

Федеральное агентство по образованию
Ульяновский государственный технический университет

Н. И. Шанченко

ЭКОНОМЕТРИКА

Лабораторный практикум

Методические указания для студентов
специальности 351400

Ульяновск 2004

УДК 330.43 (076)
ББК 65в6я73
Ш 20

Рецензент: д-р физ.-мат. наук, проф. УлГУ, А. С. Андреев

Одобрено секцией методических пособий научно-методического совета университета

Шанченко, Н. И.

Ш 20 Эконометрика: лабораторный практикум/ Н. И. Шанченко – Ульяновск: УлГТУ, 2004. – 79 с.

Содержит указания по выполнению лабораторных работ по дисциплине «Эконометрика», методические материалы, примеры решения типовых задач, варианты лабораторных работ.

Могут быть использованы преподавателями для организации лабораторных работ по дисциплине «Эконометрика», а также обучающимися для выполнения лабораторных работ.

Предназначены для студентов очной, вечерней, заочной и дистанционной форм обучения.

Работа подготовлена на кафедре «Информационные системы».

УДК 330.43 (076)
ББК 65в6я73

© Н. И. Шанченко, 2004
© Оформление. УлГТУ, 2004

1. Парная регрессия и корреляция

1.1. Понятие регрессии

Парной регрессией называется уравнение связи двух переменных y и x вида

$$y = f(x),$$

где y – зависимая переменная (результативный признак); x – независимая, объясняющая переменная (признак-фактор).

Различают линейные и нелинейные регрессии.

Линейная регрессия описывается уравнением: $y = a + b \cdot x + \varepsilon$.

Нелинейные регрессии делятся на два класса: регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам, и регрессии, нелинейные по оцениваемым параметрам.

Примеры регрессий, **нелинейных по объясняющим переменным**, но линейных по оцениваемым параметрам:

- полиномы разных степеней $\hat{y}_x = a + b_1 \cdot x + b_2 \cdot x^2 + b_3 \cdot x^3 + \varepsilon$;
- равнобочная гиперболола $\hat{y}_x = a + \frac{b}{x} + \varepsilon$.

Примеры регрессий, **нелинейных по оцениваемым параметрам**:

- степенная $\hat{y}_x = a \cdot x^b \cdot \varepsilon$;
- показательная $\hat{y}_x = a \cdot b^x \cdot \varepsilon$;
- экспоненциальная $\hat{y}_x = e^{a+b \cdot x} \cdot \varepsilon$.

Наиболее часто применяются следующие модели регрессий:

- прямой – $\hat{y}_x = a + b \cdot x$;
- гиперболола – $\hat{y}_x = a + b/x$;
- параболы – $\hat{y}_x = a + b \cdot x + c \cdot x^2$;
- показательной функции $\hat{y}_x = a \cdot b^x$;
- степенная функция – $\hat{y}_x = a \cdot x^b$ и др.

1.2. Построение уравнения регрессии

Постановка задачи. По имеющимся данным n наблюдений за совместным изменением двух параметров x и y $\{(x_i, y_i), i=1, 2, \dots, n\}$ необходимо определить аналитическую зависимость $\hat{y}=f(x)$, наилучшим образом описывающую данные наблюдений.

Построение уравнения регрессии осуществляется в два этапа (предполагает решение двух задач):

- спецификация модели (определение вида аналитической зависимости $\hat{y}=f(x)$);

– оценка параметров выбранной модели.

1.2.1. Спецификация модели

Парная регрессия применяется, если имеется доминирующий фактор, который и используется в качестве объясняющей переменной.

Применяется три основных метода выбора вида аналитической зависимости:

- графический (на основе анализа поля корреляций);
- аналитический, т. е. исходя из теории изучаемой взаимосвязи;
- экспериментальный, т. е. путем сравнения величины остаточной дисперсии $D_{\text{ост}}$ или средней ошибки аппроксимации \bar{A} , рассчитанных для различных моделей регрессии (метод перебора).

1.2.2. Оценка параметров модели

Для оценки параметров регрессий, линейных по этим параметрам, используется **метод наименьших квадратов** (МНК). МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака y от теоретических значений \hat{y}_x при тех же значениях фактора x минимальна, т. е.

$$\sum \left(y - \hat{y}_x \right)^2 \rightarrow \min.$$

В случае **линейной регрессии** параметры a и b находятся из следующей системы нормальных уравнений метода МНК:

$$\begin{cases} na + b \sum x = \sum y, \\ a \sum x + b \sum x^2 = \sum yx. \end{cases} \quad (1.1)$$

Можно воспользоваться готовыми формулами, которые вытекают из этой системы:

$$a = \bar{y} - b \cdot \bar{x}, \quad b = \frac{\text{cov}(x, y)}{\sigma_x^2} = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{x^2 - \bar{x}^2}. \quad (1.2)$$

Для нелинейных уравнений регрессии, приводимых к линейным с помощью преобразования $(x, y) \rightarrow (x', y')$, система нормальных уравнений имеет вид (1.1) в преобразованных переменных x', y' .

Коэффициент b при факторной переменной x имеет следующую интерпретацию: он показывает, на сколько изменится в среднем величина y при изменении фактора x на 1 единицу измерения.

