейку B18 (значение *Коэффициента* при ВВП) в ячейку B25 листа «Исходные данные».

В ячейку A26 введите название «Коэф. эласт».

В ячейку B26 введите формулу для вычисления коэффициента эластичности переменной ВВП = B25*B24/A24.

5. Прогнозирование*

Точечный прогноз y находится подстановкой значений объясня-

ющей переменной $x^* = 2\,500$ в уравнение регрессии. Интервальный прогноз, или доверительный интервал прогноза, имеет вид:

$$(y^* - t_{\alpha, \nu} S^*, y^* + t_{\alpha, \nu} S^*),$$

где $t_{\alpha, \nu}$ – критическое значение t -статистики при заданном уровне

значимости α и числе степеней свободы ν ; $S^* = S \cdot \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{(n-1) \cdot \sigma_x^2}}$ –

средняя стандартная ошибка прогноза; S – стандартная ошибка регрессии; σ_x^2 – исправленная дисперсия независимой переменной x .

Вернитесь на лист «Регрессия». Введите:

- в ячейку F23 – название «Точечный прогноз»;
- в ячейку F24 – формулу = B17 + B18*2 500 – точечный прогноз для ВВП, равного 2 500 (из условия);
- в ячейку F26 – название «Интервальный прогноз»;
- в ячейку F27 – формулу для вычисления левого конца интервала прогноза

$$= F24 - D19*B7*КОРЕНЬ(1 + (1/20) + ((2\,500 - СРЗНАЧ('Исходные данные'!A2:A21))^2/((20 - 1)*ДИСП('Исходные данные'!A2:A21))));$$

- в ячейку G27 – формулу для вычисления правого конца интервала прогноза

$$= F24 + D19*B7*КОРЕНЬ(1 + (1/20) + ((2\,500 - СРЗНАЧ('Исходные данные'!A2:A21))^2/((20 - 1)*ДИСП('Исходные данные'!A2:A21)))).$$

Примечания:

1. Запись 'Исходные данные'!A2:A21 означает, что диапазон A2:A21 находится на листе «Исходные данные». Поэтому для заполнения диалоговых окон функции СРЗНАЧ и ДИСП перейдите на лист «Исходные данные» и выделите указанный диапазон ячеек. Набор и редактирование формулы осуществляется в строке формул.

2. Формулы для концов интервала прогноза отличаются знаком после F24, поэтому можно воспользоваться копированием текста формулы.

Эконометрический анализ построения модели парной регрессии

1. Постановочный этап

Из экономической теории известно, что экспорт зависит от многих факторов, например, от ВВП, ценовой политики, сертификации и стандартизации продукции, политики других государств, таможенных пошлин и т. д.

Выделим один фактор – ВВП, который является наиболее существенным. Он называется объясняющим фактором (x) для результирующего (объясняемого) фактора (y) – экспорта. Поэтому возникает задача количественного описания зависимости указанных экономических показателей уравнением парной регрессии $y = f(x) + \varepsilon$ на основе

20 наблюдений экономических показателей.

2. Спецификация модели

Корреляционным полем называется множество точек на плоскости с координатами (x_i, y_i) , $i = 1, 2, \dots, n$, n – объем выборки.

Вид регрессии визуально определяется по корреляционному полю, которое изображено на листе «Регрессия» на графике подбора черными точками по данным 20 наблюдений из листа «Исходные данные» (рисунок 1).

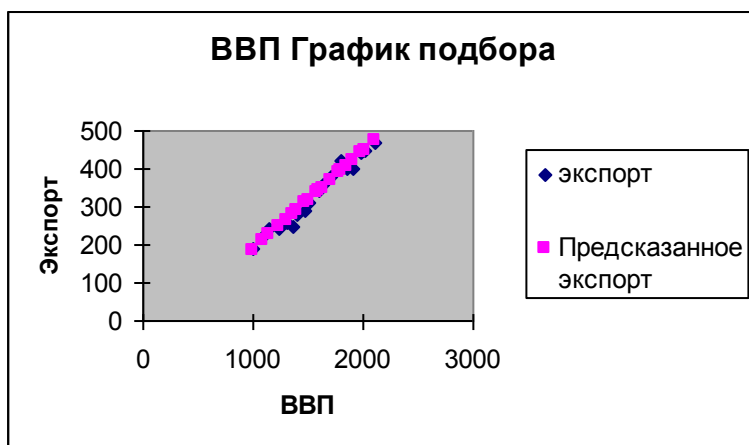


Рисунок 1 – Корреляционное поле (график подбора)

Так как точки сгруппированы вдоль прямой (не горизонтальной),

то можно предположить, что зависимость экспорта (y) от ВВП (x) описывается парной линейной регрессионной моделью $y = b_0 + b_1x + \varepsilon$, где b_0, b_1 – неизвестные параметры модели; ε – случайная переменная, которая включает в себя суммарное влияние всех неучтенных в модели факторов.

Для характеристики вида связи используется *ковариация*, рассчитываемая по формуле $\text{cov}(x, y) = xy - \bar{x} \cdot \bar{y}$. Если $\text{cov}(x, y) > 0$, то возрастание x приводит к увеличению y и связь прямая. Если $\text{cov}(x, y) < 0$, то возрастание x приводит к уменьшению y и связь обратная. Если $\text{cov}(x, y) \rightarrow 0$, то экономические показатели не связаны.

Тесноту связи изучаемых экономических показателей для линейной регрессии оценивает *коэффициент парной корреляции* r_{xy} ($-1 \leq r_{xy} \leq 1$):
$$r_{xy} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\sigma_x \sigma_y}.$$

Коэффициент корреляции необходимо проверить на значимость (значительно ли отличается от нуля), так как он найден по выборочной совокупности, что может привести к неверным выводам о всей генеральной совокупности. Проверка значимости коэффициента корреляции осуществляет-

ся с помощью t -статистики: $t = \frac{r_{xy} \sqrt{n-2}}{\sqrt{1-r_{xy}^2}}$. Величина t имеет

распределение Стьюдента с $\nu = n - 2$ степенями свободы. По выборке находится наблюдаемое значение $t_{набл}$ статистики. Если $|t_{набл}| > t_{\alpha, \nu}$, то коэффициент корреляции значим ($t_{\alpha, \nu} = t_{кр}$ – критическая точка распределения Стьюдента, зависящая только от объема выборки).

Качественная оценка тесноты связи между величинами выявляется по шкале Чеддока (таблица 2).

Таблица 2 – Шкала Чеддока

Теснота связи	Значение коэффициента корреляции при наличии	
	прямой связи	обратной связи
Слабая	0,1–0,3	(–0,1)–(–0,3)
Умеренная	0,3–0,5	(–0,3)–(–0,5)

Заметная	0,5–0,7	(–0,5)–(–0,7)
Высокая	0,7–0,9	(–0,7)–(–0,9)
Весьма высокая	0,9–0,99	(–0,9)–(–0,99)

В Excel с помощью пакета «Анализ данных» на листе «Исходные данные» получены ковариационная и корреляционная матрицы (таблицы 3 и 4).

Таблица 3 – Ковариационная матрица

Ковариация		
	<i>ВВП</i>	<i>Экспорт</i>
ВВП	98 568,75	
Экспорт	25 558,75	6 804,75

Таблица 4 – Корреляционная матрица

Корреляция		
	<i>ВВП</i>	<i>Экспорт</i>
ВВП	1	
Экспорт	0,99	1

Ковариация между ВВП и экспортом равна $25\,558,75 > 0$, поэтому зависимость *прямая*.

Коэффициент корреляции между ВВП и экспортом равен $0,99 > 0$, поэтому зависимость *прямая* и *весьма высокая*.

Проверим на значимость коэффициент корреляции, так как он найден по выборочной совокупности только из 20 наблюдений, что может привести к неверным выводам о всей генеральной совокупности таких факторов, как ВВП и экспорт. На листе «Исходные данные» получены наблюдаемое и критическое значения *t*-статистики (таблица 5).

Таблица 5 – Значимость коэффициента корреляции

тнабл	25,93
ткр	2,10

Так как $|тнабл| = 25,93 > ткр = 2,1$, то коэффициент корреляции *значим* (значительно отличается от нуля).

Исходя из проведенного анализа можно выдвинуть предположение

о том, что зависимость экспорта от ВВП описывается линейной регрессионной моделью $y = b_0 + b_1x_1 + \varepsilon$, где b_0 , b_1 – неизвестные параметры модели; ε – случайная переменная, которая включает в себя суммарное влияние всех неучтенных в модели факторов, ошибки измерений.

3. Параметризация модели

Статистической оценкой параметра называется его приближенное значение, полученное на основе выборочных данных. Для получения точечных оценок параметров уравнения парной линейной регрессии применяют *метод наименьших квадратов* (МНК). В соответствии с МНК минимизируется сумма квадратов разностей между фактическими и расчетными значениями зависимой переменной. Оценки неизвестных параметров находятся из системы нормальных уравнений, полученной методом дифференциального исчисления.

Для расчета интервальных оценок (доверительных интервалов) параметров регрессии определяются предельные ошибки Δ для каждого показателя: $\Delta_{b_0} = t_{\alpha, \nu} S_{b_0}$, $\Delta_{b_1} = t_{\alpha, \nu} S_{b_1}$,

где $S_{b_0}^2 = \frac{\bar{x}^2 S^2}{n \sigma^2(x)}$, $S_{b_1} = \frac{S^2}{n \sigma^2(x)}$ – стандартные ошибки коэф-

фициентов регрессии; $S^2 = \frac{1}{n-2} \sum e_i^2 = \frac{1}{n-2} \sum (y_i - \tilde{y}_i)^2$ –

стандартная ошибка регрессии, которая служит мерой разброса зависимой переменной вокруг линии регрессии; $t_{\alpha, \nu}$ – критическая точка распределения Стьюдента для заданного уровня значимости α и числа степеней свободы $\nu = n - 2$. Доверительные интервалы имеют вид: для b_0 – $(b_0 - \Delta_{b_0}; b_0 + \Delta_{b_0})$, b_1 – $(b_1 - \Delta_{b_1}; b_1 + \Delta_{b_1})$, где центр интервала равен точечной оценке, концы интервалов получены прибавлением и вычитанием произведения стандартной ошибки коэффициента на критическое значение t -статистики.

Доверительный интервал с вероятностью 0,95 содержит истинное значение свободного члена уравнения регрессии. Поэтому любое значение из этого интервала может служить оценкой параметра. Если в границы доверительного интервала попадает ноль, т. е. нижняя граница отрицательна, а верхняя – положительна, то оцениваемый параметр принимается нулевым, так как он не может одновременно принимать и положительное, и отрицательное значения.

В результате проведения регрессионного анализа на листе «Регрессия» в третьей таблице получены точечные и интервальные оцен-

ки неизвестных параметров (таблица 6).

Таблица 6 – Статистика коэффициентов регрессии

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%
Y-пересечение	-73,25	16,03	-4,57	0,000 24	-106,93	-39,56
ВВП	0,26	0,01	25,93	0,000 00	0,24	0,28
		tкр	2,10			

Точечная оценка параметра b_0 равна $-73,25$. Интервальная оценка равна $(-106,93; -39,56)$. Доверительный интервал с вероятностью 0,95 содержит истинное значение свободного члена уравнения регрессии. Поэтому любое значение из этого интервала может служить оценкой параметра b_0 .

Точечная оценка параметра b_1 равна 0,24. Интервальная оценка равна $(0,24; 0,28)$. Доверительный интервал с вероятностью 0,95 содержит истинное значение коэффициента при переменной x уравнения регрессии. Поэтому любое значение из этого интервала может служить оценкой параметра b_1 .

Таким образом, уравнение регрессии имеет вид

$$y = -73,25 + 0,24x.$$

Случайная переменная отсутствует в уравнении, так как коэффициенты регрессии имеют случайный характер, т. е. неучтенные факторы повлияли на их значение при применении МНК.

4. Верификация модели

4.1. Общее качество уравнения

Оценим общее качество модели.

Качество построенной модели в целом оценивает коэффициент (индекс) детерминации, который представляет собой долю дисперсии, объясненной выбранным фактором. Коэффициент детерминации R^2 ($0 \leq R^2 \leq 1$) рассчитывается как квадрат коэффициента корреляции для парной регрессии. В Excel множественный R для парной регрессии равен коэффициенту корреляции. *Скорректированный (нормированный)*

индекс детерминации \bar{R}^2 позволяет более точно определить качество модели. Если $R^2 = 1$, то все точки наблюдения лежат на регрессионной прямой. Для определения статистической значимости коэффициента детерминации используется F -статистика, рассчитываемая по формуле: $F = \frac{R^2(n-2)}{1-R^2}$. Величина F имеет распределение Фишера с $\nu_1 = 1$, $\nu_2 = n - 2$ степенями свободы. По выборке находится наблюдаемое значение $F_{набл}$ статистики. Если $F_{набл} > F_{\alpha, \nu_1, \nu_2}$, то R^2 значим ($F_{\alpha, \nu_1, \nu_2} = F_{кр}$ – критическая точка распределения Фишера, зависящая только от объема выборки).

Указанные показатели находятся в таблице «Регрессионная статистика» листа «Регрессия» (таблица 7).

Таблица 7 – Регрессионная статистика

ВЫВОД ИТОГОВ	
<i>Регрессионная статистика</i>	
Множественный R	0,99
R-квадрат	0,97
Нормированный R-квадрат	0,97
Стандартная ошибка	14,04
Наблюдения	20

Коэффициент множественной детерминации R-квадрат равен 0,97. Так как коэффициент множественной детерминации близок к 1, то уравнение имеет высокое качество. Этот факт подтверждает также нормированный индекс множественной детерминации, равный 0,97.

Значимость коэффициента множественной детерминации R-квадрат устанавливается с помощью критерия Фишера в таблице «Дисперсионный анализ» листа «Регрессия» (таблица 8).

Таблица 8 – Дисперсионный анализ

Дисперсионный анализ					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
Регрессия	1	132 547	132 547,02	672,45	1,04E-15
Остаток	18	3 547,98	197,11		
Итого	19	136 095			
			F _{кр}	4,41	

Так как наблюдаемое значение $F_{\text{набл}} = 672,45 > F_{\text{кр}} = 4,41$, то R -квадрат значим (значительно отличается от нуля), что еще раз подтверждает высокое качество построенного уравнения регрессии.

4.2. Нормальность распределения остатков

Проанализируем нормальность распределения остатков для возможности использования критерия Стьюдента при проверке статистических гипотез. Сделать вывод о нормальности распределения остатков можно:

- 1) по гистограмме остатков;
- 2) по числовым характеристикам асимметрии и эксцессу;
- 3) по критерию Пирсона.

Остатки (отклонения наблюдаемых значений от теоретических) являются оценками случайного члена ε уравнения регрессии.

Анализируя качество модели, необходимо проверить ряд статистических гипотез, использующих критерий Стьюдента, которым можно воспользоваться в случае, когда остатки распределены по нормальному закону.

Кривая плотности нормального распределения задается

функцией $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$, где a – математическое ожидание; σ – среднее квадратическое отклонение. Например, при $a = 0$ и $\sigma = 1$ она имеет вид, приведенный на рисунке 2.

Визуально нормальность распределения остатков можно определить, сравнивая кривую плотности нормального распределения с гистограммой частот (частостей) остатков, т. е. со ступенчатой фигурой, состоящей из прямоугольников, основаниями которых служат интервалы одинаковой длины на оси Ox , а высоты равны сумме частот значений остатков, попадающих в интервал. Если линия, соединяющая середины верхних

сторон прямоугольников, близка к кривой плотности нормального распределения, то предполагают, что распределение остатков приближено к нормальному.

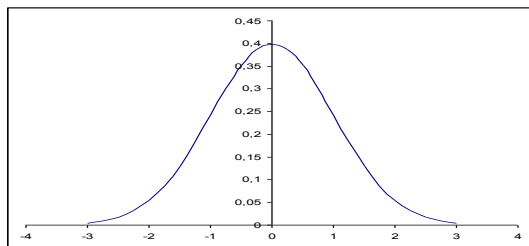


Рисунок 2 – Кривая нормального распределения

Асимметрия и эксцесс как числовые характеристики нормально распределенной случайной величины равны 0. При асимметричном распределении вершина кривой сдвинута относительно ординаты выборочной средней. Если асимметрия больше 0, то вершина сдвинута вправо (положительная асимметрия), если меньше 0, то – влево (отрицательная асимметрия) (рисунок 3).

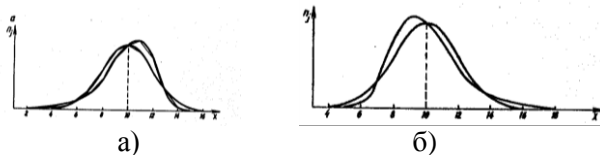


Рисунок 3 – Правосторонняя (а) и левосторонняя (б) асимметрии распределения

Эксцесс характеризует относительную остроконечность или сглаженность распределения по сравнению с нормальным распределением. Положительный эксцесс обозначает относительно остроконечное распределение, отрицательный эксцесс – относительно сглаженное распределение (рисунок 4).

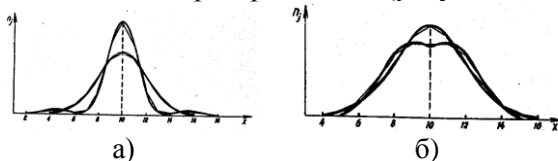


Рисунок 4 – Положительный (а) и отрицательный (б) эксцессы распределения

Оценки вида кривой Гаусса и значений асимметрии и эксцесса являются качественными характеристиками распределе-

ния. Для надежности вывода (с вероятностью 0,95) проверяется статистическая гипотеза о нормальности распределения с помощью критерия согласия Пирсона. Выдвигается гипотеза о нормальном законе распределения остатков. Для проверки данной гипотезы используется статистика $\chi^2 = \sum_{i=1}^k \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$, имеющая распределение χ^2 с $(k - r - 1)$ степенями свободы, где r – число параметров распределения $F(x)$, которые оцениваются по выборке, n – объем выборки; k – число непересекающихся интервалов выборочных значений, n_i – число значений выборки, принадлежащих i -му интервалу, $i = 0, 1, \dots, k - 1$; p_i – вероятности попадания значений случайной величины в каждый из этих интервалов. По выборке вычисляется наблюдаемое значение статистики $\chi_{набл}^2$. Для выбранного уровня значимости α по распределению χ^2 находится число $\chi_{кр}^2 = \chi^2(\alpha; k - r - 1)$. Гипотеза о нормальном распределении случайного члена принимается на заданном уровне значимости, если $\chi_{набл}^2 < \chi_{кр}^2$. Если же $\chi_{набл}^2 \geq \chi_{кр}^2$, то гипотеза отвергается.

1. Соединим середины верхних сторон прямоугольников гистограммы.

Так как ломаная линия (рисунок 5) близка к кривой нормального распределения, заданной уравнением $f(x) = \frac{1}{13,67\sqrt{2\pi}} e^{-\frac{(x-0)^2}{2 \cdot 13,67^2}}$, где

$a = 0$ – математическое ожидание; $\sigma = 13,67$ – среднее квадратическое отклонение, то остатки распределены по нормальному закону. Следовательно, по визуальному анализу гистограммы можно сделать вывод о выполнении условия нормальности распределения остатков.

2. Так как асимметричность равна $-0,3$, то наблюдается небольшая левосторонняя асимметричность эмпирической кривой относительно теоретической.

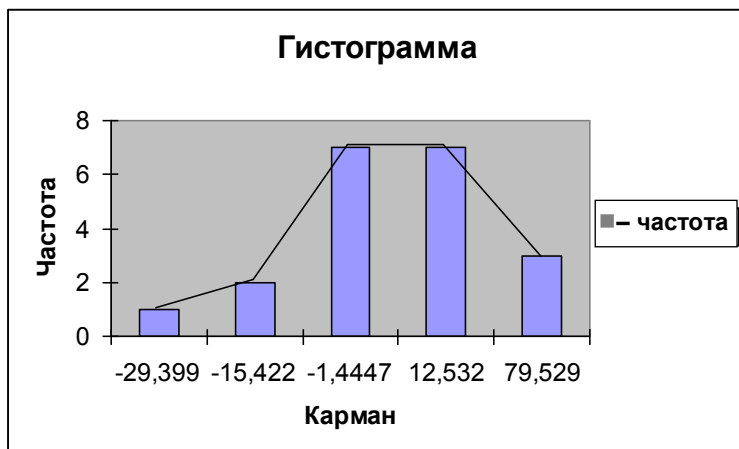


Рисунок 5 – Гистограмма остатков

Эксцесс равен 0,08, поэтому наблюдается небольшая «островершинность» эмпирической кривой. Так как характеристики плотности распределения незначительно отличаются от нуля, то можно считать распределение нормальным.

3. Докажем нормальность распределения остатков с помощью критерия Пирсона.

На листе «Условие 1 и нормальность» найдены наблюдаемое и критическое значения статистики хи-квадрат. Так как хи-квадрат наблюдаемое, равное 1,67, меньше хи-квадрат критического, равного 7,81, то остатки распределены по нормальному закону, несмотря на незначительные отклонения эмпирического распределения остатков от нормального.

4.3. Значимость коэффициентов регрессии

Проверим значимость коэффициентов регрессии.

Проверка значимости коэффициентов регрессии осуществляется с помощью *t-статистики Стьюдента*. Вычисляются наблюдаемые значения $t_{b_0} = \frac{b_0}{S_{b_0}}$ и $t_{b_1} = \frac{b_1}{S_{b_1}}$. По таблице критических точек распределения Стьюдента по заданному уровню значимости α и числу степеней свободы $\nu = n - 2$ находится

критическая точка $t_{\alpha, v} = t_{кр}$. Если $|t_{b_0}| > t_{\alpha, v}$, то коэффициент регрессии b_0 значим. Если $|t_{b_1}| > t_{\alpha, v}$, то коэффициент регрессии b_1 значим. Значимость коэффициентов подтверждает правильность выбора модели на этапе спецификации. Если хотя бы один из коэффициентов не значим, то необходимо вернуться на этап спецификации.

Значимость коэффициента регрессии b_0 оценивается с помощью t -статистики. Наблюдаемое значение статистики (таблица 6) $t_{набл}$ равно $-4,57$ (оно равно отношению точечной оценки коэффициента b_0 к его стандартной ошибке). Критическое значение $t_{кр}$ равно $2,1$. Так как $|t_{набл}| = 4,57 > t_{кр} = 2,1$, то коэффициент b_0 значим.

Для коэффициента регрессии b_1 наблюдаемое значение статистики $t_{набл}$ равно $25,93$ (оно равно отношению точечной оценки коэффициента b_1 к его стандартной ошибке). Критическое значение зависит только от объема выборки, поэтому $t_{кр}$ равно $2,1$ также как для b_0 . Так как $|t_{набл}| = 25,93 > t_{кр} = 2,1$, то коэффициент b_1 значим.

Значимость коэффициентов регрессии подтверждает выдвинутое на этапе спецификации предположение о линейной форме зависимости таких факторов, как ВВП и экспорт.

4.4. Проверка статистических свойств остатков (качества оценок коэффициентов регрессии)

По статистическим данным нельзя получить точные оценки неизвестных коэффициентов регрессии, но можно найти их приближенные оценки, качество которых характеризуется следующими тремя свойствами:

- **несмещенность** (каждый параметр регрессии можно рассматривать как среднее значение из возможного большого количества оценок);
- **эффективность** (оценки параметров характеризуются наименьшей дисперсией); в практических исследованиях это означает возможность перехода от точечного оценивания к интервальному;
- **состоятельность** (характеризует увеличение точности оценок с увеличением объема выборки).

Метод наименьших квадратов обеспечивает указанные свойства оценкам параметров регрессии при выполнении следующих предположений:

1) математическое ожидание случайной переменной равно нулю;

2) дисперсии наблюдений случайной переменной одинаковы (это условие называется *условием гомоскедастичности*);

3) независимость значений случайной переменной в любом наблюдении от ее значений во всех других наблюдениях.

Предположения 1–3 называются *условиями Гаусса–Маркова*.

Теорема Гаусса–Маркова. *Если условия 1–3 выполнены, то оценки коэффициентов парной линейной регрессии, полученные с помощью МНК, являются несмещенными, состоятельными и эффективными.*

4.4.1. Центрированность остатков

Проверим выполнение условия 1: математическое ожидание случайной переменной равно нулю, т. е. $M(\varepsilon) = 0$.

Оценкой математического ожидания случайной переменной $M(\varepsilon)$ является *среднее остатков* $\bar{\varepsilon}$. Гипотеза о значимости математического ожидания случайной переменной проверяется с помощью t -статистики, наблюдаемое значение которой определяется равенством $t = \frac{(\varepsilon - 0)\sqrt{n}}{S}$.

По распределению Стьюдента по заданному уровню значимости α и числу степеней свободы $\nu = n - 1$ находится критическая точка $t_{\alpha, \nu} = t_{кр}$. Если $t_{набл} > t_{кр}$, то $M(\varepsilon)$ значимо. Если $t_{набл} < t_{кр}$, то $M(\varepsilon)$ незначимо.

Если условие 1 нарушено, то оценки коэффициентов регрессии могут быть смещенными. Поэтому для устранения этих проблем в модель регрессии следует включать свободный член.

Среднее остатков рассчитано на листе «Условие 1 и нормальность» в таблице «Остатки» (таблица 9).

Среднее равно $-1,7E-14 = -1,7 \cdot 10^{-14}$. Оно достаточно близко к нулю, поэтому можно предположить, что математическое ожидание случайной переменной равно нулю, т. е. выполняется условие 1. Проверим гипотезу о равенстве нулю математического ожидания случайной переменной.

Таблица 9 – Числовые характеристики остатков

<i>Остатки</i>	
Среднее	-1,7E-14
Стандартная ошибка	3,02
Медиана	-1,46
Мода	#Н/Д
Стандартное отклонение	13,67
Дисперсия выборки	186,74
Экссесс	0,08
Асимметричность	-0,3
Интервал	55,91
Минимум	-29,4
Максимум	26,51
Сумма	-3,4E-13
Счет	20
Уровень надежности (95%)	6,4

На листе «Условие 1 и нормальность» рассчитаны наблюдаемое и критическое значения статистики (таблица 10).

Таблица 10 – Значимость математического ожидания случайного члена

тнабл	-5,58E-15
tkp	2,09

Так как $|тнабл| = 5,58E-15 = 5,58 \cdot 10^{-15} < tkp = 2,09$, то среднее не значимо (незначительно отличается от нуля). Следовательно, условие 1 Гаусса–Маркова выполняется.

4.4.2. Гомоскедастичность (гетероскедастичность) остатков

Проверим выполнение условия 2 Гаусса–Маркова.

Предположение о постоянстве и конечности дисперсии остатков называется *свойством гомоскедастичности остатков*. Если оно не выполняется, то такое явление называется *гетероскедастичностью*. Гетероскедастичность часто вызывается ошибками спецификации, когда не учитывается в модели существенная переменная. Гетероскедастичность приводит к

тому, что оценки коэффициентов регрессии не являются эффективными, увеличиваются дисперсии распределений оценок коэффициентов, появляется вероятность неверного вычисления оценок стандартных ошибок коэффициентов регрессии. В результате можно сделать неверный вывод о значимости коэффициента. Для оценки нарушения гомоскедастичности наиболее часто используются графический анализ отклонений, тест ранговой корреляции Спирмена и тест Голдфелда–Квандта.

При применении теста Спирмена предполагается, что абсолютные величины остатков и значения объясняющей переменной коррелированы. Эту корреляцию можно измерять с помощью коэффициента ранговой корреляции Спирмена:

$$r = 1 - \frac{6 \sum D_i^2}{n(n^2 - 1)},$$

где D_i – разность между рангом x_i и рангом модуля остатка $e_i = y_i - \tilde{y}_i$.

Тест проводится по следующей схеме:

1. Строится линейная модель регрессии.
2. Определяются ранги значений x_i независимой переменной и соответствующие ранги модулей остатков $|e_i|$ (ранг – это порядковый номер значения в ранжированном ряду).
3. Находится коэффициент ранговой корреляции Спирмена.
4. Осуществляется проверка гипотезы об отсутствии гетероскедастичности с помощью t -статистики, наблюдаемое значение которой определяется равенством $t_{набл} = r\sqrt{n-1}$. Если $t_{набл} > t_{кр}$, то гетероскедастичность присутствует, значит, МНК-оценки неэффективны.

Так как $M(\varepsilon)$ равно нулю, то МНК-оценки параметров являются несмещенными и состоятельными, поэтому их позволено использовать, например, для точечного прогнозирования даже в случае гетероскедастичности. Однако в этом случае МНК-оценки не являются эффективными, а следовательно, результаты (доверительные интервалы для коэффициентов и прогнозных значений), основанные на анализе дисперсии, неверны.

Существуют два подхода к решению проблемы гетероскедастичности:

- 1) преобразование выборочных данных;

2) применение взвешенного и обобщенного МНК (ОМНК).

Первый подход предполагает такое преобразование исходных данных, чтобы для них модель уже обладала свойством гомоскедастичности. Используют такие преобразования, как логарифмирование данных, переход к безразмерным величинам путем деления на некоторые известные величины той же размерности, что и исходные данные, стандартизация исходных данных.

Второй подход устранения гетероскедастичности состоит в построении моделей, учитывающих гетероскедастичность ошибок наблюдений.

На листе «Условие 2» рассчитаны наблюдаемое и критическое значения t -статистики (таблица 11).

Таблица 11 – Проверка гипотезы об отсутствии гетероскедастичности

$t_{\text{набл}}$	-0,74
$t_{\text{кр}}$	2,10

Так как $|t_{\text{набл}}| = 0,74 > t_{\text{кр}} = 2,10$, то гетероскедастичность отсутствует. Следовательно, условие 2 Гаусса–Маркова выполняется, значит, МНК-оценки параметров регрессии будут эффективными.

4.4.3. Автокорреляция остатков

Проверим выполнение условия 3 Гаусса–Маркова.

Условие 3 Гаусса–Маркова требует *независимости* значений случайной переменной в любом наблюдении от ее значений во всех других наблюдениях. Если данное условие не выполняется, то говорят, что случайная переменная подвержена *автокорреляции*. В этом случае коэффициенты регрессии, получаемые по МНК, оказываются *неэффективными*, хотя и *несмещенными*, а их стандартные ошибки рассчитываются некорректно (занижаются).

Существуют несколько методов определения автокорреляции остатков, два из которых приведены ниже.

Первый метод – это построение графика зависимостей остатков от номера наблюдений и визуальное определение наличия автокорреляции остатков (рисунок 6).

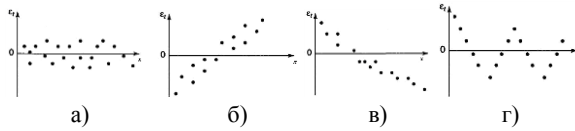


Рисунок 6 – Модели зависимости остатков от номера наблюдения: случайные остатки (а); наличие зависимости в остатках (б, в, г)

Второй метод – проверка гипотезы об отсутствии автокорреляции остатков с помощью критерия Дарбина–Уотсона, т. е.

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

наблюдаемое значение которого рассчитывается как отношение суммы квадратов разностей последовательных значений остатков к сумме квадратов остатков.

Это же значение критерия может быть вычислено по формуле

$$d = 2(1 - r_1^{\varepsilon}),$$

где $r_1^{\varepsilon} = \frac{\sum_{t=2}^n (\varepsilon_t - \bar{\varepsilon}_1)(\varepsilon_{t-1} - \bar{\varepsilon}_2)}{\sqrt{\sum_{t=2}^n (\varepsilon_t - \bar{\varepsilon}_1)^2 \sum_{t=2}^n (\varepsilon_{t-1} - \bar{\varepsilon}_2)^2}}$ – коэффициент авто-

корреляции первого порядка, $\bar{\varepsilon}_1 = \frac{\sum_{t=2}^n \varepsilon_t}{n-1}$, $\bar{\varepsilon}_2 = \frac{\sum_{t=2}^n \varepsilon_{t-1}}{n-1}$.

Альтернативные гипотезы – гипотезы о наличии положительной или отрицательной автокорреляции в остатках.

Если автокорреляция *отсутствует*, то $r_1^{\varepsilon} = 0$ и значение статистики $d \approx 2$. При *положительной* автокорреляции $r_1^{\varepsilon} > 0$ и $0 \leq d < 2$, а при *отрицательной* – $r_1^{\varepsilon} < 0$ и $2 < d \leq 4$. Следовательно, $0 \leq d \leq 4$.

Однако существуют области неопределенности, связанные с тем, что распределение статистики Дарбина–Уотсона зависит не только от числа наблюдений и числа объясняющих переменных, но и от значений объясняющих переменных. В этом случае используются другие способы проверки (например, визуальный и др.).

По таблице, фрагмент которой приведен в таблице 12, определяются критические значения критерия Дарбина–Уотсона d_1 и d_2 для заданного числа наблюдений n , числа объясняющих переменных k и заданного уровня значимости 0,05.

Таблица 12 – Статистика Дарбина–Уотсона: d_1 и d_2 при уровне значимости 5% (фрагмент таблицы)

n	$k = 1$		$k = 2$		$k = 3$		$k = 4$	
	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83

По этим значениям отрезок $[0; 4]$ разбивается на пять областей (рисунок 7).

Положительная автокорреляция	Область неопределенности	Автокорреляция отсутствует	Область неопределенности	Отрицательная автокорреляция
0	d_1	d_2	$4 - d_2$	$4 - d_1$
				4

Рисунок 7 – Критические области статистики Дарбина–Уотсона

В зависимости от того, в какую область попадает наблюдаемое значение критерия, принимают или отвергают гипотезу.

На листе «Регрессия» в ячейках D23:E47 получены данные для проверки условия 3. В ячейке E47 рассчитано наблюдаемое значение d -статистики: $d_{\text{набл}} = 1,26$. По таблице критических значений d -статистики для числа наблюдений 20, числа объясняющих переменных 1 и заданного уровня значимости 0,05 значения $d_1 = 1,20$ и $d_2 = 1,41$ (таблица 12), которые отрезок $[0; 4]$ разбивает на пять областей (рисунок 8).

Положительная автокорреляция	Область неопределенности	Автокорреляция отсутствует	Область неопределенности	Отрицательная автокорреляция	
0	1,20	1,41	2,59	2,80	4

Рисунок 8 – Критические области статистики Дарбина–Уотсона

Так как $1,20 < \text{днабл} = 1,26 < 1,41$, т. е. наблюдаемое значение попало в область неопределенности, то ничего нельзя сказать о наличии автокорреляции, используя критерий Дарбина–Уотсона. В этом случае воспользуемся визуальным способом.

Визуально наличие автокорреляции остатков можно определить по графику зависимости остатков от номера наблюдения. Так как каждому номеру наблюдения соответствует определенное значение независимого фактора, то можно анализировать зависимость остатков от этого фактора. Эта зависимость представлена на графике «ВВП График остатков» листа «Регрессия» (рисунок 9).

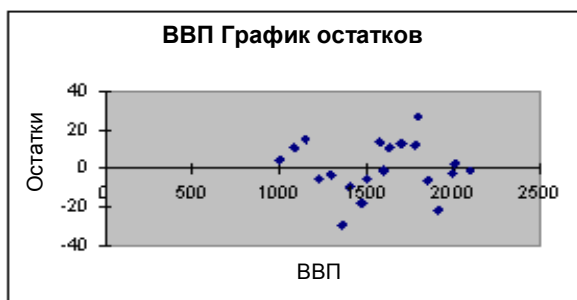


Рисунок 9 – График остатков

Так как на графике остатков (рисунок 9) точки разбросаны вдоль оси Ox хаотично без видимой закономерности (сравните, например, с рисунком 6), то зависимости между остатками не наблюдается, ковариация остатков равна нулю. Поэтому условие 3 выполняется.

4.5. Анализ свойств модели: средний коэффициент эластичности

Средний коэффициент эластичности показывает, на сколько процентов в среднем изменяется зависимая переменная при изменении на 1% независимого фактора от своей средней величины.

На листе «Исходные данные» рассчитан средний коэффициент эластичности, равный 0,06. Таким образом, изменение ВВП на 1% от его среднего значения приведет к изменению экспорта от его среднего на 0,06%.

Общий вывод по результатам этапа верификации: так как выполняются все условия верификации, то модель является качественной. Таким образом, прогноз, выполненный по ней, является качественным: несмещенным, состоятельным и эффективным.

5. Прогнозирование

Если выполняются все условия верификации, то модель является качественной. В противном случае ее надо усовершенствовать: уточнить либо спецификацию, либо выборку.

Под *прогнозированием* в эконометрике понимается нахождение оценки зависимой переменной для любого допустимого значения независимой переменной. Различают *точечное* и *интервальное* прогнозирование.

Точечная оценка y^* зависимой переменной определяется путем подстановки в найденное уравнение регрессии заданного допустимого значения независимой переменной x^* .

Интервальный прогноз, или доверительный интервал прогноза, имеет вид

$$(y^* - t_{\alpha, v} S^*, y^* + t_{\alpha, v} S^*),$$

где $t_{\alpha, v}$ – критическое значение t -статистики при заданном уровне значимости α и числе степеней свободы v ;

$$S^* = S \cdot \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x - \bar{x})^2}} = S \cdot \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{(n-1) \cdot \sigma_x^2}} \quad \text{— средняя}$$

стандартная ошибка прогноза; S – стандартная ошибка регрессии; σ_x^2 – исправленная дисперсия независимой переменной x .

Так как модель является качественной, то прогноз, выполненный по ней, является качественным: несмещенным, состоятельным и эффективным. На листе «Регрессия» для заданного значения ВВП, равного 2 500, рассчитан точечный прогноз экспорта, который равен 575, и интервальный прогноз, равный (539,04; 610,96). Этот доверительный интервал значений экспорта содержит с вероятностью 0,95 истинное значение экспорта при заданном значении ВВП.

Вопросы для самоконтроля

1. Каковы этапы построения эконометрической модели?
2. Характеристика цели каждого этапа.
3. Знания из каких научных дисциплин необходимы на каждом из этапов эконометрического моделирования?
4. В чем состоит спецификация модели парной регрессии?
5. Почему в уравнении регрессии присутствует случайная переменная?
6. Как определить силу и направленность взаимодействия факторов?
7. Что означает положительная ковариация между исследуемыми факторами?
8. Что означает отрицательная корреляция между исследуемыми факторами?
9. Как может быть представлено графически взаимодействие факторов?
10. Что означает значимость коэффициента корреляции?
11. Как проверить на значимость коэффициент корреляции?
12. В чем суть МНК для нахождения оценок параметров регрессии?
13. Почему с помощью МНК находят оценки параметров, а не их точные значения?
14. Какая оценка параметра называется точечной?
15. В чем суть интервальной оценки параметров?
16. Как найти интервальные оценки коэффициентов регрессии?
17. Как используются стандартные ошибки регрессии и стандартные ошибки коэффициентов регрессии при анализе оценок параметров парной регрессии?
18. В чем экономический смысл параметров модели парной регрессии?
19. Как можно оценить общее качество уравнения регрессии?
20. Какова суть коэффициента детерминации, нормированного коэффициента детерминации? В каких пределах они изменяются?
21. Какова связь между коэффициентом детерминации, коэффициентом корреляции и множественным коэффициентом корреляции для парной регрессии?
22. Для какой цели в парной регрессии используется критерий Фишера?
23. Как получить остатки для модели парной регрессии?
24. Какому условию должны удовлетворять остатки, чтобы для проверки статистических гипотез можно было использовать критерий Стьюдента?

25. Какое распределение называется нормальным? Каковы его параметры?
26. Какими способами можно проверить нормальность распределения остатков?
27. Каковы значения асимметрии и эксцесса для нормально распределенной случайной величины?
28. Как используется критерий согласия Пирсона для проверки гипотезы о нормальном законе распределения остатков?
29. Для чего проверяется значимость коэффициентов регрессии?
30. В чем суть статистической значимости коэффициентов регрессии?
31. Какими свойствами должны обладать оценки параметров регрессии?
32. Каковы основные предпосылки применения МНК для построения регрессионной модели?
33. Каковы последствия невыполнимости предпосылок применения МНК?
34. Как сформулировать теорему Гаусса–Маркова?
35. Как проверить центрированность остатков (условие 1) теоремы Гаусса–Маркова?
36. В чем суть гетероскедастичности остатков?
37. Каковы причины гетероскедастичности остатков?
38. Каковы последствия гетероскедастичности остатков?
39. Как определить гомоскедастичность остатков визуально?
40. Для чего и как используется тест ранговой корреляции Спирмена?
41. В чем суть автокорреляции остатков? Каковы ее последствия?
42. Методы определения автокорреляции остатков?
43. Как проверить гипотезу об отсутствии автокорреляции остатков с помощью критерия Дарбина–Уотсона?
44. Как проверить гипотезу об отсутствии автокорреляции остатков визуально?
45. Что характеризует коэффициент эластичности?
46. Что понимают под прогнозированием?
47. Как получить точечный прогноз зависимого фактора?
48. Как получить интервальный прогноз зависимого фактора?

Индивидуальные задания

Исследовать зависимость объема экспорта (y , усл. ед.) от ВВП (x , усл. ед.), используя данные 20 наблюдений, приведенные в таблице 1, которые необходимо изменить следующим образом: к каждому значению ВВП таблицы 1 надо прибавить $12 \cdot k$, где k – номер студента в журнале группы. Построить регрессионную модель $y = f(x) + \varepsilon$. Спрогнозировать объем экспорта при ВВП, равном 1 970 усл. ед. Оформить отчет (приложение Б).

Тема 2. МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

Постановка задачи

Исследовать зависимость заработной платы (y , тыс. р.) от возраста (x_1 , лет) и стажа по данной специальности (x_2 , лет), используя данные наблюдений, приведенные в таблице 13. Построить регрессионную модель $y = f(x_1, x_2) + \varepsilon$. Рассчитать значение заработной платы (ЗП) для работника в возрасте 35 лет со стажем работы по данной специальности 10 лет.

Таблица 13 – Данные наблюдений

ЗП	Возраст	Стаж
684,48	28	10
674,45	26	8
729,62	30	14
748,86	35	15
761,44	41	16
773,42	45	17
628,07	27	3
731,84	35	13
698,81	30	10
645,92	23	5
664,64	29	7
711,18	33	11
798,07	40	20
833,82	41	24

ЗП	Возраст	Стаж
607,61	23	2
711,76	32	12
728,62	37	13
666,15	31	7
683,70	30	9

Технология вычислений в MS Excel для построения и анализа линейной множественной регрессии

Примечание – Вычисления на каждом этапе эконометрического моделирования линейной множественной регрессии аналогичны проведенным в теме 1, поэтому суть каждого этапа не приводится.

1. Постановочный этап

Введите исходные данные (таблица 13). Для этого в ячейку A1 введите название первого столбца – «ЗП», в ячейку B1 – название второго столбца – «Возраст», в ячейку C1 – название третьего столбца – «Стаж». В ячейки A2, A3,..., A21 введите данные первого столбца таблицы 13, в ячейки B2, B3,..., B21 – данные второго столбца, в ячейки C2, C3,..., C21 – данные третьего столбца.

Введите новое название листа «Исходные данные». Сохраните рабочую книгу (*Файл* → *Сохранить как* → ...).

Примечание – В приложении В приведен пример оформления вычислений на листах MS Excel.

2. Спецификация

На листе «Исходные данные» выберите меню *Сервис* → *Анализ данных* → *Корреляция* → *ОК*. Значения параметров окна установите следующим образом: *Входной интервал* – введите ссылки на ячейки A1:C21, *Метки* – установите флажок, *Параметры вывода* – установите переключатель на *Выходной интервал* и в поле установите ссылку на ячейку E2. Нажмите кнопку *ОК*. Скопируйте G5 в H4. В ячейку E1 введите название «Корреляционная матрица».

В ячейку E6 введите название «Значимость коэффициентов корреляции».

В ячейку E7 введите обозначение $t_{ЗП,В}$ набл.

В ячейку F7 введите формулу для вычисления t набл для коэффи-

циента корреляции таких факторов, как ЗП и возраст $= F4 * \text{КОРЕНЬ}((20 - 2)/(1 - F4^2))$, где 20 – число наблюдений, 2 – число факторов.

В ячейку E8 введите $t_{ЗП,С}$ набл.

В ячейку F8 введите формулу для вычисления $t_{набл}$ для коэффициента корреляции факторов ЗП и стаж $= F5 * \text{КОРЕНЬ}((20 - 2)/(1 - F5^2))$, где 20 – число наблюдений, 2 – число факторов.

В ячейку E9 введите $t_{кр}$.

В ячейке F9 вычислите критическое значение $t_{кр}$ следующим образом:

- нажмите на f_x (вставка функций);
- в поле *Категория* окна *Мастер функций* выберите *статистические*, из предложенных ниже функций выделите *СТЮДРАСПОБР* и нажмите кнопку *ОК*. Откроется окно *Аргументы функций*. Заполните поля:

– *Вероятность* – наберите значение 0,05;

– *Степени свободы* – введите $20 - 2$, где 20 – число наблюдений, 2 – число факторов. Нажмите кнопку *ОК*.

3. Параметризация модели

Примечание – Предполагаем, что между факторами существует линейная зависимость. Далее находится уравнение линейной регрессии. Если доказана нелинейность зависимости, то проводится процедура линеаризации [6].

Выберите в меню *Сервис* → *Анализ данных* → *Регрессия* → *ОК*. Значения параметров в окне установите следующим образом:

- *Входной интервал Y* – введите ссылки на ячейки A1:A21.
- *Входной интервал X* – введите ссылки на ячейки B1:C21.
- *Метки* – установите флажок.
- *Уровень надежности* – установите флажок.
- *Константа ноль* – оставьте пустым.
- *Параметры вывода* – установите переключатель на *Новый рабочий лист* и в соответствующее поле введите его название «Регрессия».
- *Остатки* – установите флажок.
- *Стандартизированные остатки* – оставьте пустым.
- *График остатков* – установите флажок.
- *График подбора* – установите флажок.
- *График нормальной вероятности* – оставьте пустым. Нажмите кнопку *ОК*.

Расположите диаграммы рядом (на поле диаграммы нажмите левую кнопку мышки, затем поместите курсор на белое поле и при

нажатой левой кнопке передвигайте диаграмму вниз) и растяните (на поле диаграммы нажмите левую кнопку мышки, нижнюю линию границы диаграммы при нажатой левой клавише протяните вниз).

4. Верификация модели

4.1. Общее качество уравнения

В ячейку D15 листа «Регрессия» введите $F_{кр}$. Вычислите критическое значение в свободной ячейке E15 следующим образом:

- нажмите на f_x (вставка функций);
- в поле *Категория* окна *Мастер функций* выберите *статистические*, из предложенных ниже функций выделите *FRASПЮБР* и нажмите кнопку *OK*. Откроется окно *Аргументы функций*. Заполните поля:
 - *Вероятность* – наберите значение 0,05;
 - *Степени свободы 1* – установите курсор в поле и выделите ячейку B12 столбца *df* таблицы «Дисперсионный анализ»;
 - *Степени свободы 2* – установите курсор в поле и выделите ячейку B13 столбца *df* таблицы «Дисперсионный анализ». Нажмите кнопку *OK*.

4.2. Нормальность распределения остатков

На листе «Регрессия» выберите в меню *Сервис* → *Анализ данных* → *Описательная статистика* → *OK*. Значения параметров в диалоговом окне установите следующим образом:

- *Входной интервал* – введите ссылки на ячейки C25:C45 (столбец «Остатки» с названием).
 - *Группирование по столбцам* – установите флажок.
 - *Метки в первой строке* – установите флажок.
 - *Выходной диапазон* – установите переключатель на *Выходной интервал*, установите курсор в поле напротив и выделите указателем мыши ячейку D26.
 - *Итоговая статистика* – установите флажок.
- Уровень надежности (95%)*, *K-ый наибольший*, *K-ый наименьший* оставьте пустыми. Нажмите кнопку *OK*.

В ячейку A49 введите название «Критерий Пирсона». Выберите в меню *Сервис* → *Анализ данных* → *Гистограмма* → *OK*. Значения параметров окна установите следующим образом:

- *Входной интервал* – введите ссылки на ячейки C25:C45 (столбец «Остатки» с названием).
- *Интервал карманов* – не заполняйте.
- *Метки* – установите флажок.
- *Параметры вывода* – установите переключатель на *Выходной*

интервал и введите ссылку на ячейку A50.

- *Парето* – оставьте пустым.
- *Интегральный процент* – оставьте пустым.
- *Вывод графика* – установите флажок. Нажмите кнопку *ОК*.

Перенесите гистограмму вниз и растяните ее.

Удалите слово *Еще* в столбце «Карман» и в этой же ячейке введите формулу =МАКС(С26:С45)*3, т. е. значение максимума остатков увеличили в три раза.

В ячейку С50 введите значение 0.

В ячейки С51:С55 введите *формулу массива*
{=НОРМРАСП(А51:А55;Е28;Е32;ИСТИНА)}.

В ячейки D51:D55 введите *формулу массива*
{=С51:С55 – С50:С54}.

В ячейки Е51:Е55 введите *формулу массива* {=Е40*D51:D55}.

В ячейки F51:F55 введите *формулу массива*
{=(B51:B55 – E51:E55)^2/E51:E55}.

В ячейку А57 введите обозначение хи-кв набл.

В ячейку В57 введите формулу =СУММ(F51:F55) (для вычисления хи-квадрат набл).

В ячейку А58 введите обозначение хи-кв кр.

В ячейку В58 введите формулу =ХИ2ОБР(0,05; 6 – 2 – 1), где 6 = 5 + 1(число значений в кармане 5 плюс 1), 2 – число параметров нормального распределения (для вычисления хи-квадрат кр).

4.3. Значимость коэффициентов регрессии

В ячейку С20 листа «Регрессия» введите ткр. Вычислите критическое значение ткр в свободной ячейке D20 следующим образом:

- нажмите на f_x (вставка функций);
- в поле *Категория* окна *Мастер функций* выберите *статистические*, из предложенных ниже функций выделите СТЬЮДРАСПОБР и нажмите кнопку *ОК*. Откроется окно *Аргументы функций*. Значения параметров следующие:
 - *Вероятность* – наберите значение 0,05;
 - *Степени свободы* – введите 20 – 2 – 1, где 20 – число наблюдений, 2 – число факторов возраст, стаж в уравнении регрессии, 1 – число свободных членов (b_0) в уравнении регрессии. Нажмите кнопку *ОК*.

4.4. Проверка статистических свойств остатков (качества оценок коэффициентов регрессии)

4.4.1. Центрированность остатков

На листе «Регрессия» в ячейку D25 введите название «Условие 1».

В ячейку D41 введите тнабл. В ячейку E41 введите формулу $= (E28 - 0) * \text{КОРЕНЬ}(E40) / E32$ (для подсчета наблюдаемого значения статистики тнабл). В ячейку D42 введите ткр. В ячейку E42 введите формулу $= \text{СТЮДРАСПОБР}(0,05; 20 - 1)$ (для подсчета критической точки распределения Стьюдента ткр).

4.4.2. Гомоскедастичность (гетероскедастичность) остатков

На листе «Исходные данные» ячейки A1:A21 скопируйте в ячейку A1 нового листа и назовите лист «Условие 2». В ячейку B1 скопируйте из листа «Регрессия» столбец «Остатки» вместе с названием. В ячейку C1 введите название «Модуль ост». В ячейки C2:C21 введите **формулу массива** (выделите ячейки C2:C21, нажмите F2, введите формулу, нажмите комбинацию клавиш *Ctrl+Shift+Enter*) $\{= \text{ABS}(B2:B21)\}$.

Выберите в опциях меню *Сервис* → *Анализ данных* → *Ранг и перцентиль* → *ОК* и заполните диалоговое окно следующим образом: *Входной интервал* – введите ссылки на ячейки A1:A21; *Метки* – установите флажок; *Выходной интервал* – ячейка D1. Нажмите кнопку *ОК*.

Выберите в опциях меню *Сервис* → *Анализ данных* → *Ранг и перцентиль* → *ОК* и заполните диалоговое окно следующим образом: *Входной интервал* – введите ссылки на ячейки C1:C21; *Метки* – установите флажок; *Выходной интервал* – ячейка H1. Нажмите кнопку *ОК*.

Выделите ячейки D2:G21 и нажмите кнопку *Сортировка по возрастанию* на панели инструментов. Выделите ячейки H2:K21 и нажмите кнопку *Сортировка по возрастанию*. В ячейки L2:L21 введите **формулу массива** $\{= (F2:F21 - J2:J21)^2\}$. Введите в L1 название «Квадрат разности рангов».

В ячейку K22 введите тнабл. В ячейку L22 введите формулу $= (1 - 6 * \text{СУММ}(L2:L21) / (20 * (20^2 - 1))) * \text{КОРЕНЬ}(20 - 1)$ (для вычисления тнабл).

В ячейку K23 введите ткр. В ячейку L23 введите формулу $= \text{СТЮДРАСПОБР}(0,05; 20 - 2)$ (для вычисления ткр).

4.4.3. Автокорреляция остатков

На листе «Регрессия» в ячейку F25 введите название «Условие 3».

В ячейки F26:F44 введите **формулу массива**
 $\{= (C26:C44 - C27:C45)^2\}$.

В ячейку F46 введите формулу $= \text{СУММ}(F26:F44)$.

В ячейки G26:G45 введите **формулу массива** $\{= (C26:C45)^2\}$.

В ячейку G46 введите формулу $= \text{СУММ}(G26:G45)$.

В ячейку F47 введите днабл.

В ячейку G47 введите формулу $= F46 / G46$ (для вычисления днабл).

Если критерий Дарбина–Уотсона не дает ответа о наличии автокорреляции, то можно воспользоваться визуальным способом анализа графика зависимости остатков от номера наблюдения, построенного с помощью диаграммы.

Для построения графика остатков выполните следующие действия:

- на панели инструментов активизируйте кнопку *Мастер диаграмм* (шаг 1 из 4), в одноименном диалоговом окне среди стандартных типов выберите *Точечная*, в первой строке первый вид и нажмите кнопку *Далее*>;

- открывается диалоговое окно *Мастер диаграмм* (шаг 2 из 4), в котором во вкладке *Диапазон данных* в поле *Диапазон* введите ссылку на диапазон ячеек C26:C45; нажмите кнопку *Далее*>;

- откройте диалоговое окно *Мастер диаграмм* (шаг 3 из 4), в котором во вкладке *Заголовки* в поле *Название диаграммы* введите название «График остатков», в поле *Ось X(категорий)* – название «Номер наблюдения», в поле *Ось Y(значений)* – название «Остатки»; во вкладке *Легенда* снимите флажок *Добавить легенду* и нажмите кнопку *Далее*>;

– в открывшемся диалоговом окне *Мастер диаграмм* (шаг 4 из 4) в поле *имеющаяся* установите флажок. Нажмите кнопку *Готово*.

4.5. Анализ свойств модели

4.5.1. Мультиколлинеарность факторов: выявление зависимости объясняющих факторов

Для проверки гипотезы об отсутствии мультиколлинеарности используется статистика хи-квадрат с $v = n(n-1)/2$ степенями свободы, наблюдаемое значение которой определяется по формуле

$$\chi^2 = n - 1 - \frac{1}{6}(2p + 5)\lg \Delta r,$$
 где Δr – определитель матрицы парных

коэффициентов корреляции между факторами; n – количество наблюдений; p – число независимых переменных.

На листе «Исходные данные» в ячейку E11 введите название «Мультиколлинеарность».

В ячейку E12 введите название «Определитель».

В ячейку F12 введите математическую формулу
=МОПРЕД(G4:H5).

В ячейку E13 введите название «хи-кв набл».

В ячейку F13 введите формулу = 20 – 1 – 9*LOG(F12;10)/6 (для нахождения хи-квадрат наблюдаемого по выборке).

В ячейку E14 введите название «хи-кв кр».

В ячейку F14 введите формулу = ХИ2ОБР(0,05;20*(20 – 1)/2) (для нахождения хи-квадрат критического).

4.5.2. Эластичность

На листе «Исходные данные» в ячейку A23 введите название «Эластичность».

В ячейку A24 введите название «ЗП_ср».

В ячейку A25 введите формулу = СРЗНАЧ(A2:A21).

В ячейку B24 введите название «Возраст_ср».

В ячейку B25 введите формулу = СРЗНАЧ(B2:B21).

В ячейку C24 введите название «Стаж_ср».

В ячейку C25 введите формулу = СРЗНАЧ(C2:C21).

В ячейку A26 введите название «Коэф. Возраст». Из листа «Регрессия» скопируйте коэффициент при переменной «Возраст» из ячейки B18 в ячейку B26 листа «Исходные данные».

В ячейку A27 введите название «Эластичность» по фактору возраст.

В ячейку B27 введите формулу для вычисления средней частной эластичности по переменной «Возраст» = B26*B25/A25.

В ячейку A28 введите название «Коэф. Стаж». Из листа «Регрессия» скопируйте коэффициент при переменной «Стаж» из ячейки B19 в ячейку B28 листа «Исходные данные».

В ячейку A29 введите название «Эластичность» по фактору стаж.

В ячейку B29 введите формулу для вычисления средней частной эластичности по переменной «Стаж» = B28*C25/A25.

4.5.3. Частные коэффициенты корреляции

Коэффициент частной корреляции первого порядка для переменной x_1 при неизменном значении переменной x_2 находится по формуле

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{(1 - r_{yx_2}^2)(1 - r_{x_1x_2}^2)}} \quad (\text{через коэффициенты парной корреляции факторов}).$$

Для его нахождения выполните следующие действия:

- в ячейку E16 введите название «Частные коэф. корр.»;
- в ячейку E17 введите название «rЗП,В-С»;
- в ячейку F17 введите формулу
= (F4 – F5*G5)/КОРЕНЬ((1 – F5^2)*(1 – G5^2)).

Аналогично найдите $r_{yx_2 \cdot x_1}$:

- в ячейку E18 введите название «rЗП,С-В»;
- в ячейку F18 введите формулу
= (F5 – F4*G5)/КОРЕНЬ((1 – F4^2)*(1 – G5^2)).

Проверка значимости частных коэффициентов осуществляется

сравнением наблюдаемых и критического значений t -статистики аналогично проверке значимости парных коэффициентов корреляции на этапе спецификации (см. выше).

5. Прогнозирование

На листе «Регрессия» в ячейке E1 введите название «Точечный прогноз», в ячейку E2 введите формулу = B17 + B18*35 + B19*10 для расчета точечной оценки заработной платы при возрасте работника 35 лет и стаже работы 10 лет из условия задачи.

Примечание – Стандартная ошибка прогноза считается по формуле $S^* = S \cdot \sqrt{X_p^T (X^T X)^{-1} X_p}$, где X – матрица наблюдений независимых переменных; X_p – матрица значений независимых переменных для прогноза; S – стандартная ошибка регрессии; T – операция транспонирования матрицы.

В ячейку B2 нового листа «Интервальный прогноз» скопируйте ячейки B2:C21 листа «Исходные данные». Заполните ячейки A2:A21 единицами (это значения переменной при свободном члене). Для простоты дальнейших ссылок в объединенные ячейки A1:C1 введите название «Массив 1» (массив X, содержащий значения переменной при свободном члене, возраста, стажа, – ячейки A2:C21), в ячейку D1 – название «Массив 2» (массив X_p , содержащий данные для прогноза, – ячейки D2:D4). В ячейку D2 введите 1, в D3 – 35 (возраст), в D4 – 10 (стаж).

Пример оформления промежуточных вычислений стандартной ошибки прогноза и интервального прогноза приведен на рисунке 10.

Для транспонирования массива 2 введите в ячейки A23:C23 формулу массива {= ТРАНСП(D2:D4)}.

Для транспонирования массива 1 введите в ячейки A25:T27 формулу массива {= ТРАНСП(A2:C21)}.

Результатом произведения транспонированного массива 1 размерностью 3 на 20 и массива 1 размерностью 20 на 3 является массив 3 размерностью 3 на 3, поэтому в ячейки A29:C31 введите формулу массива {= МУМНОЖ(A25:T27;A2:C21)}.

Результатом вычисления обратной матрицы полученного массива 3 является матрица размерностью 3 на 3, которая находится в ячейках A33:C35 по формуле массива {= МОБР(A29:C31)} (массив 4).

Результатом произведения транспонированного массива 2 размерностью 1 на 3 и массива 4 размерностью 3 на 3 является массив 5 размерностью 1 на 3, поэтому в ячейки A37:C37 введите формулу массива {= МУМНОЖ(A23:C23;A33:C35)}.

Результатом произведения массива 5 размерностью 1 на 3 и массива

ва 2 размерностью 3 на 1 является массив 6 размерностью 1 на 1, поэтому в ячейку A39 введите формулу = МУМНОЖ(A37:C37;D2:D4).

Стандартную ошибку прогноза посчитайте в ячейке A41 по формуле = регрессия!В7*КОРЕНЬ(A39).

Интервальный прогноз величины заработной платы рассчитайте в ячейках A43, B43 соответственно по формулам:

= 'регрессия'!E2 – регрессия!D20*'Интервальный_прогноз'!A41
(для левого конца интервала);

= 'регрессия'!E2 + 'регрессия'!D20*'Интервальный_прогноз'!A41
(для правого конца интервала).

Примечание – Запись 'регрессия'!E2 означает, что ячейка E2 находится на листе «Регрессия». Набор и редактирование формулы осуществляется в строке формул.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Массив 1			Массив 2																
2	1	28	10	1																
3	1	26	8	35																
4	1	30	14	10																
5	1	35	15																	
6	1	41	16																	
7	1	45	17																	
8	1	27	3																	
9	1	35	13																	
10	1	30	10																	
11	1	23	5																	
12	1	29	7																	
13	1	33	11																	
14	1	40	20																	
15	1	41	24																	
16	1	41	6																	
17	1	23	2																	
18	1	32	12																	
19	1	37	13																	
20	1	31	7																	
21	1	30	9																	
22	Массив 2 трансп.																			
23	1	35	10																	
24	Массив 1 трансп																			
25	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
26	28	26	30	35	41	45	27	35	30	23	29	33	40	41	41	23	32	37	31	30
27	10	8	14	15	16	17	3	13	10	5	7	11	20	24	6	2	12	13	7	9
28	Массив 3 (произведение массива 1 трансп. и массива 1)																			
29	20	657	222																	
30	657	22349	7799																	
31	222	7799	3062																	
32	Массив 4 (обратный к массиву 3)																			
33	1,88412	-0,06943	0,04023																	
34	-0,06943	0,00296	-0,00251																	
35	0,04023	-0,00251	0,0038																	
36	Массив 5 (произведение массива 2 трансп. и массива 4)																			
37	-0,14353	0,00912	-0,00957																	
38	Массив 6 (произведение массива 5 и массива 2)																			
39	0,08014																			
40	Стандартная ошибка прогноза																			
41	1,01838																			
42	Интервальный прогноз																			
43	697,383	701,68																		

**Рисунок 10 – Пример оформления вычислений
интервальной оценки прогноза
Эконометрический анализ построения модели
множественной регрессии**

1. Постановочный этап

Из экономической теории известно, что заработная плата зависит от многих факторов, например, от возраста, квалификации, стажа по данной специальности, общего стажа работы, производительности труда и т. д. Выделим два фактора – возраст и стаж по данной специальности, которые являются объясняющими факторами для результа-

тивного (объясняемого) фактора – заработная плата. Поэтому возникает задача количественного описания зависимости указанных экономических показателей уравнением множественной регрессии $y = f(x_1, x_2) + \varepsilon$ на основе 20 наблюдений экономических показателей.

2. Спецификация модели

Определим наличие зависимости фактора заработная плата от возраста и стажа, а также форму этой зависимости.

Тесноту связи и наличие линейной зависимости изучаемых экономических показателей оценивает *коэффициент парной корреляции* r_{xy} (см. тему 1).

На листе «Исходные данные» получена таблица 14.

Таблица 14 – Корреляционная матрица

Корреляционная матрица			
	<i>ЗП</i>	<i>Возраст</i>	<i>Стаж</i>
<i>ЗП</i>	1		
<i>Возраст</i>	0,79	1	
<i>Стаж</i>	0,99	0,75	1

Коэффициент корреляции факторов ЗП и возраст равен $0,79 > 0$, поэтому зависимость между ними прямая и высокая. Коэффициент корреляции факторов ЗП и стаж равен $0,99 > 0$, поэтому зависимость между ними прямая и весьма высокая (таблица 14).

Проверим на значимость коэффициенты парной корреляции. На листе «Исходные данные» вычислены наблюдаемые и критическое значения *t*-статистики (таблица 15).

Таблица 15 – Значимость коэффициентов корреляции

Значимость коэффициентов корреляции	
tЗП,В набл	5,56
tЗП,С набл	43,78
tкр	2,10

Так как $|t_{ЗП,В \text{ набл}}| = 5,56 > t_{кр} = 2,1$, то коэффициент корреляции значим (значительно отличается от нуля). Поэтому подтверждается наличие линейной зависимости между факторами ЗП и возраст.

Так как $|t_{ЗП,С \text{ набл}}| = 43,79 > t_{кр} = 2,1$, то коэффициент корреляции значим. Поэтому также подтверждается наличие линейной зависимости между факторами ЗП и стаж.

Исходя из проведенного анализа можно выдвинуть предположение о том, что зависимость заработной платы (y) от возраста (x_1) и стажа по данной специальности (x_2) описывается линейной регрессионной моделью $y = b_0 + b_1x_1 + b_2x_2 + \varepsilon$, где b_0, b_1, b_2 – неизвестные параметры модели; ε – случайный член, который включает в себя суммарное влияние всех неучтенных в модели факторов, ошибки измерений.

3. Параметризация модели

Найдем оценки неизвестных параметров модели.

Для точечной оценки параметров уравнения линейной множественной регрессии применяют *метод наименьших квадратов*, суть которого описана в пункте 3 раздела «Эконометрический анализ ...» темы 1.

Доверительный интервал параметра (интервальная оценка) b_2 множественной регрессии имеет такой же вид, как и для b_1 в теме 1.

В результате проведения регрессионного анализа на листе «Регрессия» получены точечные и интервальные оценки неизвестных параметров модели (таблица 16).

Таблица 16 – Статистика коэффициентов регрессии

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%
Y-пересечение	570,74	4,94	115,58	4,63E-26	560,32	581,16
Возраст	1,03	0,19	5,26	6,39E-05	0,62	1,44
Стаж	9,28	0,22	41,85	1,37E-18	8,81	9,74

Точечная оценка параметра b_0 (Y-пересечение) равна 570,74, ее интервальная оценка – (560,32; 581,16).

Точечная оценка параметра b_1 при переменной x_1 равна 1,03, ее интервальная оценка – (0,62; 1,44).

Точечная оценка параметра b_2 при переменной x_2 равна 9,28, ее интервальная оценка – (8,81; 9,74).

Таким образом, уравнение регрессии имеет вид

$$y = 570,74 + 1,03 x_1 + 9,26 x_2.$$

Так как любое значение из доверительного интервала может служить оценкой параметра, то уравнение регрессии также может иметь вид

$$y = 568 + 0,8 x_1 + 9 x_2.$$

4. Верификация модели

4.1. Общее качество уравнения

Оценим общее качество модели по коэффициенту (индексу) детерминации и нормированному индексу детерминации (см. в пункте 4 раздела «Эконометрический анализ ...» темы 1).

Проанализируем показатели в таблице «Регрессионная статистика» листа «Регрессия» (таблица 17).

Таблица 17 – Регрессионная статистика

<i>Регрессионная статистика</i>	
Множественный R	0,998
R-квадрат	0,996
Нормированный R-квадрат	0,996
Стандартная ошибка	3,597
Наблюдения	20

Коэффициент множественной детерминации R-квадрат равен 0,996. Так как он близок к 1, то уравнение имеет высокое качество. Этот факт подтверждает также нормированный индекс множественной детерминации, равный 0,996.

В таблице «Дисперсионный анализ» листа «Регрессия» рассчитаны наблюдаемое и критическое значения критерия Фишера (таблица 18).

Таблица 18 – Дисперсионный анализ

Дисперсионный анализ					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
Регрессия	2	61 917,59	30 958,8	2 392,35	1,47E-21
Остаток	17	219,99	12,94		
Итого	19	62 137,59			

			Ф _{кр}	3,59	
--	--	--	-----------------	------	--

Так как наблюдаемое значение $F_{\text{набл}} = 2\,392,35 > F_{\text{кр}} = 3,59$, то R-квадрат значим, что еще раз подтверждает высокое качество построенного уравнения линейной множественной регрессии.

4.2. Нормальность распределения остатков

Проанализируем нормальность распределения остатков:

- 1) по гистограмме остатков;
- 2) по числовым характеристикам асимметрии и эксцессу;
- 3) по критерию Пирсона.

Теоретический материал приводится в пункте 4 раздела «Эконометрический анализ ...» темы 1.

1. Построим гистограмму остатков. Соединим середины верхних сторон прямоугольников гистограммы и получим полигон распределения, по которому визуалью можно предположить закон распределения (рисунок 11).

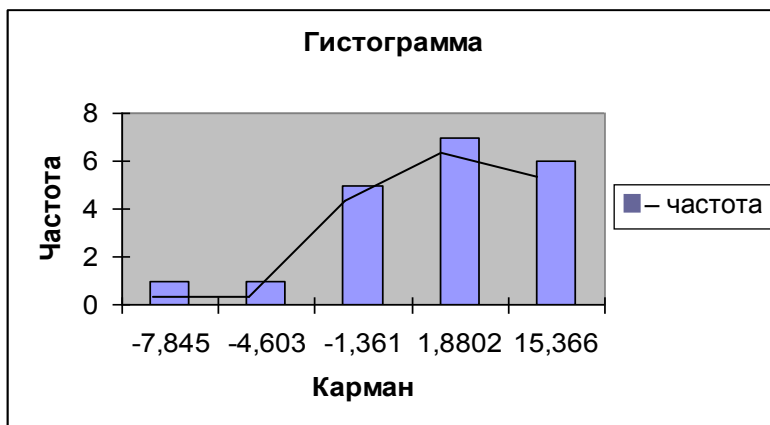


Рисунок 11 – Гистограмма

Так как ломаная линия на рисунке 11 близка к кривой нормального

распределения, заданной уравнением $f(x) = \frac{1}{3,4 \cdot \sqrt{2\pi}} e^{-\frac{(x-0)^2}{2 \cdot (3,4)^2}}$ (срав-

ните с рисунком 2), то остатки распределены по нормальному закону. Следовательно, по визуальному анализу гистограммы можно предпо-

ложить нормальность распределения остатков.

2. Асимметричность равна $-0,36$ (левосторонняя асимметричность эмпирической кривой относительно теоретической), эксцесс равен $0,1$ («островершинность» эмпирической кривой), т. е. характеристики плотности распределения асимметричность и эксцесс незначительно отличаются от нуля, поэтому можно считать распределение нормальным.

3. Подтвердим нормальность распределения с помощью критерия Пирсона.

На листе «Регрессия» найдены наблюдаемое и критическое значения статистики хи-квадрат (таблица 19).

Таблица 19 – Проверка критерия Пирсона

хи-кв набл	3,16
хи-кв кр	7,81

Наблюдаемое значение, равное $3,16$, меньше хи-квадрат критического, равного $7,81$, поэтому остатки распределены по нормальному закону.

4.3. Значимость коэффициентов регрессии

Проверим значимость коэффициентов регрессии.

Проверка значимости коэффициентов регрессии описана в теме 1.

Значимость коэффициентов регрессии оценивается с помощью t -статистики, значения которой получены на листе «Регрессия» (см. таблицу 16).

Наблюдаемое значение статистики для коэффициента b_0 тнабл равно $115,59$ (оно равно отношению точечной оценки коэффициента b_0 к его стандартной ошибке). Критическое значение $t_{кр}$ равно $2,1$. Так как $|t_{набл}| = 115,59 > t_{кр} = 2,1$, то коэффициент b_0 значим.

Аналогично для коэффициента b_1 имеем: $t_{набл} = 5,26$, $t_{кр} = 2,1$. Так как $|t_{набл}| = 5,26 > t_{кр} = 2,1$, поэтому коэффициент b_1 значим. Для коэффициента b_2 имеем: $|t_{набл}| = 41,85 > t_{кр} = 2,1$, поэтому коэффициент b_2 значим.

Значимость коэффициентов регрессии подтверждает выдвинутое на этапе спецификации предположение о линейной форме зависимости факторов.

4.4. Проверка статистических свойств остатков (качества оценок коэффициентов регрессии)

Чтобы оценки параметров регрессии были *несмещенными, эффективными и состоятельными*, необходимо выполнение требований к остаткам, изложенных в теме 1.

4.4.1. Центрированность остатков

Проверим выполнение условия 1 о равенстве математического ожидания остатков нулю.

Среднее из числовых характеристик остатков рассчитано на листе «Регрессия» в таблице «Остатки» (таблица 20).

Таблица 20 – Числовые характеристики остатков

<i>Условие 1</i>	
<i>Остатки</i>	
Среднее	-2,84E-14
Стандартная ошибка	0,76
Медиана	-0,72
Мода	#Н/Д
Стандартное отклонение	3,40
Дисперсия выборки	11,58
Эксцесс	0,10
Асимметричность	-0,36
Интервал	12,97
Минимум	-7,84
Максимум	5,12
Сумма	-5,68E-13
Счет	20
tнабл	-3,74E-14
tкр	2,09

Среднее равно $-2,84E-14 = -2,84 \cdot 10^{-14}$. Оно достаточно близко к нулю, поэтому можно предположить выполнимость условия 1 Гаусса–Маркова. Проверим значение среднего на значимость, т. е. гипотезу о равенстве нулю математического ожидания случайной переменной.

Сравним рассчитанное наблюдаемое и критическое значения статистики. Так как $|t_{\text{набл}}| = 3,74E-14 = 3,74 \cdot 10^{-14} < t_{\text{кр}} = 2,09$, то среднее

незначимо (т. е. незначительно отличается от нуля). Следовательно, условие 1 Гаусса–Маркова выполняется.

4.4.2. Гомоскедастичность (гетероскедастичность) остатков

Проверим выполнение условия 2 о постоянстве и конечности дисперсии остатков, т. е. гомоскедастичность остатков.

На листе «Условие 2» рассчитаны наблюдаемое и критическое значения t -статистики (таблица 21).

Таблица 21 – Проверка гипотезы об отсутствии гетероскедастичности

$t_{\text{набл}}$	-1,25
$t_{\text{кр}}$	2,10

Так как $|t_{\text{набл}}| = 1,25 < t_{\text{кр}} = 2,1$, то гетероскедастичность присутствует. Следовательно, условие 2 Гаусса–Маркова выполняется, значит, МНК-оценки параметров регрессии будут эффективными. Поэтому модель можно использовать при точечном и интервальном прогнозировании.

4.4.3. Автокорреляция остатков

Проверим выполнение условия 3 о независимости случайного члена в любом наблюдении от его значений во всех других наблюдениях.

На листе «Регрессия» в ячейке E47 рассчитано наблюдаемое значение d -статистики: $d_{\text{набл}} = 1,38$. По таблице критических значений d -статистики для числа наблюдений 20, числа объясняющих переменных 2 и заданного уровня значимости 0,05 значения $d_1 = 1,10$ и $d_2 = 1,54$, которые разбивают отрезок $[0; 4]$ на пять областей (рисунок 12).

Положительная автокорреляция	Область неопределенности	Автокорреляция отсутствует	Область неопределенности	Отрицательная автокорреляция	
0	1,10	1,54	2,46	2,90	4

Рисунок 12 – Критические области d -статистики Дарбина–Уотсона

Так как $1,10 < d_{\text{набл}} = 1,38 < 1,54$, т. е. наблюдаемое значение попало в зону неопределенности, то ничего нельзя сказать о наличии автокорреляции, используя критерий Дарбина–Уотсона.

Визуально наличие автокорреляции остатков можно определить по графику остатков, полученному на листе «Регрессия» (рисунок 13).

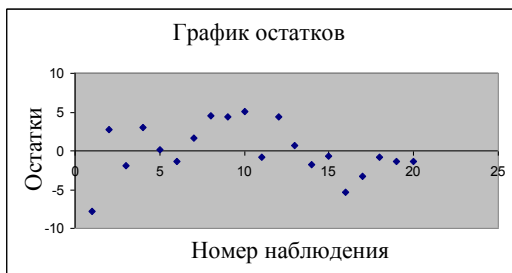


Рисунок 13 – График остатков

Так как на графике остатков точки разбросаны вдоль прямой $y = 0$ хаотично без видимой закономерности, то зависимости между остатками не наблюдается. Поэтому условие 3 выполняется.

4.5. Анализ свойств модели

4.5.1. Мультиколлинеарность факторов

Проверим мультиколлинеарность факторов.

При построении уравнения множественной регрессии может возникнуть проблема *мультиколлинеарности* факторов, их тесной, линейной связанности. Считается, что две переменные явно *коллинеарны*, т. е. находятся между собой в линейной зависимости, если их парный коэффициент корреляции больше или равен 0,7. При наличии мультиколлинеарности МНК-оценки формально существуют, но обладают рядом недостатков. В частности, оценки имеют большие стандартные ошибки, малую значимость, в то время как модель в целом является значимой с высоким значением коэффициента детерминации.

Для отбора факторов в модель регрессии можно использовать корреляционную матрицу. Однако по величине парных коэффициентов корреляции обнаруживается лишь явная коллинеарность факторов. Поэтому при оценке мультиколлинеарности факторов предполагается использовать определитель Δr матрицы парных коэффициентов корреляции между факторами.

Если бы факторы не коррелировали между собой, то матрица парных коэффициентов корреляции между факторами была бы единичной матрицей, поскольку все недиа-

гональные элементы были бы равны нулю. Если же, наоборот, между факторами существует полная линейная зависимость и все коэффициенты корреляции равны 1, то определитель такой матрицы равен 0, т. е. $\Delta r = 0$. Таким образом, чем ближе к 0 определитель матрицы межфакторной корреляции, тем сильнее мультиколлинеарность факторов и ненадежнее результаты множественной регрессии. И, наоборот, чем ближе к 1 определитель матрицы межфакторной корреляции, тем меньше мультиколлинеарность факторов.

Мультиколлинеарность факторов выявляется проверкой гипотезы $H_0 : \Delta r = 1$ с помощью статистики хи-квадрат с $\nu = \frac{1}{2}n(n-1)$ степенями свободы. Наблюдаемое значение статистики определяется по формуле $\chi_{набл}^2 = n-1 - \frac{1}{6}(2p+5)\lg \Delta r$, где n – количество наблюдений; p – число переменных. Если $\chi_{набл}^2 > \chi_{\alpha, \nu}^2$, то гипотеза H_0 отклоняется и наличие мультиколлинеарности объясняющих факторов считается доказанной.

На листе «Исходные данные» найдены парные коэффициенты корреляции и определитель матрицы парных коэффициентов корреляции объясняющих факторов возраст и стаж. Так как парный коэффициент корреляции факторов возраст и стаж $r_{В,С} = 0,75 < 0,8$, то зависимость между факторами существует, но она незначительная. Докажем это предположение проверкой гипотезы об отсутствии мультиколлинеарности с помощью статистики хи-квадрат, наблюдаемое и критическое значения которой найдены на листе «Регрессия» (таблица 22).

Таблица 22 – Мультиколлинеарность

Мультиколлинеарность	
Определитель	0,44
хи-кв набл	19,53
хи-кв кр	223,16

Так как хи-квадрат наблюдаемое равно 19,53 и меньше хи-квадрат критического, равного 223,16, то мультиколлинеарность факторов отсутствует.

4.5.2. Эластичность

Оценим влияние каждого объясняющего фактора на результирующий фактор ЗП.

Частные средние коэффициенты эластичности показывают, на сколько процентов от среднего значения изменяется зависимая переменная с изменением на 1% фактора x_j от своего среднего при фиксированном значении других факторов. *Частные коэффициенты эластичности* по каждой объясняющей переменной для линейной регрессии рассчитываются по формуле

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

На листе «Исходные данные» найдены коэффициенты эластичности факторов возраст и стаж (таблица 23).

Таблица 23 – Эластичность факторов

Эластичность		
ЗП_ср	Возраст_ср	Стаж_ср
707,521 5	32,85	11,1
Коэф. регрессии фактора возраст	1,03	
Коэф. эласт. фактора возраст	0,05	
Коэф. регрессии фактора стаж	9,28	
Коэф. эласт. фактора стаж	0,15	

С изменением значения фактора возраст на 1% при фиксированном значении фактора стаж значение фактора ЗП увеличивается на 0,05%. Аналогично с изменением значения фактора стаж на 1% при фиксированном значении фактора возраст значение фактора ЗП увеличивается на 0,15%. Значит, влияние фактора стаж больше, чем фактора возраст.

4.5.3. Частные коэффициенты корреляции: целесообразность

включения в модель факторов

Определим степень влияния факторов на результирующий фактор ЗП при устранении влияния других факторов.

Частные коэффициенты (или индексы) корреляции характеризуют тесноту связи между результатом и соответствующим фактором при неизменном уровне других факторов, включенных в уравнение регрессии. Они широко используются при решении проблемы отбора факторов, ранжировании факторов, участвующих в множественной линейной регрессии. При нелинейной взаимосвязи исследуемых признаков эту функцию выполняют частные индексы детерминации.

Порядок частного коэффициента корреляции определяется количеством факторов, влияние которых исключается.

Например, $r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1 \cdot x_2}}{\sqrt{(1 - r_{yx_2}^2)(1 - r_{x_1 \cdot x_2}^2)}}$ – коэффициент част-

ной корреляции первого порядка для переменной x_1 при неизменном значении переменной x_2 . Аналогично определяется $r_{yx_2 \cdot x_1}$ через коэффициенты парной корреляции факторов.

Значимость частных коэффициентов корреляции оценивается с помощью t -статистики $t = r\sqrt{(n-2)/(1-r^2)}$, которая имеет распределение Стьюдента. Если $|t| \geq t_{\alpha, k}$, то проверяемый коэффициент значим.

На листе «Исходные данные» найдены частные коэффициенты корреляции (таблица 24).

Таблица 24 – Частные коэффициенты корреляции

Частные коэф. коррел.		Значимость	
гЗП,В-С	0,79	тнаблЗП,В-С	5,41
гЗП,С-В	0,995	тнаблЗП,С-В	43,06

Так как $0,78 < 0,995$, то из двух факторов большее влияние оказывает фактор стаж.

Оба частных коэффициента корреляции значимы: $|t_{\text{наблЗП,В-С}}| = 5,4 > t_{\text{кр}} = 2,1$, $|t_{\text{наблЗП,С-В}}| = 43,06 > t_{\text{кр}} = 2,1$.

Общий вывод по результатам этапа верификации: так как выполняются все условия верификации, то модель является качественной. Таким образом, прогноз, выполненный по ней, является качественным: несмещенным, состоятельным и эффективным.

5. Прогнозирование

Если выполняются все условия верификации, то модель является качественной. В противном случае ее надо усовершенствовать: либо на этапе спецификации, либо варьировать выборку. По качественной модели можно прогнозировать значение зависимой переменной при заданных значениях независимых переменных. Точечный прогноз получается подстановкой заданных значений независимых переменных в уравнение регрессии. Интервальный прогноз – это интервал значений зависимой переменной, содержащий с вероятностью 0,95 истинное ее значение при заданных значениях независимых переменных. Центр интервала равен точечному прогнозу, концы интервалов получены прибавлением и вычитанием произведения стандартной ошибки прогноза на критическое значение t -статистики, т. е. $y^* \pm t_{\alpha, \nu} S^*$. Средняя стандартная ошибка прогноза S^* в матричной форме имеет вид

$$S^* = S \cdot \sqrt{X_p^T (X^T X)^{-1} X_p},$$

где S – стандартная ошибка регрессии; $X_p = \begin{pmatrix} 1 \\ p_1 \\ p_2 \end{pmatrix}$ – матрица

заданных для построения прогноза значений независимых

переменных $x_1 = p_1$, $x_2 = p_2$; $X = \begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} \end{pmatrix}$ – матрица,

составленная из столбца n единиц, столбца n значений переменной x_1 и столбца n значений переменной x_2 из исходных данных; индекс T обозначает операцию транспонирования матрицы.

Так как выполняются все условия верификации, то модель является качественной, следовательно, прогноз, выполненный по ней, является качественным: несмещенным, состоятельным и эффективным. На листе «Регрессия» рассчитан точечный прогноз заработной платы, который равен 699,53. На листе «Интервальный прогноз» получен интервальный прогноз (697,38; 701,68), который означает, что с вероятностью 0,95 любое значение из этого интервала является оценкой заработной платы.

Вопросы для самоконтроля

1. Этапы построения эконометрической модели.
2. Краткая характеристика цели каждого этапа.
3. Знания из каких научных дисциплин необходимы на каждом из этапов эконометрического моделирования?
4. В чем состоит спецификация модели множественной регрессии?
5. Почему в уравнении регрессии присутствует случайная переменная?
6. Как определить силу и направленность взаимодействия факторов?
7. Что означает значимость коэффициента корреляции?
8. Как проверить на значимость коэффициент корреляции?
9. В чем суть МНК для нахождения оценок параметров регрессии?
10. Почему с помощью МНК находят оценки параметров, а не их точные значения?
11. Какая оценка параметра называется точечной?
12. В чем суть интервальной оценки параметров?
13. Как найти интервальные оценки коэффициентов регрессии?
14. Как используются стандартные ошибки регрессии и стандартные ошибки коэффициентов регрессии при анализе оценок параметров регрессии?
15. В чем экономический смысл параметров модели регрессии?
16. Как можно оценить общее качество уравнения регрессии?
17. Какова суть коэффициента детерминации, нормированного коэффициента детерминации? В каких пределах они изменяются?

18. Какова связь между коэффициентом детерминации, коэффициентом корреляции и множественным коэффициентом корреляции для множественной регрессии?

19. Для какой цели в парной регрессии используется критерий Фишера?

20. Как получить остатки для модели парной регрессии?

21. Какому условию должны удовлетворять остатки, чтобы для проверки статистических гипотез можно было использовать критерий Стьюдента?

22. Какое распределение называется нормальным? Каковы его параметры?

23. Какими способами можно проверить нормальность распределения остатков?

24. Как используется критерий согласия Пирсона для проверки гипотезы о нормальном законе распределения остатков?

25. Для чего и как проверяется значимость коэффициентов регрессии?

26. Какими свойствами должны обладать оценки параметров регрессии?

27. Основные предпосылки применения МНК для построения регрессионной модели.

28. Каковы последствия невыполнимости предпосылок применения МНК?

29. Как сформулировать теорему Гаусса–Маркова?

30. Как проверить центрированность остатков?

31. В чем суть гетероскедастичности (гомоскедастичности) остатков?

32. Каковы причины и последствия гетероскедастичности остатков?

33. Как определить гомоскедастичность остатков?

34. В чем суть автокорреляции остатков? Каковы ее последствия?

35. Как проверить гипотезу об отсутствии автокорреляции остатков с помощью критерия Дарбина–Уотсона?

36. Как проверить гипотезу об отсутствии автокорреляции остатков визуально?

37. Что характеризуют средние коэффициенты эластичности?

38. Для чего используются частные коэффициенты корреляции? Как они рассчитываются?

39. Каковы последствия мультиколлинеарности остатков?

40. Что представляет собой прогнозирование?

41. Как получить точечный прогноз зависимого фактора?

42. Как получить интервальный прогноз зависимого фактора?

Индивидуальные задания

Исследовать зависимость заработной платы (y , тыс. р.) от возраста (x_1 , лет) и стажа по данной специальности (x_2 , лет), используя данные наблюдений, приведенные в таблице 13, прибавив к заработной плате значение $10 \cdot k$, где k – номер в журнале. Построить регрессионную модель $y = f(x_1, x_2) + \varepsilon$. Рассчитать значение заработной платы для работника в возрасте 35 лет со стажем работы по данной специальности 10 лет. Оформить отчет.

Тема 3. ВРЕМЕННЫЕ РЯДЫ

Постановка задачи

Динамика выпуска продукции некоторого предприятия характеризуется данными, представленными в таблице 25.

Таблица 25 – Исходные данные для анализа структуры временного ряда, млн р.

Год	Выпуск продукции	Год	Выпуск продукции
1990	5 665	2000	19 037
1991	9 570	2001	21 748
1992	11 172	2002	23 298
1993	10 150	2003	26 570
1994	12 704	2004	26 080
1995	12 588	2005	27 446
1996	13 018	2006	29 658
1997	13 471	2007	32 573
1998	15 017	2008	36 435
1999	17 356	2009	38 100

Проанализируйте структуру временного ряда, проверьте гипотезу о структурной стабильности ряда, проведите аналитическое выравнивание временного ряда, сделайте прогноз на 2011 г.

Технология вычислений в MS Excel при построении модели временного ряда

1. Постановочный этап

Введите подготовленные исходные данные. В ячейку A1 введите

название первого столбца «Год», в ячейку В1 – название второго столбца «Выпуск продукции». В ячейки А2, А3, ..., А21 введите данные первого столбца исходной таблицы, в ячейки В2, В3, ..., В21 – данные второго столбца согласно варианту.

Введите новое название листа «Исходные данные». Сохраните рабочую книгу под названием «Временные ряды».

2. Спецификация: *определение вида аналитической модели временного ряда*

2.1. Анализ структуры временного ряда

2.1.1. Оценка наличия тенденции во временном ряде с помощью корреляционного поля

Для построения корреляционного поля выполните следующие действия:

- на панели инструментов активизируйте кнопку *Мастер диаграмм (шаг 1 из 4)*, в одноименном диалоговом окне среди стандартных типов выберите *График* и во второй строке вид графика с маркерами, помечающими точки данных, и нажмите кнопку *Далее*>;

- открывается диалоговое окно *Мастер диаграмм (шаг 2 из 4)*, в котором во вкладке *Диапазон данных* в поле *Диапазон* введите ссылку на ячейки В1:В21, по умолчанию установится переключатель *Ряды в столбцах*; во вкладке *Ряд* в поле *Подписи оси X* введите ссылку на ячейки А2:А21, нажмите кнопку *Далее*>;

- открывается диалоговое окно *Мастер диаграмм (шаг 3 из 4)*, в котором во вкладке *Заголовки* в поле *Название диаграммы* введите название «Корреляционное поле временного ряда»; в поле *Ось X(категорий)* – название «Год», в поле *Ось Y(значений)* – название «Выпуск продукции»; во вкладке *Легенда* снимите флажок *Добавить легенду* и нажмите кнопку *Далее*>;

- в открывшемся диалоговом окне *Мастер диаграмм (шаг 4 из 4)* в поле *имеющемся* установите флажок; нажмите кнопку *Готово*.

2.1.2. Оценка структуры временного ряда: наличие тренда, сезонности, цикличности, случайной компоненты – по автокорреляционной функции временного ряда и коррелограмме

Для обеспечения статистической достоверности коэффициентов автокорреляции определите число периодов, по которым рассчитыва-

ется коэффициент автокорреляции по правилу: максимальный лаг должен быть не больше $n/4=20/4=5$. На листе «Исходные данные» в ячейку D1 введите название «Период». В ячейки D2:D6 введите значения 1,2,..., 5 (число периодов τ).

В ячейку E1 введите название «Коэф. автокор.».

В ячейку E2 введите формулу коэффициента корреляции r_1
= КОРРЕЛ(B2:B20;B3:B21).

В ячейку E3 введите формулу коэффициента корреляции r_2
= КОРРЕЛ(B2:B19;B4:B21).

В ячейку E4 введите формулу коэффициента корреляции r_3
= КОРРЕЛ(B2:B18;B5:B21).

В ячейку E5 введите формулу коэффициента корреляции r_4
= КОРРЕЛ(B2:B17;B6:B21).

В ячейку E6 введите формулу коэффициента корреляции r_5
= КОРРЕЛ(B2:B16;B7:B21).

Для проверки на значимость коэффициентов автокорреляции выполните следующие действия:

- в ячейку G1 введите название «Значимость коэффициентов автокорреляции»;

- в ячейки G2, G3, G4, G5, G6 введите t_1, t_2, t_3, t_4, t_5 соответственно;

- в ячейку H2 введите формулу = E2*КОРЕНЬ((20 - 2)/(1 - E2^2));

- в ячейки H3:H6 скопируйте формулу из ячейки H2 автозаполнением.

В ячейку D7 введите МАХ.

В ячейку E7 введите формулу = МАКС(E2:E6).

Вычислите наблюдаемое значение статистики:

- в ячейку G7 введите t_{kr} ;

- в ячейку H7 введите формулу = СТЬЮДРАСПОБР(0,05;20 - 2).

Если наиболее высоким (первым в списке) окажется коэффициент автокорреляции порядка $\tau > 1$, то ряд содержит сезонные компоненты с периодом τ , тренд и случайную компоненту; если $\tau = 1$, то ряд содержит только тренд и случайную компоненту.

Для построения коррелограммы выполните следующие действия:

- на панели инструментов активизируйте кнопку *Мастер диаграмм (шаг 1 из 4)*, в одноименном диалоговом окне среди стандартных типов выберите *График* и во второй строке первый график с маркерами, помечающими точки данных, и нажмите кнопку *Далее*>;

- открывается диалоговое окно *Мастер диаграмм (шаг 2 из 4)*, в котором во вкладке *Диапазон данных* в поле *Диапазон* введите ссылку на ячейки E2:E6; во вкладке *Ряд* в поле *Подписи оси X* введите ссылку на ячейки D2:D6 значений τ , в поле *Имя* – название «Коэффициенты автокорреляции уровней»; нажмите кнопку *Далее*>;

- откройте диалоговое окно *Мастер диаграмм (шаг 3 из 4)*, в котором во вкладке *Заголовки* в поле *Ось X(категорий)* введите название «Периоды», в поле *Ось Y(значений)* – название «Коэффициенты автокорреляции уровней»; во вкладке *Легенда* снимите флажок *Добавить легенду* и нажмите кнопку *Далее*>;

- в открывшемся диалоговом окне *Мастер диаграмм (шаг 4 из 4)* в поле *имеющемся* установите флажок; нажмите кнопку *Готово*.

2.2. Определение вида модели (аддитивная или мультипликативная) по корреляционному полю

Примечание – Подробное описание выбора вида модели приведено в разделе «Эконометрический анализ ...».

3. Аналитическое выравнивание временного ряда

3.1. Структурная стабильность временного ряда

Определяется по виду корреляционного поля, построенного в подпункте 2.1.1.

3.2. Проведение аналитического выравнивания временного ряда

Для аналитического выравнивания временного ряда выполните следующие действия:

- на панели инструментов активизируйте кнопку *Мастер диаграмм (шаг 1 из 4)*, в одноименном диалоговом окне среди стандартных типов выберите *График* и вид диаграммы *График с маркерами, помечающими точки данных*, нажмите кнопку *Далее*>;

- открывается диалоговое окно *Мастер диаграмм (шаг 2 из 4)*, в котором во вкладке *Диапазон данных* в поле *Диапазон* вводится ссылка на диапазон ячеек B2:B21; во вкладке *Ряд* в поле *Подписи оси X* вводится ссылка на ячейки A2:A21; нажмите кнопку *Далее*>;

- открывается диалоговое окно *Мастер диаграмм (шаг 3 из 4)*, в котором во вкладке *Заголовки* в поле *Название диаграммы* введите название «Выпуск продукции», в поле *Ось X(категорий)* – название «Годы», в поле *Ось Y(значений)* – название «Объем продукции»; во вкладке *Легенда* снимите флажок *Добавить легенду* и нажмите кнопку *Далее*>;

- в открывшемся диалоговом окне *Мастер диаграмм (шаг 4 из 4)* в поле *имеющемся* установите флажок; нажмите кнопку *Готово*.

Выделите полученную кривую нажатием левой кнопки мыши, затем на ней же нажмите правую кнопку мыши и выберите *Добавить линию тренда*. В диалоговом окне выделите линейный тип линии

тренда, во вкладке *Параметры* установите флажок в полях *Показать уравнение линии тренда, показать ... R²*. Нажмите кнопку *ОК*.

Последнее действие *Добавить линию тренда* проделайте для логарифмической линии тренда, сравните их показатели R^2 , удалите линию с меньшим R^2 (выделите и нажмите клавишу *Delete*). Аналогично, добавляя полиномиальные степени 2 и 3, степенную и экспоненциальную линии, каждый раз попарно сравнивая R^2 , удаляйте линию с наименьшим показателем R^2 .

4. Верификация

В результате предыдущих действий останется линия тренда с максимальным R^2 . Соответствующее ей уравнение будет наилучшей формой тренда.

5. Прогнозирование

В ячейку A23 введите название года 2011. В ячейке B23 рассчитайте для выбранной формы прогнозируемое значение выпуска продукции по формуле трендовой модели при $x = 22$ (порядковый номер года – 2011, считая 1990 г. первым по порядку).

Эконометрический анализ построения модели временного ряда

1. Постановочный этап

Временной ряд – это совокупность значений какого-либо показателя за ряд последовательных периодов времени.

Модель, построенная по данным временного ряда, называется моделью временного ряда.

Выделяют следующие основные задачи исследования временных рядов:

- характеристика структуры временных рядов;
- определение вида аналитической модели временного ряда;
- определение аналитической модели временного ряда;
- анализ качества модели;
- прогнозирование по модели временного ряда.

Динамика выпуска продукции представлена в таблице 25 наблюдениями, характеризующими некоторое предприятие за 20 последовательных лет. Возникает задача количественного описания динамики выпуска продукции с использованием модели временного ряда, где t – независимая переменная, y – зависимая переменная.

2. Спецификация (определение вида аналитической модели временного ряда)

2.1. Анализ структуры временного ряда

Уровни временного ряда – это значения $y_t (t = 1, 2, \dots, n)$ наблюдаемого показателя в каждом из n временных периодов (дней, недель, месяцев, кварталов, лет).

Каждый уровень формируется под воздействием большого числа факторов, которые условно можно подразделить на три группы:

1) факторы, формирующие тренд ряда, т. е. изменение динамики значений исследуемого показателя под совместным долговременным воздействием множества факторов;

2) факторы, формирующие циклические колебания (периодические колебания, выходящие за рамки более короткого периода, например, одного года) и сезонные колебания (периодические колебания, в рамках более короткого периода, например, года);

3) случайные факторы, которые нерегулярно воздействуют на временной ряд (например, факторы резкого и внезапного действия), ошибки наблюдений.

Под воздействием этих факторов проявляется тенденция изменения (возрастания, убывания, стабильности) временного ряда, которая содержит все три компоненты: случайную ошибку, тренд, сезонность (циклическость). Некоторые временные ряды не содержат тренда и циклическую компоненту, а каждый следующий их уровень образуется как сумма среднего уровня ряда и некоторой случайной компоненты.

Уровень временного ряда можно представить как сумму или произведение трендовой, циклической и случайной структурных компонент ряда.

Модель, в которой временной ряд представлен как сумма перечисленных компонент, называется аддитивной моделью временного ряда.

Модель, в которой временной ряд представлен как произведение перечисленных компонент, называется мультипликативной моделью временного ряда.

Выбор одной из двух моделей проводится на основе анализа структуры сезонных колебаний. Если амплитуда колебаний приблизительно постоянна, строят аддитивную мо-

дель временного ряда, в которой значения сезонной компоненты предполагаются постоянными для различных циклов. Если амплитуда сезонных колебаний возрастает или уменьшается, строят мультипликативную модель временного ряда, которая ставит уровни ряда в зависимость от значений сезонной компоненты.

По реальным данным строится модель, содержащая хотя бы одну компоненту.

2.1.1. Оценка наличия тенденции во временном ряде с помощью корреляционного поля

Корреляционным полем временного ряда называется множество точек на плоскости с координатами $(t; y)$. По виду корреляционного поля оценивается тенденция, характеризующая совокупное долговременное воздействие множества факторов на динамику изучаемого показателя, которые в совокупности формируют его возрастание или убывание либо их отсутствие.

По виду корреляционного поля временной ряд содержит возрастающую тенденцию (рисунок 14).

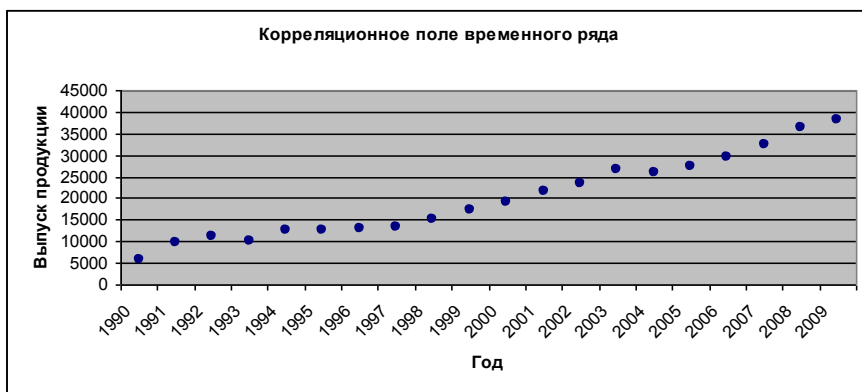


Рисунок 14 – Корреляционное поле временного ряда

2.1.2. Оценка структуры временного ряда: наличие тренда, сезонности, цикличности, случайной компоненты – по автокорреляционной функции временного ряда и коррелограмме

При наличии тенденции значение каждого последующего уровня ряда зависит от предыдущих значений. Корреляционную зависимость между последовательными уровнями временного ряда называют автокорреляцией уровней ряда. Количественно ее можно измерить с помощью линейного коэффициента корреляции между уровнями y_t исходного временного ряда и уровнями $y_{t-\tau}$ этого ряда, сдвинутыми на τ шагов во времени, по следующей формуле:

$$r_{\tau} = \frac{\sum_{t=1}^{n-\tau} (y_t - \bar{y}_1) \cdot (y_{t+\tau} - \bar{y}_2)}{\sqrt{\sum_{t=1}^{n-\tau} (y_t - \bar{y}_1)^2 \cdot \sum_{t=\tau+1}^n (y_{t+\tau} - \bar{y}_2)^2}}$$

где $\bar{y}_1 = \frac{\sum_{t=1}^{n-\tau} y_t}{n-\tau}$, $\bar{y}_2 = \frac{\sum_{t=\tau+1}^n y_t}{n-\tau}$.

Число периодов τ , по которым рассчитывается коэффициент автокорреляции, называется лагом. Считается целесообразным для обеспечения статистической достоверности коэффициентов автокорреляции использовать правило: максимальный лаг должен быть не больше $n/4$.

Коэффициент автокорреляции является аналогом линейного коэффициента корреляции, изменяется в пределах от -1 до 1 . Его значимость оценивается с помощью t -статистики:

$t = r_{\tau} \sqrt{(n-2)/(1-r_{\tau}^2)}$, которая имеет распределение Стьюдента с $k = n-2$ степенями свободы. Если $|t| \geq t_{\alpha,k}$, то проверяемый коэффициент значим.

На основании автокорреляционных коэффициентов осуществляется анализ структуры временного ряда. Если наиболее высоким оказался коэффициент автокорреляции первого порядка, исследуемый ряд содержит только трендовую и случайную компоненты. Если наиболее высоким оказался коэффициент автокорреляции порядка τ , то ряд содержит циклические колебания с периодичностью τ . Если ни один из коэффициентов автокорреляции не является значимым, можно сделать предположение относительно структуры этого ряда: ряд не содержит тренда и циклических колебаний, а толь-

ко случайную компоненту.

По коэффициенту автокорреляции можно судить о наличии линейного тренда. Если временной ряд имеет сильную нелинейную тенденцию, то коэффициент автокорреляции уровней исходного ряда может приближаться к нулю. По знаку коэффициента автокорреляции нельзя делать вывод о возрастающей или убывающей тенденции в уровнях ряда.

Последовательность коэффициентов автокорреляции уровней первого, второго и следующих порядков называют *автокорреляционной функцией временного ряда*. График зависимости ее значений от величины лага (порядка коэффициента автокорреляции) называется *коррелограммой*. При помощи анализа автокорреляционной функции и коррелограммы можно выявить структуру ряда.

Так как максимальным является коэффициент автокорреляции первого порядка (таблица 26 и рисунок 15), то исследуемый ряд не содержит сезонной компоненты, т. е. содержит только трендовую и случайную структурные компоненты.

Таблица 26 – Значения автокорреляционной функции

Период	Кэф. автокор.
1	0,99
2	0,98
3	0,98
4	0,97
5	0,97

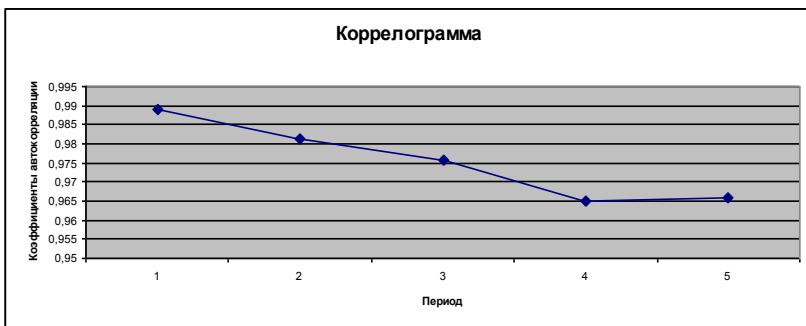


Рисунок 15 – Автокорреляционная функция

Так как все наблюдаемые значения t -статистики больше критического (таблица 27), то все коэффициенты автокорреляции значимы. Это подтверждает предположение о том, что ряд содержит трендовую и случайную компоненты.

Таблица 27 – Значимость коэффициентов автокорреляции

Значимость	
t1	28,15
t2	21,57
t3	18,97
t4	15,66
t5	15,80
tkp	2,10

2.2. Определение вида модели (аддитивная или мультипликативная) по корреляционному полю

Если имеются несколько компонент, то необходимо определить, как объединить их в модели: умножением или сложением. Для этого анализируется корреляционное поле. Если амплитуда колебаний временного ряда затухает, то временной ряд описывается мультипликативной моделью, если амплитуда колебаний постоянна, то – аддитивной моделью.

Анализ рисунка 14 показывает, что колебания временного ряда отсутствуют. Следовательно, модель аддитивна и представляет собой сумму трендовой и случайной компоненты.

3. Аналитическое выравнивание временного ряда

3.1. Структурная стабильность временного ряда

Существуют единовременные изменения характера тенденции временного ряда, вызванные структурными изменениями в экономике или иными факторами. Они отличаются от сезонных и циклических колебаний тем, что, начиная с некоторого момента времени, происходит изменение характера динамики изучаемого показателя. Это приводит к изменению параметров тренда, описывающего эту динамику (рисунок 16).

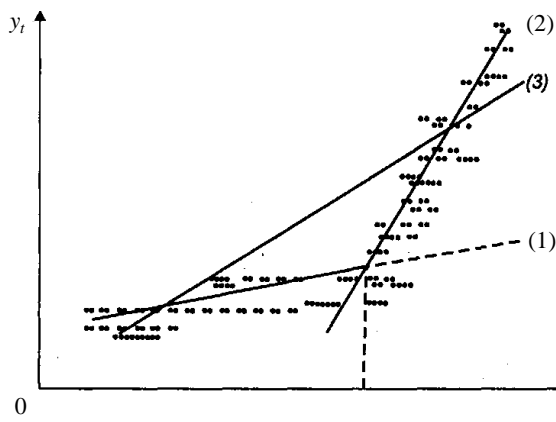


Рисунок 16 – Изменение характера тенденции временного ряда

В момент времени t^* значительно изменяется влияние ряда факторов на изучаемый показатель (например, изменения структуры экономики: начало крупных экономических реформ, изменение экономического курса, нефтяные кризисы и прочие факторы). Возникает вопрос о том, значимо ли повлияли общие структурные изменения на характер тенденции изменения изучаемого показателя.

Если это влияние значимо, то для моделирования тенденции данного временного ряда следует использовать кусочно-линейные модели регрессии, т. е. разделить исходную совокупность на две подсовкупности (до момента времени t^* и после него) и построить отдельно по каждой подсовкупности уравнения линейной регрессии (на рисунке 16 этим уравнениям соответствуют прямые (1) и (2)). Если структурные изменения незначительно влияли на характер тенденции ряда, то ее можно описать с помощью единого для всей совокупности данных уравнения (на рисунке 16 этому уравнению соответствует прямая (3)).

Остаточная сумма квадратов кусочно-линейной модели меньше по сравнению с общим для всей совокупности уравнением тренда. Однако разделение исходной совокупности на две части ведет к потере числа наблюдений и, следовательно, к снижению числа степеней свободы в каждом уравнении кусочно-линейной модели. Построение единого для всей совокупности уравнения тренда, напротив, позволяет

сохранить число наблюдений n исходной совокупности.

Выбор одной из двух моделей (кусочно-линейной или единого уравнения тренда) будет зависеть от соотношения между снижением остаточной дисперсии и потерей числа степеней свободы при переходе от единого уравнения регрессии к кусочно-линейной модели.

Гипотеза о наличии структурной стабильности временного ряда проверяется по тесту Чоу при справедливости гипотезы о равенстве генеральных дисперсий для выделенных подвыборок.

По корреляционному полю на рисунке 14 определяем, что изменения характера динамики временного ряда не наблюдается, так как точки расположены вдоль одной линии. Значит, нет необходимости использовать кусочно-линейные модели регрессии.

3.2. Проведение аналитического выравнивания временного ряда

Одним из наиболее распространенных способов моделирования временного ряда является построение тренда или аналитической функции, характеризующей зависимость уровней ряда от времени. Этот способ называют *аналитическим выравниванием временного ряда*.

Для аналитического выравнивания могут применяться следующие функции:

- линейная $\tilde{y}_t = a + b \cdot t$;
- гиперболическая $\tilde{y}_t = a + \frac{b}{t}$;
- экспоненциальная $\tilde{y}_t = e^{a+b \cdot t}$;
- степенная $\tilde{y}_t = a \cdot t^b$;
- полиномы второго и более высоких порядков $\tilde{y}_t = a + b_1 \cdot t + b_2 \cdot t^2 + \dots + b_k \cdot t^k$.

Параметры каждого из перечисленных выше трендов можно определить обычным МНК, используя в качестве независимой переменной время, а в качестве зависимой переменной – фактические уровни временного ряда y_t . Для нелинейных трендов предварительно проводят стандартную процедуру их линеаризации.

Известно несколько способов определения типа трендов. К наиболее распространенным относятся качественный анализ изучаемого процесса, построение и визуальный анализ графика зависимости уровней ряда от времени, расчет некоторых основных показателей динамики, коэффициенты автокорреляции уровней ряда. Тип тренда можно определить путем сравнения коэффициентов автокорреляции первого порядка, рассчитанных по исходным и преобразованным уровням ряда. Если временной ряд имеет линейный тренд, то его соседние уровни тесно коррелируют. В этом случае коэффициент автокорреляции первого порядка уровней исходного ряда должен быть высоким. Если временной ряд содержит нелинейный тренд, например, в форме экспоненты, то коэффициент автокорреляции первого порядка по логарифмам уровней исходного ряда будет выше, чем соответствующий коэффициент, рассчитанный по уровням ряда. Чем сильнее выражен нелинейный тренд в изучаемом временном ряде, тем в большей степени будут различаться значения указанных коэффициентов.

4. Верификация

Выбор наилучшего уравнения в случае, если ряд содержит нелинейный тренд, можно осуществить путем перебора основных форм тренда, расчета по каждому уравнению скорректированного коэффициента детерминации R_2 , значимость которого оценивается по критерию Фишера, и выбора уравнения тренда с максимальным значением скорректированного коэффициента детерминации. Реализация этого метода относительно проста при компьютерной обработке данных.

При наличии неявного нелинейного тренда следует дополнять описанные выше методы выбора наилучшего уравнения тренда качественным анализом динамики изучаемого показателя, чтобы избежать ошибок спецификации при выборе вида тренда. Качественный анализ предполагает изучение проблем возможного наличия в исследуемом временном ряде поворотных точек и изменения темпов прироста, начиная с определенного момента (периода) времени под влиянием ряда факторов. В случае, если уравнение тренда выбрано неверно при больших значениях выборки (ошибка спецификации), результаты анализа и прогнозирования динамики временного ряда с использованием выбранного уравнения будут недостоверными.

Так как наибольшее значение коэффициента детерминации 0,98 имеет уравнение, заданное кубическим полиномом, то в качестве модели можно использовать это уравнение (рисунок 17). Однако, значение коэффициента детерминации линейного тренда равно 0,96, что также дает право использовать его для прогноза. Как правило, при прогнозировании предпочтение отдается линейному тренду, если по качеству он незначительно уступает нелинейному.

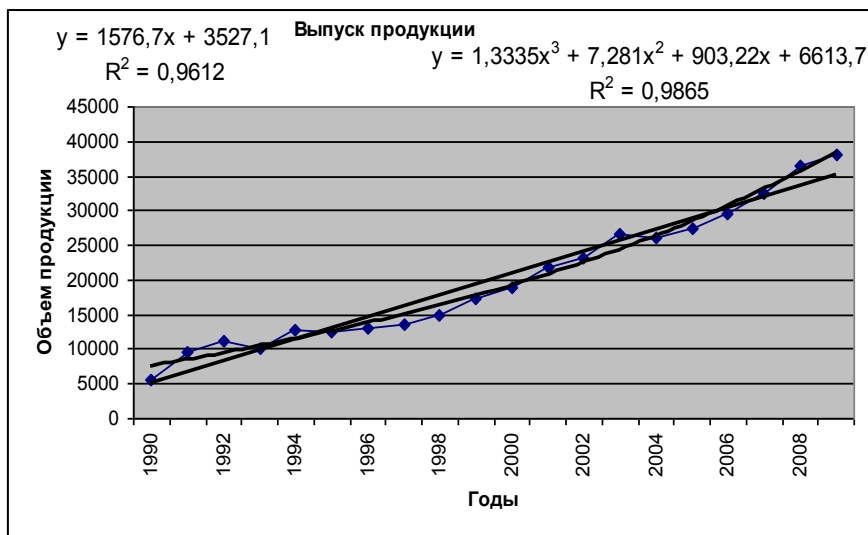


Рисунок 17 – Подбор линии тренда

5. Прогнозирование

Используя линию тренда (кубический полином), осуществляется прогноз выпуска продукции, который составит в 2011 г. 44 208 ед. Прогноз выпуска продукции по линейному тренду составит 38 214,5 ед. Заметим, что полином лучше описывает имеющуюся выборку, но прогнозное значение резко увеличивается по сравнению с наблюдаемыми значениями. Прогноз по линейному тренду более достоверен.

Вопросы для самоконтроля

1. Каково определение модели временного ряда?

2. Какие известны основные компоненты временного ряда?
3. Каковы основные цели исследования временных рядов?
4. Как использовать автокорреляционную функцию при анализе структуры временного ряда?
5. Как рассчитывается коэффициент автокорреляции пятого порядка?
6. Как строится коррелограмма?
7. Каков общий вид мультипликативной и аддитивной моделей временного ряда?
8. С какой целью проводится анализ структуры сезонных колебаний временного ряда?
9. Какие тесты используются для проверки гипотезы о структурной стабильности временного ряда?
10. В каком случае нарушается структурная стабильность временного ряда?
11. Что понимается под аналитическим выравниванием временного ряда?
12. Каковы известны наиболее распространенные модели, используемые для аналитического выравнивания временного ряда?
13. Что понимается под линейризирующими преобразованиями? Как они используются в МНК?
14. Как оценивается качество построенной модели?
15. Как осуществляется точечный прогноз по модели временного ряда?

Индивидуальные задания

Динамика выпуска продукции некоторого предприятия характеризуется данными, представленными в таблице 25 (в каждом варианте к объему выпускаемой продукции надо прибавить число $120 \cdot k$, где k – порядковый номер студента в журнале группы). Требуется выполнить следующее:

- проанализировать структуру временного ряда;
- проверить гипотезу о структурной стабильности ряда;
- провести аналитическое выравнивание временного ряда;
- сделать прогноз на 2011 г.;
- оформить отчет.

Тема 4. ЗАВИСИМОСТЬ ПЕРЕМЕННЫХ, ЗАДАННЫХ ВРЕМЕННЫМИ РЯДАМИ

Постановка задачи

Построить модель зависимости фактора Y от фактора X по данным наблюдений за 22 месяца, предполагая, что на причинно-следственную связь между ними влияет ложная корреляция, вызванная наличием тренда в каждом из временных рядов факторов (таблица 28). Сделать прогноз значения Y для $X = 15$.

Таблица 28 – Данные наблюдений

Месяц	Фактор X	Фактор Y	Месяц	Фактор X	Фактор Y
1	9,098	5,490	12	11,305	5,905
2	9,137	5,540	13	11,430	6,125
3	9,095	5,305	14	11,450	6,185
4	9,280	5,505	15	11,697	6,225
5	9,230	5,420	16	11,870	6,495
6	9,348	5,320	17	12,018	6,720
7	9,525	5,540	18	12,525	6,920
8	9,755	5,690	19	12,055	6,470
9	10,280	5,870	20	12,088	6,395
10	10,665	6,157	21	12,215	6,555
11	11,020	6,342	22	12,495	6,755

Технология вычислений в MS Excel взаимосвязи двух временных рядов

Ввод данных. В ячейку A1 введите название первого столбца «Месяц», в ячейку B1 – название второго столбца «Фактор X», в ячейку C1 – название третьего столбца «Фактор Y». В ячейки A2, A3,..., A23 введите данные первого столбца исходной таблицы, в ячейки B2, B3,..., B23 – второго столбца, в ячейки C2, C3,..., C23 – третьего.

Дайте название рабочему листу «Предварительный анализ рядов». Сохраните рабочую книгу под названием «Зависимость переменных, заданных временными рядами».

Предварительный анализ временных рядов каждого наблюдаемого фактора для выявления их структуры и устранения сезон-

ных (циклических) колебаний по их автокорреляционным функциям. Построение модели $Y=f(X)$ зависимости фактора Y от фактора X требует устранения сезонности во временных рядах этих факторов и влияния фактора времени на каждую переменную в соответствующей модели.

Установите порядок наибольшего коэффициента автокорреляции уровней каждого из рядов. Для этого выполните действия, описанные в подпункте 2.1.2 раздела «Технология вычислений...» темы 3, для фактора X в ячейках E1:F8, а затем для фактора Y в ячейках H1:I8.

Если порядок наибольшего коэффициента автокорреляции уровней ряда равен 1, то сезонность отсутствует. Если же временные ряды содержат сезонную компоненту, то ее влияние исключается следующим образом: для каждого уровня находится сезонная составляющая методом скользящей средней, которая вычитается в случае аддитивной модели из наблюдаемых значений, в случае мультипликативной модели наблюдаемые значения делятся на сезонную компоненту [8]. Полученные временные ряды факторов X , Y должны содержать только трендовую и случайную компоненты.

Устранение влияния фактора времени на каждую переменную в модели методом отклонений от тренда. Постройте регрессионную модель зависимости фактора X от времени. Для этого выполните действия, приведенные ниже.

В ячейку A26 введите название «Модель временного ряда X ».

В меню *Сервис* выберите дополнение *Анализ данных*, укажите *Регрессия*. Нажмите кнопку *ОК*. Установите следующим образом значения параметров:

- *Входной интервал Y* – введите ссылки на ячейки B1:B23.
- *Входной интервал X* – введите ссылки на ячейки A1:A23.
- *Метки* – установите флажок.
- *Уровень надежности* – установите флажок.
- *Константа-ноль* – не активизируйте.
- *Параметры вывода* – установите флажок на *Выходной интервал*, поставьте курсор в поле напротив и активизируйте ячейку A27.
- *Остатки* – установите флажок.
- *Стандартизированные остатки* – оставьте пустым.
- *График остатков* – оставьте пустым.
- *График подбора* – оставьте пустым.
- *График нормальной вероятности* – оставьте пустым.

Нажмите кнопку *ОК*.

Аналогично постройте регрессионную модель зависимости факто-

ра Y от времени в ячейке K26.

Остатки каждой модели (отклонения от трендов) проверьте на автокорреляцию. В ячейку C50 введите название «Остатки по X», в M50 – «Остатки по Y» (для конкретизации при дальнейшем использовании).

Найдите коэффициенты автокорреляции остатков первого порядка. В ячейку C74 введите формулу =КОРРЕЛ(C51:C71;C52:C72).

В ячейку M74 введите формулу =КОРРЕЛ(M51:M71;M52:M72).

В ячейки B74 и L74 введите название «Коэффициент корреляции остатков».

Для проверки на значимость коэффициентов автокорреляции выполните следующие действия:

- в ячейки B75 и L75 введите $t_{наблX}$ и $t_{наблY}$ соответственно;

- в ячейку C75 введите формулу

$$= C74 * \text{КОРЕНЬ}((22 - 2) / (1 - C74^2));$$

в ячейку M75 введите формулу

$$= M74 * \text{КОРЕНЬ}((22 - 2) / (1 - M74^2));$$

- в ячейки B76 и L76 введите $t_{кр}$;

- в ячейки C76 и M76 введите формулу

$$= \text{СТЮДРАСПОБР}(0,05;22 - 2).$$

Сравните наблюдаемое и критическое значения статистики. Если имеется автокорреляция остатков по любой из них, то она устраняется.

Устранение автокорреляции остатков во временном ряде X(t).

На новый лист «Устранение автокорреляции по X» в ячейку A1 скопируйте остатки по X из ячеек C50:C72 листа «Предварительный анализ рядов». В ячейку B1 введите название «Остатки по X со сдвигом». В ячейку B2 скопируйте содержимое ячеек A3:A23. Постройте регрессионную модель (*Сервис* → *Анализ данных* → *Регрессия*), взяв в качестве входного интервала Y ячейки B1:B22, входного интервала X – A1:A22, установив флажок только в метках. В качестве выходного интервала задайте ячейку A25. Если свободный член (Y-пересечение) не значим, то необходимо построить уточненную регрессионную модель в ячейке A45, в которой константа равна нулю (в *Регрессии* установив флажок в поле *Константа-ноль*).

Скопируйте содержимое ячеек A1:B23 листа «Предварительный анализ рядов» в ячейку A64 листа «Устранение автокорреляции по X». В ячейку C64 введите X^* , в ячейку D64 – t^* . В ячейки C65:C85 введите формулу массива {=B66:B86 – B62*B65:B85} (формула авторегрессионного преобразования первого порядка AR(1)). В ячейки D65:D85 введите формулу массива {=A66:A86 – B62*A65:A85} (формула AR(1)).

По полученным значениям X^* и t^* постройте в ячейке F64 регрессионную модель $X^*(t^*)$ (*Сервис* → *Анализ данных* → *Регрессия*), взяв в качестве входного интервала Y ячейки C64:C85, входного интервала X – ячейки D64:D85, установив флажок в метках и остатках. Убедитесь, что автокорреляция остатков этой модели отсутствует (с помощью критерия Дарбина–Уотсона или выборочного коэффициента корреляции остатков первого порядка).

Построение регрессионной модели по отклонениям от трендов.

На новом листе «Модель по отклонениям» выполните следующее:

- введите в ячейку A1 название «Месяц», B1 – «Остатки модели $X^*(t^*)$ », C1 – «Остатки модели $Y(t)$ »;
- скопируйте в ячейку A2 ячейки A3:A23 листа «Предварительный анализ рядов», в ячейку B2 – ячейки H88:H108 листа «Устранение автокорреляции по X», в ячейку C2 – ячейки M52:M72 листа «Предварительный анализ рядов».

В ячейке A25 постройте регрессионную модель по отклонениям от трендов (*Сервис* → *Анализ данных* → *Регрессия*), взяв в качестве входного интервала Y ячейки C1:C22, входного интервала X – ячейки B1:B22, установив флажок в метках и остатках. Оцените качество модели (см. тему 1) и, если необходимо, уточните модель (например, в *Регрессии* установите флажок *Константа-ноль*). Убедитесь, что автокорреляция остатков этой модели отсутствует.

Прогнозирование. Создайте новый лист и дайте ему название «Прогноз».

Примечание – Оформление вычислений на листе «Прогноз» представлено в приложении Г.

На листе «Прогноз» выполните следующее:

- введите в ячейку A1 название «Прогноз для $X = 15$ », в ячейку A2 – «Модель временного ряда $X(t)$ », в ячейку D2 – «Модель временного ряда $Y(t)$ »;
- скопируйте в ячейку A3 ячейки A42:B44 листа «Предварительный анализ рядов», в ячейку D3 – ячейки K42:L44 листа «Предварительный анализ рядов»;
- введите в ячейку A7 название «Модель временного ряда $X^*(t^*)$ »;
- скопируйте в ячейку A8 ячейки F79:G81 листа «Устранение автокорреляции по X»;
- введите в ячейку A12 название «Модель авторегрессии времен-

ного ряда $X(t)$ »;

- скопируйте в ячейку A13 ячейки A40:B42 листа «Устранение автокорреляции по X»;
- введите в ячейку A17 название «Модель по отклонениям»;
- скопируйте в ячейку A18 ячейки A88:B90 листа «Модель по отклонениям».

Ячейки A22:B31 заполните согласно рисунку 18.

22	t находится по модели $X(t)$ при $X=15$	$= (15 - B4) / B5$
23	t^* находится из AR(1)	$= B22 - B15 * (B22 - 1)$
24	X^* модельное находится по модели $X^*(t^*)$	$= B9 + B10 * B23$
25	X при t-1 по модели $X(t)$	$= B4 + B5 * (B22 - 1)$
26	X из условия	15
27	X^* набл., используя AR(1)	$= B26 - B15 * B25$
28	отклонение Δx^*	$= B27 - B24$
29	остаток по модели по отклонениям	$= B19 + B20 * B28$
30	Упрогнозируемый по модели $Y(t)$	$= E4 + E5 * B22$
31	Упрогнозируемый с учетом исключения трендов	$= B30 + B29$

Рисунок 18 – Вычисление прогнозного значения фактора Y

Эконометрический анализ построения модели взаимосвязи двух временных рядов

Предварительный анализ временных рядов.

Установление причинно-следственных зависимостей временных, представленных в форме временных рядов, требует предварительного анализа структуры изучаемых рядов, поскольку наличие сезонных (или циклических) колебаний может исказить корреляционно-регрессионный анализ: если оба ряда содержат циклические колебания одинаковой периодичности, то сила тесноты связи изучаемых временных рядов будет завышена; если сезонные (или циклические) колебания содержит только один из рядов или периодичность колебаний в рассматриваемых временных рядах различна, то эти показатели будут занижены.

Перед проведением дальнейшего исследования взаимосвязи необходимо устранить сезонную (или циклическую)

компоненту из уровней каждого ряда. Для этого рассчитывается сезонная компонента методом скользящей средней и строится мультипликативная или аддитивная модель временного ряда [8, с. 311].

После устранения сезонных (или циклических) компонент для количественной характеристики зависимости между рядами x и y используется линейный коэффициент корреляции. Если рассматриваемые временные ряды имеют тренд, коэффициент корреляции по абсолютной величине будет высоким (положительным – в случае совпадения, отрицательным – в случае противоположной направленности трендов рядов x и y). Однако из этого нельзя делать вывод о том, что x причина y или наоборот. Высокий коэффициент корреляции в данном случае – это результат того, что x и y зависят от времени или содержат тренд. При этом одинаковый или противоположный тренд могут иметь ряды, совершенно не связанные друг с другом причинно-следственной зависимостью. В этом случае говорят о так называемой *ложной корреляции*, вызванной наличием тренда в каждом ряде. Фактор времени воздействует на зависимую и независимую переменную, но в модели не учитывается непосредственно. Влияние времени выражается в корреляционной зависимости между остатками, которая называется *автокорреляцией в остатках*. Избавление от тренда осуществляется с помощью одного из методов исключения тренда. Суть всех методов исключения тренда заключается в том, чтобы устранить или зафиксировать воздействие фактора времени на формирование уровней ряда. К основным методам относятся метод последовательных разностей, метод включения фактора времени в модель, метод отклонений от тренда.

Определим структуру ряда фактора X , исследуя автокорреляционную функцию, полученную на листе «Предварительный анализ рядов» (таблица 29).

Теория представлена в пункте 2.1 темы 3.

Таблица 29 – Анализ автокорреляционной функции фактора X

Период	R_x
1	0,99

2	0,97
3	0,94
4	0,91
5	0,89
6	0,87

Так как максимальное значение коэффициента корреляции имеет порядок 1, то сезонность временного ряда фактора X отсутствует.

Аналогично, временной ряд фактора Y не содержит сезонной составляющей (таблица 30).

Таблица 30 – Анализ автокорреляционной функции фактора Y

Период	Ry
1	0,91
2	0,82
3	0,77
4	0,73
5	0,72
6	0,77
макс	0,91

Устранение влияния фактора времени на каждую переменную в модели методом отклонений от тренда.

Метод отклонений от тренда предполагает построение регрессионной модели отклонений (наблюдаемых значений от трендовых) исследуемых факторов. Для парной регрессии, например, проводят аналитическое выравнивание временных рядов изучаемых показателей, рассчитывая параметры модели $\tilde{y}_t = f(t)$ и $\tilde{x}_t = f(t)$ по временным данным. Затем проводят расчет отклонений от трендов: $y_t - \tilde{y}_t$ и $x_t - \tilde{x}_t$. Для дальнейшего анализа используют не исходные данные, а отклонения от трендов, при условии, что последние не содержат тренда.

Анализ на автокорреляцию остатков регрессионной модели зависимости факторов X и Y от времени. Проанализируем остатки

временного ряда фактора X, полученного на листе «Предварительный анализ рядов», на наличие автокорреляции (таблица 31).

Для проверки гипотезы о наличии автокорреляции остатков используется кроме статистики Дарбина–Уотсона (см. тему 1) на больших выборках выборочный коэффициент автокорреляции остатков первого порядка $r_{e(t), e(t-1)}$, так как $DW \approx 2(1 - r_{e(t), e(t-1)})$.

Таблица 31 – Анализ модели X(t) на наличие автокорреляции остатков

Коэффициент корреляции остатков	0,75
tнабл	5,01
tкр	2,09

Так как коэффициент автокорреляции первого порядка значим (tнабл > tкр), то автокорреляция присутствует.

Для фактора Y автокорреляция отсутствует (таблица 32).

Таблица 32 – Анализ модели Y(t) на наличие автокорреляции остатков

Коэффициент корреляции остатков	0,42
tнабл	2,09
tкр	2,09

Устранение автокорреляции остатков во временном ряде X(t).

Поскольку во временном ряде фактора X присутствует автокорреляция, то модель этого ряда нельзя использовать в дальнейших исследованиях. Устраним автокорреляцию.

Для устранения автокорреляции можно воспользоваться *обобщенным методом наименьших квадратов*. Для применения ОМНК необходимо специфицировать модель автокорреляции регрессионных остатков. Обычно в качестве такой модели используется авторегрессионный процесс первого порядка AR(1). Для простоты изложения ограничимся случаем парной регрессии.

Пусть исходное уравнение регрессии $y = a + bx + \varepsilon$ содержит автокорреляцию случайных членов.

Допустим, что автокорреляция подчиняется авторегрес-

сионной схеме первого порядка $\varepsilon_i = \rho\varepsilon_{i-1} + u_i$, где ρ – коэффициент авторегрессии; u_i – случайный член, удовлетворяющий предпосылкам МНК.

Данная схема является *авторегрессионной*, так как ε определяется значениями этой же величины с запаздыванием, и *схемой первого порядка*, так как в этом случае запаздывание равно единице.

Величина ρ есть коэффициент корреляции между двумя соседними ошибками. Пусть ρ известно.

Обозначим $y_i^* = y_i - \rho y_{i-1}$, $x_i^* = x_i - \rho x_{i-1}$, $a^* = (1 - \rho)a$.

Это преобразование и называется *авторегрессионным преобразованием первого порядка AR(1)*, или *преобразованием Бокса–Дженкинса*.

Тогда преобразованное уравнение принимает вид

$$y_i^* = a^* + bx_i^* + u_i,$$

где $i \geq 2$. Это уравнение не содержит автокорреляцию и для оценки его параметров используется обычный МНК.

На практике величина ρ неизвестна. Наиболее простой способ оценить ρ – применить обычный МНК к регрессионному уравнению $\varepsilon_i = \rho\varepsilon_{i-1} + u_i$. Коэффициент ρ можно также приближенно оценить, используя статистику Дарбина–Уотсона: $DW \cong 2 \cdot (1 - \rho)$.

На листе «Устранение автокорреляции по X» получена модель $\varepsilon_i = \rho\varepsilon_{i-1} + u_i$ (таблица 33), в которой свободный член не значим (см. тему 1). Поэтому модель была уточнена (таблица 34).

Таблица 33 – Результаты расчета параметров модели случайной составляющей ряда фактора X

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение
Y-пересечение	-0,03	0,04	-0,56	0,58
Остатки по X	0,75	0,15	4,88	0,000 1

Таблица 34 – Результаты расчета параметров уточненной модели случайной составляющей ряда фактора X

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>
Y-пересечение	0	#Н/Д	#Н/Д	#Н/Д
Остатки по X	0,74	0,15	4,94	7,8E-05

Из таблицы 34 находим значение параметра $\rho = 0,74$. Используя авторегрессионное преобразование первого порядка AR(1), пересчитаны значения независимой переменной t^* и зависимой переменной X^* (таблица 35) и по данным значениям построена модель $X^* = 2,17 + 0,19 t^*$ (таблица 36), в которой отсутствует автокорреляция остатков (таблица 37). В дальнейшем будем использовать ее при построении модели методом отклонения от трендов.

Таблица 35 – Преобразованные значения переменных X^* и t^*

X^*	t^*
2,38	1,26
2,31	1,52
2,53	1,77
2,34	2,03
2,5	2,29
2,59	2,55
2,68	2,80
3,04	3,06
3,03	3,32
3,10	3,58
3,12	3,83
3,04	4,09
2,97	4,35
3,2	4,61
3,19	4,87
3,21	5,12

Окончание таблицы 35

X^*	t^*
3,60	5,38
2,76	5,64
3,14	5,9
3,24	6,15
3,43	6,41

Таблица 36 – Результаты расчета параметров модели $X^*(t^*)$

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>
Y-пересечение	2,18	0,12	18,27	1,64E-13
t^*	0,19	0,03	6,77	1,84E-06

Таблица 37 – Анализ модели $X^*(t^*)$ на наличие автокорреляции остатков

Коэффициент корреляции остатков	0,17
tнабл	0,75
tкр	2,09

Построение регрессионной модели по отклонениям от трендов.

Исходные данные (таблица 38) для построения модели по отклонениям и ее параметры (таблица 39) представлены на листе «Модель по отклонениям».

Таблица 38 – Отклонения факторов

Месяц	Остатки модели $X^*(t^*)$	Остатки модели $Y(t)$
2	-0,04	0,18
3	-0,16	-0,13
4	0,01	0,001
5	-0,23	-0,16
6	-0,13	-0,34
7	-0,09	-0,18
8	-0,04	-0,10
9	0,27	0,01
10	0,21	0,22
11	0,23	0,34

Окончание таблицы 38

Месяц	Остатки модели $X^*(t^*)$	Остатки модели $Y(t)$
12	0,20	-0,17
13	0,06	-0,02
14	-0,06	-0,04
15	0,12	-0,07
16	0,06	0,13

17	0,03	0,28
18	0,38	0,41
19	-0,52	-0,11
20	-0,19	-0,26
21	-0,13	-0,17
22	0,002	-0,04

Параметры модели $\Delta Y = f(\Delta X^*)$ представлены в таблице 39.

Таблица 39 – Результаты расчета параметров модели $\Delta Y = f(\Delta X^*)$

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>
Y-пересечение	-0,01	0,04	-0,27	0,79
Переменная ΔX^*	0,61	0,18	3,41	0,002

Проверка статистической значимости коэффициентов модели (см. тему 1) показывает, что модель требует уточнения.

Параметры уточненной модели $\Delta Y = f(\Delta X^*)$ представлены в таблице 40.

Таблица 40 – Результаты расчета параметров уточненной модели $\Delta Y = f(\Delta X^*)$

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>
Переменная ΔX^*	0,61	0,17	3,51	0,002
Y-пересечение	0	#Н/Д	#Н/Д	#Н/Д

В модели $\Delta Y = 0,61\Delta X^*$ автокорреляция отсутствует (таблица 41).

Таблица 41 – Анализ модели по отклонениям на наличие автокорреляции остатков

Коэффициент корреляции остатков	0,20
tнабл	0,9
tkp	2,09

Прогнозирование. Из уравнения модели $X = 0,19 t + 8,61$ для $X = 15$ находится значение $t = 33,53$.

Используя авторегрессионное преобразование первого порядка $AR(1)$ $t_i^* = t_i - \rho t_{i-1}$, при $\rho = 0,746$, $t_i = 33,53$, $t_{i-1} = 32,53$ находится $t^* = 9,25$. По модели $X^* = 0,19 t^* + 2,18$ вычисляется X^* модельное, равное 3,98. Для нахождения X^* набл используется авторегрессионное преобразование первого порядка $AR(1)$ $x_i^* = x_i - \rho x_{i-1} = 15 - 0,75 * 14,81 = 3,94$, где x_{i-1} рассчитывается по модели $X = 0,19 t + 8,61$ при $t-1 = 32,53$, $x_i = 15$ (из условия). Отклонение по фактору X определяется как разность между наблюдаемым и модельным значениями $\Delta X^* = X^*$ набл $- X^*$ модельное $= - 0,03$.

Используя модель $\Delta Y = 0,61 \Delta X^*$, находится $\Delta Y = 0,61 * (- 0,03) = - 0,02$. По модели $Y = 5,22 + 0,07 t$ при $t = 33,5$ находится прогнозируемое значение $Y = 7,62$. С учетом исключения трендов прогнозируемое значение $Y = 7,62 - 0,02 = 7,6$.

Все вычисления занесены в таблицу 42.

Таблица 42 – **Вычисление прогнозного значения**

t находится по модели $X(t)$ при $X = 15$	33,53
t^* находится из $AR(1)$	9,25
X^* модельное находится по модели $X^*(t^*)$	3,98
X при $t-1$ по модели $X(t)$	14,81
X из условия	15
X^* набл, используя $AR(1)$	3,94
Отклонение Δx^*	-0,03
Остаток по модели по отклонениям	-0,02
Y прогнозируемый по модели $Y(t)$	7,62
Y прогнозируемый с учетом исключения трендов	7,60

Вопросы для самоконтроля

1. Что понимают под ложной корреляцией?
2. Какова суть метода отклонений от тренда?
3. В каком случае используется ОМНК?
4. Каковы основные этапы ОМНК?
5. Как осуществляется прогнозирование?

Индивидуальные задания

Построить модель зависимости фактора Y от фактора X , предполагая, что на причинно-следственную связь между ними влияет ложная корреляция, используя данные наблюдений в таблице 28, которые необходимо изменить следующим образом: к каждому значению фактора X надо прибавить $0,3 \cdot \kappa$, где κ – номер студента в журнале группы. Сделать прогноз значения Y для $X = 17$.

Тема 5. СИСТЕМЫ ВЗАИМОЗАВИСИМЫХ УРАВНЕНИЙ

Постановка задачи

Имеются данные годового потребления свинины на единицу населения, представленные в таблице 43.

Таблица 43 – Данные наблюдений

Год	Годовое потребление свинины на единицу населения, кг (y_1)	Оптовая цена за 1 кг, усл. ед. (y_2)	Доход на душу населения, усл. ед. (x_1)	Отношение расходов по обработке мяса к цене, % (x_2)
2005	60	5,0	1 300	60
2006	62	4,0	1 400	56
2007	65	4,2	1 500	57
2008	63	5,0	1 600	63
2009	66	3,8	1 800	50
2010	67	4,3	1 700	58

Построить линейную структурную модель вида

$$\begin{cases} y_1 = f(y_2, x_1) = b_{12}y_2 + a_{11}x_1 + \varepsilon_1, \\ y_2 = f(y_1, x_2) = a_{22}x_2 + b_{21}y_1 + \varepsilon_2. \end{cases} \quad (1)$$

Для решения этой задачи необходимо:

1. По виду модели (1) определить количество эндогенных, экзогенных переменных в каждом уравнении системы и сделать вывод о ее идентифицируемости.

2. Используя косвенный МНК, оценить по выборочным данным, представленным в таблице 43, коэффициенты приведенной формы модели

$$\begin{cases} y_1 = \delta_{11}x_1 + \delta_{12}x_2, \\ y_2 = \delta_{21}x_1 + \delta_{22}x_2. \end{cases} \quad (2)$$

2.1. Методом МНК найти уравнения регрессии системы (2), используя тему 2.

2.2. По найденным приведенным коэффициентам рассчитать структурные коэффициенты, используя следующее правило:

$$\begin{cases} a_{11} = \delta_{11} - \frac{\delta_{12}\delta_{21}}{\delta_{22}}, & b_{12} = \frac{\delta_{12}}{\delta_{22}}, \\ a_{22} = \delta_{22} - \frac{\delta_{12}\delta_{21}}{\delta_{11}}, & b_{21} = \frac{\delta_{21}}{\delta_{11}}. \end{cases} \quad (3)$$

Записать структурные уравнения модели (1).

Технология вычислений в MS Excel для модели системы одновременных уравнений

1. Подсчитайте число эндогенных переменных в каждом уравнении системы:

- в первом уравнении две ($H_1 = 2$) эндогенные переменные: y_1, y_2 ;
- во втором уравнении тоже две ($H_2 = 2$) эндогенные переменные:

y_1, y_2 .

Подсчитайте отсутствующие экзогенные переменные в системе:

- в первом уравнении одна ($D_1 = 1$) переменная x_2 ;
- во втором уравнении тоже одна ($D_2 = 1$) переменная x_1 .

Проверьте условие идентифицируемости для каждого уравнения системы:

- для первого уравнения $D_1 + 1 = H_1$, т. е. $1 + 1 = 2$, значит, условие выполняется;

- для второго уравнения $D_2 + 1 = H_2$, т. е. $1 + 1 = 2$, значит, условие выполняется.

Следовательно, оба уравнения точно идентифицируемы. Поэтому система точно идентифицируема.

2. Постройте уравнения приведенной формы (2), воспользовавшись множественной регрессией (см. тему 2) для переменных, выраженных в отклонениях от среднего уровня (т. е. вместо x используйте разность $x - \bar{x}$, вместо y – разность $y - \bar{y}$) (рисунок 19).

	A	B	C	D	E
1	год	y1	y2	x1	x2
2	2005	60	5	1300	60
3	2006	62	4	1400	56
4	2007	65	4,2	1500	57
5	2008	63	5	1600	63
6	2009	66	3,8	1800	50
7	2010	67	4,3	1700	58
8					
9	Средние	=СРЗНАЧ(B2:B7)	=СРЗНАЧ(C2:C7)	=СРЗНАЧ(D2:D7)	=СРЗНАЧ(E2:E7)
10		y1	y2	x1	x2
11	Отклонения от средних	=B2-\$B\$9	=C2-\$C\$9	=D2-\$D\$9	=E2-\$E\$9
12		=B3-\$B\$9	=C3-\$C\$9	=D3-\$D\$9	=E3-\$E\$9
13		=B4-\$B\$9	=C4-\$C\$9	=D4-\$D\$9	=E4-\$E\$9
14		=B5-\$B\$9	=C5-\$C\$9	=D5-\$D\$9	=E5-\$E\$9
15		=B6-\$B\$9	=C6-\$C\$9	=D6-\$D\$9	=E6-\$E\$9
16		=B7-\$B\$9	=C7-\$C\$9	=D7-\$D\$9	=E7-\$E\$9

Рисунок 19 – Отклонения исходных данных от их средних значений

Параметры уравнений приведенной формы (2) представлены на рисунке 20.

20	ВЫВОД ИТОГОВ			ВЫВОД ИТОГОВ	
21					
22	<i>Регрессионная статистика</i>			<i>Регрессионная статистика</i>	
23	Множественный R	0,8732		Множественный R	0,89723
24	R-квадрат	0,7626		R-квадрат	0,80501
25	Нормированный	0,4532		Нормированный	0,50627
26	Стандартная ошибка	1,438		Стандартная ошибка	0,2506
27	Наблюдения	6		Наблюдения	6
28					
29	Дисперсионный анализ			Дисперсионный анализ	
30		<i>df</i>			<i>df</i>
31	Регрессия	2		Регрессия	2
32	Остаток	4		Остаток	4
33	Итого	6		Итого	6
34					
35		<i>Коэффициенты</i>		<i>Коэффициенты</i>	
36	Y-пересечение	0		Y-пересечение	0
37	x1	0,0118		x1	-0,00012
38	x2	-0,0445		x2	0,10172

а)

б)

Рисунок 20 – Итоги параметризации уравнений приведенной формы:

$$\text{а) } y_1 = \delta_{11}x_1 + \delta_{12}x_2; \text{ б) } y_2 = \delta_{21}x_1 + \delta_{22}x_2$$

Составьте матрицу системы приведенных уравнений из коэффициентов при переменных x_1, x_2 копированием соответствующих чисел (рисунок 21).

40		при x1	при x2
41	Матрица системы	0,0118026517083	-0,0444926058133
42		-0,0001241713411	0,1017210606833

Рисунок 21 – Матрица системы

Найдите структурные коэффициенты, используя формулы, приведенные на рисунке 22.

44	a11	=B41-B42*C41/C4	b12	=C41/C42
45	a22	=C42-B42*C41/B4	b21	=B42/B41

Рисунок 22 – Формулы расчета структурных коэффициентов по приведенным коэффициентам

3. Запишите уравнения системы в структурной форме.
4. Рассчитайте свободные члены для уравнений с исходными переменными по формулам, приведенным на рисунке 23.

Свободные члены уравнений	
a10	=B9-D45*C9-B45*D9
a20	=C9-D46*B9-B46*E9

Рисунок 23 – Формулы расчета свободных членов уравнений

5. Запишите структурную модель.

Эконометрический анализ построения модели системы одновременных уравнений

Для описания сложных экономических процессов недостаточно отдельных уравнений регрессий, так как изменение одной из переменных влечет за собой изменения во всей системе взаимосвязанных переменных. Поэтому возникает задача описания структуры связей между переменными несколькими уравнениями или системой структурных уравнений. При этом в одних уравнениях определенная переменная может рассматриваться как объясняющая, а в других – как зависимая (объясняемая) переменная. При рассмотрении систем эконометрических уравнений различают эндогенные (зависимые) и экзогенные (независимые) переменные.

В эконометрике рассматривается несколько видов систем

Структурная модель называется *идентифицированной*, если по коэффициентам приведенных уравнений можно однозначно определить значения коэффициентов структурных уравнений. Модель считается *идентифицируемой*, если каждое ее уравнение идентифицируемо, т. е. число параметров уравнения равно числу параметров его приведенной формы.

Структурная модель называется *неидентифицируемой*, если по коэффициентам приведенных уравнений можно получить несколько вариантов значений коэффициентов структурных уравнений, т. е. число приведенных коэффициентов хотя бы одного уравнения меньше числа его структурных коэффициентов.

Структурная модель называется *сверхидентифицируемой*, если по коэффициентам приведенных уравнений невозможно определить значения коэффициентов структурных уравнений, т. е. число приведенных коэффициентов хотя бы одного уравнения больше числа его структурных коэффициентов.

Необходимое условие идентифицируемости модели: для того чтобы уравнение было идентифицируемо, необходимо чтобы число экзогенных переменных, отсутствующих в данном уравнении, но присутствующих в системе, было равно числу эндогенных переменных в данном уравнении без одного. Это условие можно записать в виде следующего счетного правила:

$D + 1 = H$ – уравнение идентифицируемо;

$D + 1 < H$ – уравнение неидентифицируемо;

$D + 1 > H$ – уравнение сверхидентифицируемо;

где H – число эндогенных переменных в уравнении;

D – число экзогенных переменных, отсутствующих в уравнении, но присутствующих в системе.

Достаточное условие идентифицируемости модели: определитель матрицы, составленной из коэффициентов при переменных, отсутствующих в исследуемом уравнении, не равен нулю, и ранг этой матрицы не менее числа эндогенных переменных системы без единицы.

Для параметризации идентифицируемого уравнения применяется *косвенный метод наименьших квадратов*, для сверхидентифицированного – *двухшаговый метод наименьших*

квадратов. Эти методы дают состоятельные оценки параметров.

Косвенный МНК для идентифицируемой модели состоит в следующем:

- составляют приведенную форму модели и определяют численные значения параметров каждого ее уравнения обычным МНК по исходным данным;
- путем алгебраических преобразований полученной системы уравнений переходят к уравнениям в структурной форме модели, получая тем самым численные оценки структурных параметров.

Двухшаговый МНК для сверхидентифицируемой модели заключается в следующем:

- составляют приведенную форму модели и определяют численные значения параметров каждого ее уравнения обычным МНК по исходным данным;
- для сверхидентифицируемого уравнения получают теоретические значения эндогенных переменных, содержащихся в правой части уравнения, по соответствующим уравнениям приведенной формы. Далее, подставив их вместо фактических значений, обычным МНК определяют параметры структурного уравнения.

Напомним, что условия идентифицируемости выполняются для каждого уравнения системы:

- для первого уравнения $D_1 + 1 = H_1$, т. е. $1 + 1 = 2$;
- для второго уравнения $D_2 + 1 = H_2$, т. е. $1 + 1 = 2$.

Поэтому система точно идентифицируема.

Для построения системы уравнений найдем отклонения от средних, которые будем использовать в качестве новых переменных (названия сохраняются) (рисунок 24).

	A	B	C	D	E
1	год	y1	y2	x1	x2
2	2005	60	5	1300	60
3	2006	62	4	1400	56
4	2007	65	4,2	1500	57
5	2008	63	5	1600	63
6	2009	66	3,8	1800	50
7	2010	67	4,3	1700	58
8					
9	Средние	64	4	1550	57
10		y1	y2	x1	x2
11	Отклонения от ср	-4	1	-250	3
12		-2	0	-150	-1
13		1	0	-50	0
14		-1	1	50	6
15		2	-1	250	-7
16		3	0	150	1

Рисунок 24 – Значения отклонений от средних

По результатам параметризации (рисунок 20) имеем следующие уравнения приведенной формы:

$$\begin{cases} y_1 = 0,01 x_1 - 0,04 x_2, \\ y_2 = -0,0001 x_1 + 0,1 x_2. \end{cases}$$

По матрице коэффициентов системы приведенных уравнений из коэффициентов при переменных x_1 , x_2 найдены структурные коэффициенты (рисунок 25).

40		при x_1	при x_2	
41	Матрица коэффициентов	0,0118	-0,04449	
42	приведенной модели	-0,00012	0,10172	
43				
44	Коэффициенты структурной модели			
45	a11	0,01175	b12	-0,4374
46	a22	0,10125	b21	-0,01052

Рисунок 25 – Параметры приведенной и структурной моделей

Следовательно, уравнения структурной модели по отклонениям от средних значений имеют следующий вид:

$$\begin{cases} y_1 = -0,44 y_2 + 0,012 x_1, \\ y_2 = 0,1 x_2 - 0,01 y_1. \end{cases}$$

Перейдем от переменных в виде отклонений от среднего уровня к исходным переменным y и x . Свободные члены уравнений определим по следующим формулам:

$$a_{10} = \bar{y}_1 - b_{12}\bar{y}_2 - a_{11}\bar{x}_1 = 63,83 + 0,44 \cdot 4,38 - 0,012 \cdot 1550 = 47,54,$$

$$a_{20} = \bar{y}_2 - b_{21}\bar{y}_1 - a_{22}\bar{x}_2 = 4,38 + 0,01 \cdot 63,83 - 0,1 \cdot 57,33 = -0,75.$$

Таким образом, структурная модель имеет вид

$$\begin{cases} y_1 = 47,54 - 0,44 y_2 + 0,012 x_1, \\ y_2 = -0,75 - 0,1 x_2 + 0,01 y_1. \end{cases}$$

Оценка значимости модели дается через F -критерий и коэффициент детерминации для каждого уравнения в отдельности.

Вопросы для самоконтроля

1. Каковы основные причины использования систем одновременных уравнений?
2. В чем состоит основное различие между структурной и приведенной формами?
3. Почему не применим МНК для оценки структурных коэффициентов модели?
4. Для оценки каких систем возможно применение МНК?
5. Что понимают под экзогенными и эндогенными переменными?
6. Что представляет собой сверхидентифицируемая система?
7. Как сформулировать необходимое условие идентификации?
8. Как сформулировать достаточное условие идентификации?
9. В чем состоит суть двухшагового метода наименьших квадратов?
10. Этапы косвенного метода наименьших квадратов.
11. Существует ли единый критерий для оценки общего качества всей системы одновременных уравнений в целом?
12. Какой вид имеет модель «спрос – предложение»?
13. Какой метод оценок параметров целесообразен для точно идентифицируемого уравнения?
14. Какова система рекурсивных уравнений?

Индивидуальные задания

Имеются данные годового потребления свинины на единицу населения, представленные в таблице 44.

Таблица 44 – Данные к индивидуальным заданиям (к – номер студента в журнале)

Год	Годовое потребление свинины на единицу населения, кг (y_1)	Оптовая цена за 1 кг, усл. ед. (y_2)	Доход на душу населения, усл. ед. (x_1)	Отношение расходов по обработке мяса к цене, % (x_2)
2005	$60 + k$	$5,0 + 0,1k$	$1\,300 + 100k$	$60 + k$
2006	$62 + k$	$4,0 + 0,1k$	$1\,400 + 100k$	$56 + k$
2007	$65 + k$	$4,2 + 0,1k$	$1\,500 + 100k$	$57 + k$
2008	$63 + k$	$5,0 + 0,1k$	$1\,600 + 100k$	$63 + k$
2009	$66 + k$	$3,8 + 0,1k$	$1\,800 + 100k$	$50 + k$
2010	$67 + k$	$4,3 + 0,1k$	$1\,700 + 100k$	$58 + k$

Построить линейную структурную модель вида

$$\begin{cases} y_1 = f(y_2, x_1) = b_{12}y_2 + a_{11}x_1 + \varepsilon_1, \\ y_2 = f(y_1, x_2) = a_{22}x_2 + b_{21}y_1 + \varepsilon_2. \end{cases}$$

ПРИЛОЖЕНИЯ

Приложение А

Результаты расчетов по теме «Парная регрессия»

	A	B	C	D	E	F
1	ВВП	Экспорт		Ковариация		
2		1000	190		ВВП	Экспорт
3		1090	220	ВВП	98568,75	
4		1150	240	Экспорт	25558,75	6804,75
5		1230	240			
6		1300	260	Корреляция		
7		1360	250		ВВП	Экспорт
8		1400	280	ВВП	1	
9		1470	290	Экспорт	0,986879	1
10		1500	310			
11		1580	350	тнал.	25,93168	
12		1600	340	ткр.	2,100922	
13		1630	360			
14		1700	380			
15		1780	400			
16		1800	420			
17		1850	400			
18		1910	400			
19		1990	440			
20		2010	450			
21		2100	470			
22						
23	Среднее ВВП	Среднее экспорт				
24	1572,5	334,5				
25	Коэффициент ВВП	0,2592987				
26	Коэффициент эластичности	0,0551577				
27	И Регрессия Исходные данные					

Рисунок А.1 – Лист «Исходные данные»

	A	B	C	D	E	F	G	H	I	
1	ВЫВОД ИТОГОВ									
2										
3	Регрессионная статистика									
4	Множественный	0,98687898								
5	R-квadrat	0,97393012								
6	Нормированный	0,97248179								
7	Стандартная ошибка	14,03956878								
8	Наблюдения	20								
9										
10	Дисперсионный анализ									
11		σ^2	SS	MS	F	Значимость F				
12	Регрессия	1	132547,02	132547,0195	672,46194	1,04396E-15				
13	Остаток	18	3547,98047	197,1100261						
14	Итого	19	136095							
15				Фкр.	4,4138734					
16		Коэффициент стандартной ошибки		t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95,0%	Верхние 95,0%	
17	У-пересечение	-73,247226	16,0342342	-4,568177389	0,00023836	-106,9339018	-39,56055	-106,9339	-39,56055	
18	ВВП	0,25929871	0,0099993	25,93167832	1,044E-15	0,238290956	0,28030647	0,23829096	0,28030647	
19			ткр.	2,100922037						
20										
21										
22	ВЫВОД ОСТАТКА									
23										
24		Условие3			Точечный прогноз					
25	Наблюдение	казанное	Экст	Остатки	Числителе	Знам.	674,9995561			
26	1	186,051487		3,94851309	44,39711277	15,5907557				
27	2	209,388371		10,6116289	19,73205012	112,606669	Интервальный прогноз			
28	3	224,946294		15,0537062	430,3092638	226,614069	539,0388423	610,96027		
29	4	246,690191		-5,6901909	3,419134206	32,378272				
30	5	263,841101		-3,8411008	653,2074163	14,754055				
31	6	279,399024		-29,399024	385,2604052	864,302584				
32	7	289,921882		-9,770972	66,43733216	95,4718946				
33	8	307,921882		-17,921882	149,3537848	321,193852				
34	9	315,700843		-5,7008433	370,7975017	32,4996146				
35	10	336,44474		13,5552597	230,6138141	183,745064				
36	11	341,630715		-1,6307146	149,3537848	2,65923012				
37	12	349,409676		10,590324	3,419134206	112,154963				
38	13	367,560586		12,4394141	0,553382786	154,739024				
39	14	388,304483		11,6955171	219,4653587	136,78512				
40	15	393,490467		26,5095428	1086,686982	702,755861				
41	16	406,465393		-6,4653928	242,0489609	41,6720963				
42	17	422,013316		-22,013316	370,7975017	484,586063				
43	18	442,757213		-2,7572126	23,17484386	7,60222135				
44	19	447,943187		2,05681314	11,13479586	4,23048028				
45	20	471,280071		-1,280071		1,63858181				
46			суммы		4460,15256	3547,98047				
47				днабл.		1,25709614				
48										

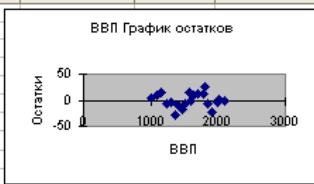


Рисунок А.2 – Лист «Регрессия»

	A	B	C	D	E	F	G
1	Карман	Частота	0				
2	-29,399	1	0,015723	0,015723	0,314454493	1,494565	
3	-15,4219	2	0,129542	0,11382	2,276392379	0,033559	
4	-1,44474	7	0,4579	0,328358	6,56716186	0,028528	
5	12,5324	7	0,820457	0,362557	7,251130767	0,008697	
6	79,52863	3	1	0,179543	3,590860442	0,097224	
7					хи-кв набл.	1,662573	
8					хи-кв кр.	7,814728	
9							
10							
11							
12	Остатки		Остатки				
13			3,948513				
14	Среднее	-1,7E-14	10,61163	тнабл.	-5,58088E-15		
15	Стандартн	3,055616	15,05371	ткр.	2,09302405		
16	Медиана	-1,45539	-5,69019				
17	Мода	#Н/Д	-3,8411				
18	Стандартн	13,66513	-29,399				
19	Дисперсия	186,7358	-9,77097				
20	Эксцесс	0,076676	-17,9219				
21	Асимметри	-0,29745	-5,70084				
22	Интервал	55,90857	13,55526				
23	Минимум	-29,399	-1,63071				
24	Максимум	26,50954	10,59032				
25	Сумма	-3,4E-13	12,43941				
26	Счет	20	11,69552				
27	Уровень н	6,395478	26,50954				
28			-6,45539				
29			-22,0133				
30			-2,75721				
31			2,056813				
32			-1,28007				

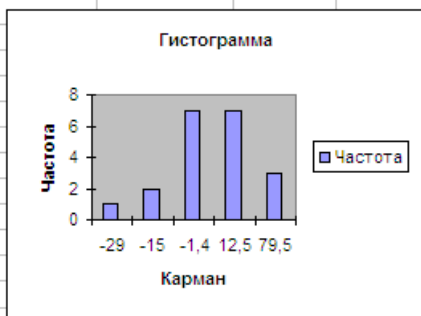


Рисунок А.3 – Лист «Условие 1-нормальность»

	A	B	C	D	E	F	G	H	I	J	K	L
1	Экспорт	Остатки	Модуль ост.	Точка	Экспорт	Ранг	Процент	Точка	мод ост	Ранг	Процент	Квадрат разности рангов
2	190	3,9485131	3,94851309	1	190	20	0,00%	1	3,948513	15	26,30%	25
3	220	10,611629	10,6116289	2	220	19	5,20%	2	10,61163	9	57,80%	100
4	240	15,053706	15,0537062	3	240	17	10,50%	3	15,05371	5	78,90%	144
5	240	-5,690191	5,69019086	4	240	17	10,50%	4	5,690191	14	31,50%	9
6	260	-3,8411101	3,841110075	5	260	15	26,30%	5	3,8411101	16	21,00%	1
7	250	-29,39902	29,3990235	6	250	16	21,00%	6	29,39902	1	100,00%	225
8	280	-9,770972	9,77097204	7	280	14	31,50%	7	9,770972	11	47,30%	9
9	290	-17,92188	17,9218819	8	290	13	36,80%	8	17,92188	4	84,20%	81
10	310	-5,700843	5,70084332	9	310	12	42,10%	9	5,700843	13	36,80%	1
11	350	13,55526	13,5552597	10	350	10	52,60%	10	13,55526	6	73,60%	16
12	340	-1,630715	1,6307146	11	340	11	47,30%	11	1,630715	19	5,20%	64
13	360	10,590324	10,590324	12	360	9	57,80%	12	10,59032	10	52,60%	1
14	380	12,439414	12,4394141	13	380	8	63,10%	13	12,43941	7	68,40%	1
15	400	11,695517	11,6955171	14	400	5	68,40%	14	11,69552	8	63,10%	9
16	420	26,509543	26,5095428	15	420	4	84,20%	15	26,50954	2	94,70%	4
17	400	-6,455393	6,45539281	16	400	5	68,40%	16	6,455393	12	42,10%	49
18	400	-22,01332	22,0133156	17	400	5	68,40%	17	22,01332	3	89,40%	4
19	440	-2,757213	2,75721261	18	440	3	89,40%	18	2,757213	17	15,70%	196
20	450	2,0568131	2,05681314	19	450	2	94,70%	19	2,056813	18	10,50%	256
21	470	-1,280071	1,28007102	20	470	1	100,00%	20	1,280071	20	0,00%	361
22											Коэффициент ранговой корреляции	-0,169925
23											тнбл.	-0,740685
24											ткр.	2,100922

Рисунок А.4 – Лист «Условие 2»

Примерный макет отчета по теме «Парная регрессия»

**ОТЧЕТ
по теме «Парная регрессия»**

студента _____ гр. _____

1. Постановочный этап

Из экономической теории известно, что экспорт зависит от ВВП и многих других факторов, например, от _____.

Выделим один фактор – ВВП, который является наиболее существенным. Он называется объясняющим фактором для результативного (объясняемого) фактора – экспорта. Поэтому возникает задача количественного описания зависимости указанных экономических показателей уравнением парной регрессии.

2. Спецификация модели

Вид регрессии визуально определяется по корреляционному полю, которое изображено на графике подбора черными точками по данным 20 наблюдений. Так как точки сгруппированы вдоль прямой (не горизонтальной), то можно предположить, что зависимость экспорта (y) от ВВП (x) описывается парной линейной регрессионной моделью $y = b_0 + b_1x + \varepsilon$, где b_0, b_1 – неизвестные параметры модели; ε – случайный член, который включает в себя суммарное влияние всех неучтенных в модели факторов.

Ковариация равна _____, поэтому зависимость (*прямая или обратная*) _____.

Коэффициент корреляции равен _____. Так как он (*больше или меньше*) _____ нуля, то зависимость (*прямая или обратная*) _____. (*Вывод о силе линейной зависимости*) _____, так как _____.

Корреляция _____ (*значима/незначима*), так как $|t_{\text{набл}}| = | \dots \dots t_{\text{кр}} = \dots \dots$.
(*знак*)

3. Параметризация модели

Для оценки параметров уравнения парной регрессии применяют МНК. В результате проведения регрессионного анализа получены точечные и интервальные оценки неизвестных параметров.

Точечная оценка параметра b_0 равна _____. Интервальная оценка равна (_____, _____).

Точечная оценка параметра b_1 равна _____. Интервальная оценка равна (_____, _____).

Таким образом, уравнение регрессии имеет вид _____.

4. Верификация модели

4.1. Качество построенной модели в целом оценивает коэффициент (индекс) детерминации, который представляет собой долю дисперсии, объясненной выбранным фактором.

Коэффициент детерминации R-квадрат равен _____. (Вывод об общем качестве уравнения) _____.

Значимость коэффициента детерминации R-квадрат устанавливается с помощью критерия Фишера. Так как наблюдаемое значение $F_{\text{набл}} = \frac{\text{MSR}}{\text{MSE}}$, $F_{\text{кр}} = \frac{\text{MSR}_{\text{кр}}}{\text{MSE}}$, то R-квадрат _____ (знак) _____ (значим или незначим).

Нормированный индекс детерминации \bar{R}^2 позволяет более точно определить качество модели.

Нормированный индекс множественной детерминации равен _____ . (Вывод об общем качестве уравнения) _____.

4.2. Сделать вывод о нормальности распределения остатков можно: 1) по гистограмме остатков; 2) по числовым характеристикам асимметрии и эксцессу; 3) по критерию Пирсона.

1. Соединим середины верхних сторон прямоугольников гистограммы. Так как полученная ломаная линия _____, то остатки распределены _____ (по нормальному или не по нормальному закону). (Вывод о выполнении условия нормальности остатков по визуальному анализу гистограммы) _____.

2. Так как асимметрия равна _____, эксцесс равен _____, то (вывод о выполнении условия нормальности остатков по анализу числовых характеристик) _____.

3. Так как хи-квадрат наблюдаемое равно _____ (меньше или больше) хи-квадрат критического, равного _____, то остатки распределены (по нормальному или не по нормальному за-

кону) _____.

4.3. Значимость коэффициента регрессии b_0 оценивается с помощью t -статистики. Наблюдаемое значение статистики $t_{\text{набл}} =$ _____. Критическое значение $t_{\text{кр}} =$ _____. Так как $|t_{\text{набл}}| = |$ _____ $| \dots t_{\text{кр}} =$ _____, то коэффициент b_0 _____
(знак)
_____ (значим или незначим).

Значимость коэффициента регрессии b_1 оценивается с помощью t -статистики. Наблюдаемое значение статистики $t_{\text{набл}} =$ _____. Критическое значение $t_{\text{кр}} =$ _____. Так как $|t_{\text{набл}}| = |$ _____ $| \dots t_{\text{кр}} =$ _____, то коэффициент b_1 _____
(знак)
(значим или незначим). (Вывод о спецификации) _____

_____ .
(Вывод об общем качестве уравнения) _____

4.4. Проверка статистических свойств остатков. Для того чтобы оценки параметров линейного уравнения множественной регрессии были несмещенными, состоятельными и эффективными, необходимо выполнение следующих условий Гаусса–Маркова для случайной переменной: 1) $M(\varepsilon) = 0$; 2) постоянство дисперсии; 3) отсутствие автокорреляции.

4.4.1. Среднее остатков является оценкой математического ожидания случайного члена. Оно равно _____. Значимость среднего устанавливается с помощью критерия Стьюдента. Так как $|t_{\text{набл}}| = |$ _____ $| \dots t_{\text{кр}} =$ _____, то среднее _____
(знак)

(значимо или незначимо).
(Вывод о выполнении условия 1) _____

4.4.2. Для проверки условия гомоскедастичности остатков согласно схеме теста ранговой корреляции Спирмена найдены коэффициент ранговой корреляции Спирмена и наблюдаемое значение t -статистики. Так как $|t_{\text{набл}}| = |$ _____ $| \dots t_{\text{кр}} =$ _____, то _____
(знак)

_____ (вывод о наличии гетероскедастичности или гомоскедастичности).

(Вывод о выполнении условия 2) _____

4.4.3. Наблюдаемое значение d -статистики равно _____. По

таблице критических значений d -статистики для числа наблюдений _____, числа объясняющих переменных _____ и заданного уровня значимости 0,05 значения $d_1 =$ _____ и $d_2 =$ _____. Они разбивают отрезок $[0; 4]$ на пять областей:

Положительная автокорреляция	Область неопределенности	Автокорреляция отсутствует	Область неопределенности	Отрицательная автокорреляция
0	_____	_____	_____	4

Так как наблюдаемое значение критерия Дарбина–Уотсона $d_{набл} =$ _____ попало в область _____, то _____ (вывод о наличии автокорреляции) _____.

Так как на графике остатков точки _____ (описать расположение), то _____ (вывод о наличии зависимости между остатками).
(Вывод о выполнении условия 3) _____.

4.5. Так как коэффициент эластичности равен _____, то _____ (вывод о влиянии независимого фактора на зависимый).

Общий вывод этапа верификации: _____ (вывод о качестве модели и возможности прогнозировать по ней).

5. Прогнозирование

Точечный прогноз экспорта равен _____, интервальный прогноз равен (_____, _____). (Вывод о качестве прогноза) _____.

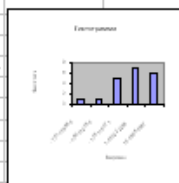
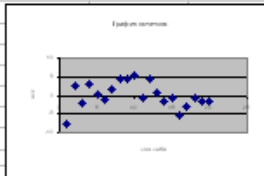
Замечания по модели _____.

Результаты расчетов по теме «Множественная регрессия»

	A	B	C	D	E	F	G	H
1	ЗП	возраст	стаж		Корреляционная матрица			
2	684,48	28	10			ЗП	возраст	стаж
3	674,45	26	8		ЗП	1		
4	729,62	30	14		возраст	0,794825	1	0,747928
5	748,86	35	15		стаж	0,995338	0,747928	1
6	761,44	41	16		Значимость коэффициентов корреляции			
7	773,42	45	17		тэл, в набл	5,556924		
8	628,07	27	3		тэл, с набл	43,7846		
9	731,84	35	13		ткр.	2,100922		
10	698,81	30	10					
11	645,92	23	5		Мультиколлинеарность			
12	664,64	29	7		Определитель	0,440604		
13	711,18	33	11		хи-кв набл.	19,53393		
14	798,07	40	20		хи-кв кр.	223,1602		
15	833,82	41	24					
16	667,97	41	6		Частные коэффициенты корреляции			
17	607,61	23	2		rЗП,В-С	0,787012		
18	711,76	32	12		rЗП,С-В	0,995181		
19	728,62	37	13		тнабл (rЗП,В-С)	5,412226		
20	666,15	31	7		тнабл (rЗП,С-В)	43,06106		
21	683,7	30	9					
22								
23	Эластичность							
24	ЗП ср	Возраст ср	Стаж ср					
25	707,5215	32,85	11,1					
26	Коэффициент возраст	1,029555						
27	Эластичность возраст	0,0478019						
28	Коэффициент стаж	9,2757834						
29	Эластичность стаж	0,1455238						
30								

Рисунок В.1 – Лист «Исходные данные»

	A	B	C	D	E	F	G	H
1	ВЫВОД ИТОГОВ				Точечный прогноз			
2					699,5316814			
3	Дисперсионная статистика							
4	Множественность	0,998228						
5	R-квадрат	0,99646						
6	Нормированная	0,996043						
7	Стандартная	3,597326						
8	Наблюдения	20						
9								
10	Дисперсионный анализ							
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>		
12	Регрессия	2	61917,595	30958,79731	2392,348568	1,4688E-21		
13	Остаток	17	219,99284	12,94075525				
14	Итого	19	62137,587					
15				Фкр.	3,591530569			
16		Коэффициент стандартная ошибка статистика		P-значение	Нижние 95%	Верхние 95%	Нижние 95,0	Верхние 95,0
17	У-пересеч	570,7394	4,9378112	115,5855106	4,62747E-26	560,321553	581,157294	560,3216
18	возраст	1,029555	0,1957426	5,259738186	6,38714E-05	0,61657413	1,44253581	0,616574
19	стаж	9,275783	0,2216551	41,8478256	1,37064E-18	8,80813205	9,74343475	8,808132
20			tкр	2,109815559				
21								
22								
23	ВЫВОД ОСТАТКА							
24								
25	Наблюдения	коэффициент	Остатки	Условие 1		Условие 3		
26	1	692,3248	-7,8447966		Остатки	111,95072	61,5408344	
27	2	671,7141	2,7358801			21,1868753	7,48503993	
28	3	731,487	-1,8670402	Среднее	-2,84217E-14	23,198111	3,4858391	
29	4	745,9106	2,9494015	Стандартная	0,760873527	8,25477964	8,69896946	
30	5	761,3637	0,0762883	Медиана	-0,717009631	1,9994053	0,00581991	
31	6	774,7577	-1,3377115	Мода	#Н/Д	9,25958803	1,78948136	
32	7	626,3648	1,7052421	Стандартное	3,402729859	7,70465596	2,90785076	
33	8	727,359	4,4809683	Дисперсия в	11,57857049	0,00301126	20,0790773	
34	9	694,3839	4,4260934	Экссесс	0,102240561	0,48414016	19,5903029	
35	10	640,7981	5,1218952	Асимметричн	-0,360269088	36,1068388	26,2338107	
36	11	665,527	-0,8870014	Интервал	12,96669187	28,2880006	0,7867715	
37	12	706,7484	4,4316461	Минимум	-7,844796646	14,4319103	19,6394782	
38	13	797,4373	0,6327097	Максимум	5,121895223	5,67720487	0,40032153	
39	14	835,57	-1,7499789	Сумма	-5,68434E-13	1,24122158	3,06242618	
40	15	668,6059	-0,6358777	Счет		20	0,4043404	
41	16	612,9708	-5,3607546	tнабл	-3,73541E-14	4,52060411	28,7376896	
42	17	714,9946	-3,2345833	tкр	2,09302405	5,93624835	10,4625294	
43	18	729,4181	-0,7981416			0,40700541	0,63703001	
44	19	667,5861	-1,4361114			0,00078334	2,06241582	
45	20	685,1081	-1,4081232				1,98281091	
46						302,975556	219,992839	
47						d набл	1,37720649	
48								
49	Критерий Пирсона							
50	Карман	Частота	0					
51	-7,844797	1	0,0105709	0,010570913	0,211418264	2,94137859		
52	-4,603124	1	0,088064	0,077493045	1,549860904	0,1950801		
53	-1,361461	5	0,3446394	0,256475472	5,129509443	0,00326984		
54	1,880222	7	0,7097186	0,365179191	7,303583821	0,01261889		
55	15,36569	6	0,9999968	0,290278223	5,805564458	0,00651189		
56								
57	хи кв наб	3,158859						
58	хи кв кр	7,814728						
59								
60								



10 < > >> Регрессия \ Исходные данные \ Условие 2 \ Интервальный_пр

Рисунок В.2 – Лист «Регрессия»

	A	B	C	D	E	F	G	H	I	J	K	L
1	ЗП	Остатки	Мод ост.	Точка	ЗП	Ранг	Процент	Точка	Мод ост.	Ранг	Процент	Квадрат разности рангов
2	684,48	-7,8447966	7,844797	1	684,48	12	42,10%	1	7,844797	1	100,00%	121
3	674,45	2,7358801	2,73588	2	674,45	14	31,50%	2	2,73588	9	57,80%	25
4	729,62	-1,8670402	1,86704	3	729,62	7	68,40%	3	1,86704	10	52,60%	9
5	748,86	2,94940154	2,949402	4	748,86	5	78,90%	4	2,949402	8	63,10%	9
6	761,44	0,07628831	0,076288	5	761,44	4	84,20%	5	0,076288	20	0,00%	256
7	773,42	-1,337715	1,337715	6	773,42	3	89,40%	6	1,337715	15	26,30%	144
8	628,07	1,70524214	1,705242	7	628,07	19	5,20%	7	1,705242	12	42,10%	49
9	731,84	4,48096835	4,480968	8	731,84	6	73,60%	8	4,480968	4	84,20%	4
10	698,81	4,42609341	4,426093	9	698,81	11	47,30%	9	4,426093	6	73,60%	25
11	645,92	5,12189522	5,121895	10	645,92	18	10,50%	10	5,121895	3	89,40%	225
12	664,64	-0,8870014	0,887001	11	664,64	17	15,70%	11	0,887001	16	21,00%	1
13	711,18	4,43164509	4,431645	12	711,18	10	52,60%	12	4,431645	5	78,90%	25
14	798,07	0,63270967	0,63271	13	798,07	2	94,70%	13	0,63271	19	5,20%	289
15	833,82	-1,7499789	1,749979	14	833,82	1	100,00%	14	1,749979	11	47,30%	100
16	667,97	-0,6358777	0,635878	15	667,97	15	26,30%	15	0,635878	18	10,50%	9
17	607,61	-5,3607546	5,360755	16	607,61	20	0,00%	16	5,360755	2	94,70%	324
18	711,76	-3,2345833	3,234583	17	711,76	9	57,80%	17	3,234583	7	68,40%	4
19	728,62	-0,7981416	0,798142	18	728,62	8	63,10%	18	0,798142	17	15,70%	81
20	666,15	-1,4361114	1,436111	19	666,15	16	21,00%	19	1,436111	13	36,80%	9
21	683,7	-1,4081232	1,408123	20	683,7	13	36,80%	20	1,408123	14	31,50%	1
22											тнбп.	-1,2453997
23											ткр.	2,10092204

Рисунок В.3 – Лист «Условие 2»

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Массив 1			Массив 2																
2	1	28	10	1																
3	1	26	8	35																
4	1	30	14	10																
5	1	35	15																	
6	1	41	16																	
7	1	45	17																	
8	1	27	3																	
9	1	35	13																	
10	1	30	10																	
11	1	23	5																	
12	1	29	7																	
13	1	33	11																	
14	1	40	20																	
15	1	41	24																	
16	1	41	6																	
17	1	23	2																	
18	1	32	12																	
19	1	37	13																	
20	1	31	7																	
21	1	30	9																	
22	Массив 2 трансп.																			
23	1	35	10																	
24	Массив 1 трансп																			
25	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
26	28	26	30	35	41	45	27	35	30	23	29	33	40	41	41	23	32	37	31	30
27	10	8	14	15	16	17	3	13	10	5	7	11	20	24	6	2	12	13	7	9
28	Массив 3 (произведение массива 1 трансп. и массива 1)																			
29	20	657	222																	
30	657	22349	7799																	
31	222	7799	3062																	
32	Массив 4 (обратный к массиву 3)																			
33	1,88412	-0,06943	0,04023																	
34	-0,06943	0,00296	-0,00251																	
35	0,04023	-0,00251	0,0038																	
36	Массив 5 (произведение массива 2 трансп. и массива 4)																			
37	-0,14353	0,00912	-0,00957																	
38	Массив 6 (произведение массива 5 и массива 2)																			
39	0,08014																			
40	Стандартная ошибка прогноза																			
41	1,01838																			
42	Интервальный прогноз																			
43	697,383	701,68																		

Рисунок В.4 – Лист «Интервальный прогноз»

Расчет прогноза по взаимосвязанным временным рядам

Лист «Прогноз»

	A	B	C	D	E
1	Прогноз Y для X=15				
2	Модель временного ряда X(t)			Модель временного ряда Y(t)	
3		<i>Кoeffициенты</i>		<i>Кoeffициенты</i>	
4	Y-пересечение	8,60654545		Y-пересечение	5,216662338
5	Месяц	0,19066008		Месяц	0,071788255
6					
7	Модель временного ряда X*(t*)				
8		<i>Кoeffициенты</i>			
9	Y-пересечение	2,176939			
10	t*	0,19474015			
11					
12	Модель авторегрессии временного ряда X(t)				
13		<i>Кoeffициенты</i>			
14	Y-пересечение	-0,0252045			
15	Остатки по X	0,74656599			
16					
17	Модель по отклонениям				
18		<i>Кoeffициенты</i>			
19	Y-пересечение	0			
20	Остатки модели X*(t*)	0,613656			
21					
22	t находится по модели X(t) при X=15	33,5332628			
23	t* находится из AR(1)	9,2450353			
24	X* модельное находится по модели X*(t)	3,97731854			
25	X при t-1 по модели X(t)	14,8093399			
26	X из условия	15			
27	X* набл., используя AR(1)	3,9438505			
28	отклонение Δx*	-0,033468			
29	остаток по модели по отклонениям	-0,0205379			
30	Yпрогнозируемый по модели Y(t)	7,62395677			
31	Yпрогнозируемый с учетом исключения трендов	7,60341891			
32					
33					

СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

1. **Аистов, А. В.** Эконометрика шаг за шагом : учеб. пособие для вузов / А. В. Аистов, А. Г. Максимов. – М. : ГУ ВШЭ, 2006. – 177 с.
2. **Айвазян, С. А.** Прикладная статистика и основы эконометрики : учеб. для вузов / С. А. Айвазян, В. С. Мхитрян. – М. : ЮНИТИ, 1998. – 1022 с.
3. **Айвазян, С. А.** Прикладная статистика. Исследование зависимостей / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. – М. : Финансы и статистика, 1985.
4. **Айвазян, С. А.** Прикладная статистика. Основы моделирования и первичная обработка данных / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. – М. : Финансы и статистика, 1983.
5. **Алексеевко, В. Б.** Математические методы исследования экономических систем : учеб. пособие для вузов / В. В. Алексеевко, В. В. Красавина. – М. : Изд-во РУДН, 2005. – 154 с.
6. **Бородич, С. А.** Эконометрика : учеб. пособие для вузов / С. А. Бородич. – Минск : Новое знание, 2001. – 408 с.
7. **Грубер, Й.** Эконометрия : учеб. пособие. В 2 т. Т. 2 : Эконометрические прогнозные и оптимизационные модели / Й. Грубер. – Киев : Нічлава, 1999. – 308 с.
8. **Доугерти, К.** Введение в эконометрику : учеб. пособие для вузов / К. Доугерти : [пер. с англ.] / К. Доугерти ; под ред. О. О. Замкова. – М. : ИНФРА-М, 1997. – 402 с.
9. **Елисеева, И. И.** Эконометрика : учеб. / И. И. Елисеева [и др.] ; под ред. И. И. Елисеевой. – М. : Финансы и статистика, 2006. – 576 с.
10. **Карминский, А. М.** Рейтинги в экономике: методология и практика : учеб. пособие / А. М. Карминский [и др.] ; под ред. А. М. Карминского. – М. : Финансы и статистика, 2005. – 240 с.
11. **Краасс, М. С.** Математические методы и модели для магистрантов и экономистов : учеб. пособие для вузов / М. С. Краасс, Б. П. Чупрынов. – СПб. : Питер, 2006. – 496 с.
12. **Калмыкова, Т. Ф.** Анализ взаимосвязи экономических показателей : учеб.-метод. пособие / Т. Ф. Калмыкова, Т. М. Моисеева. – Гомель : Бел. торгово-экон. ун-т потребит. кооп., 2006. – 52 с.
13. **Магнус, Я. Р.** Эконометрика. Начальный курс : учеб. пособие / Я. Р. Магнус [и др.]. – М. : Дело, 1997. – 248 с.
14. **Маленво, Э.** Статистические методы эконометрии / Э. Маленво. – М. : Статистика, 1976.
15. **Мардас, А. Н.** Эконометрика : учеб. пособие для вузов / А. Н. Мардас. – СПб. : Питер, 2001. – 144 с.
16. **Марченко, Л. Н.** Теория вероятностей и математическая статистика : практ. пособие / Л. Н. Марченко, Л. П. Авдашкова. – Гомель : ГГУ им. Ф. Скорины, 2007. – 275 с.
17. **Нименья, И. Н.** Эконометрика : учеб. пособие / И. Н. Нименья. – СПб. : Нева, 2004. – 224 с.
18. **Новак, Э.** Введение в методы эконометрики : сб. задач / Э. Новак ; пер. с пол. под ред. И. И. Елисеевой. – М. : Финансы и статистика, 2006. – 248 с.
19. **Переяслова, И. Г.** Статистика : учеб. пособие для вузов / И. Г. Переяслова, Е. Б. Колбачев, О. Г. Переяслова. – Ростов н/Д : Феникс, 2003. – 288 с.
20. **Переяслова, И. Г.** Основы статистики : учеб. пособие для вузов / И. Г. Переяслова, Е. Б. Колбачев. – Ростов н/Д : Феникс, 1999. – 320 с.
21. **Шелобаев, С. И.** Математические методы и модели в экономике, финансах, бизнесе : учеб. пособие для вузов / С. И. Шелобаев. – М. : ЮНИТИ-ДАНА, 2000. – 367 с.
22. **Экономико-математические методы и прикладные модели** : учеб. пособие для вузов / В. В. Федосеева [и др.] ; под ред. В. В. Федосеевой. – М. : ЮНИТИ-ДАНА, 2005. – 304 с.

Учебное издание

**ЭКОНОМЕТРИКА
И ЭКОНОМИКО-МАТЕМАТИЧЕСКИЕ
МЕТОДЫ И МОДЕЛИ**

Пособие

для студентов экономических специальностей

Авторы-составители:

Авдашкова Людмила Павловна

Грибовская Марал Атаевна

Калмыкова Тамара Федоровна и др.

Редактор **И. А. Михайлова**

Технический редактор **И. А. Козлова**

Компьютерная верстка **Н. Н. Короедова**

Подписано в печать 02.04.12. Бумага типографская № 1.

Формат 60 × 84 ¹/₁₆. Гарнитура Таймс. Ризография.

Усл. печ. л. 6,74. Уч.-изд. л. 7,10. Тираж 200 экз.

Заказ №

Учреждение образования

«Белорусский торгово-экономический университет
потребительской кооперации».

246029, г. Гомель, просп. Октября, 50.

ЛИ № 02330/0494302 от 04.03.2009 г.

Отпечатано в учреждении образования

«Белорусский торгово-экономический университет
потребительской кооперации».

246029, г. Гомель, просп. Октября, 50.

**БЕЛКООПСОЮЗ
УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ
«БЕЛОРУССКИЙ ТОРГОВО-ЭКОНОМИЧЕСКИЙ
УНИВЕРСИТЕТ ПОТРЕБИТЕЛЬСКОЙ КООПЕРАЦИИ»**

Кафедра информационно-вычислительных систем

**ЭКОНОМЕТРИКА
И ЭКОНОМИКО-МАТЕМАТИЧЕСКИЕ
МЕТОДЫ И МОДЕЛИ**

**Пособие
для студентов экономических специальностей**

Гомель 2012