Гиперболическая регрессия: $y_x = a_0 + a_1 / x$.

Линеаризующее преобразование: $x' = 1/x$; $y' = y$.

Уравнения (1.1) и формулы (1.2) принимают вид

$$\begin{cases} na + b \sum \frac{1}{x} = \sum y, \\ a \sum \frac{1}{x} + b \sum \frac{1}{x^2} = \sum y \frac{1}{x}. \end{cases}$$

$$a_0 = \frac{1}{n} \sum y - \frac{1}{n} a_1 \sum \frac{1}{x}, \quad a_1 = \frac{n \sum \frac{y}{x} - \sum \frac{1}{x} \sum y}{n \sum \frac{1}{x^2} - \left(\sum \frac{1}{x} \right)^2}.$$

Экспоненциальная регрессия: $y_x = e^{a_0 + a_1 x}$.

Линеаризующее преобразование: $x' = x$; $y' = \ln y$.

$$a_0 = \frac{1}{n} \sum \ln y - \frac{1}{n} a_1 \sum x, \quad a_1 = \frac{n \sum x \ln y - \sum x \sum \ln y}{n \sum x^2 - (\sum x)^2}.$$

Модифицированная экспонента: $y_x = K + a_0 \times a_1^x$, ($0 < a_1 < 1$).

Линеаризующее преобразование: $x' = x$; $y' = \ln |y - K|$.

$$\ln |a_0| = \frac{1}{n} \sum \ln |y - K| - \frac{1}{n} \ln a_1 \sum x, \quad a_1 = \frac{n \sum (x \ln |y - K|) - \sum x \sum \ln |y - K|}{n \sum x^2 - (\sum x)^2}.$$

Величина предела роста K выбирается предварительно на основе анализа поля корреляций либо из качественных соображений. Параметр a_0 берется со знаком «+», если $y_x > K$ и со знаком «-» в противном случае.

Степенная функция: $y_x = a_0 \times x^{a_1}$, ($a_0 > 0$).

Линеаризующее преобразование: $x' = \ln x$; $y' = \ln y$.

$$\ln a_0 = \frac{1}{n} \sum \ln y - \frac{1}{n} a_1 \sum \ln x, \quad a_1 = \frac{n \sum (\ln x \cdot \ln y) - \sum \ln x \sum \ln y}{n \sum (\ln x)^2 - (\sum \ln x)^2}.$$

Показательная функция: $y_x = a_0 \times a_1^x$.

Линеаризующее преобразование: $x' = x$; $y' = \ln y$.

$$\ln a_0 = \frac{1}{n} \sum \ln y - \frac{1}{n} \ln a_1 \sum x, \quad \ln a_1 = \frac{n \sum x \ln y - \sum x \sum \ln y}{n \sum x^2 - (\sum x)^2}.$$

Логарифмическая функция: $y_x = a_0 + a_1 \ln x$.

Линеаризующее преобразование: $x' = \ln x$; $y' = y$.

$$a_0 = \frac{1}{n} \sum y - \frac{1}{n} a_1 \sum \ln x, \quad a_1 = \frac{n \sum \ln x \cdot y - \sum \ln x \sum y}{n \sum (\ln x)^2 - (\sum \ln x)^2}.$$

Парабола второго порядка: $y_x = a_0 + a_1 x^1 + a_2 x^2$.

Парабола второго порядка имеет 3 параметра a_0 , a_1 , a_2 , которые определяются из системы трех уравнений

$$\begin{cases} n \cdot a_0 + a_1 \sum x + a_2 \sum x^2 = \sum y, \\ a_0 \sum x + a_1 \sum x^2 + a_2 \sum x^3 = \sum xy, \\ a_0 \sum x^2 + a_1 \sum x^3 + a_2 \sum x^4 = \sum x^2 y. \end{cases}$$

1.3. Оценка тесноты связи

Тесноту связи изучаемых явлений оценивает **линейный коэффициент парной корреляции** r_{xy} для линейной регрессии ($-1 \leq r_{xy} \leq 1$)

$$r_{xy} = \frac{\frac{1}{n} \sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\sigma_x \sigma_y} \quad (1.3)$$

и **индекс корреляции** ρ_{xy} для нелинейной регрессии ($0 \leq \rho_{xy} \leq 1$)

$$\rho_{xy} = \sqrt{1 - \frac{\sigma_{ост}^2}{\sigma_n^2}} = \sqrt{1 - \frac{\sum (y - \hat{y}_x)^2}{\sum (y - \bar{y})^2}} \quad (1.4)$$

Имеет место соотношение

$$r_{xy} = b \frac{\sigma_x}{\sigma_y}.$$

Долю дисперсии, объясняемую регрессией, в общей дисперсии результативного признака y характеризует **коэффициент детерминации** r^2_{xy} (для линейной регрессии) или **индекс детерминации** $R^2 = \rho^2_{xy}$ (для нелинейной регрессии).

Коэффициент детерминации – квадрат коэффициента или индекса корреляции.

Для оценки качества построенной модели регрессии можно использовать показатель (коэффициент, индекс) детерминации R^2 либо среднюю ошибку аппроксимации.

Чем выше показатель детерминации или чем ниже средняя ошибка аппроксимации, тем лучше модель описывает исходные данные.

Средняя ошибка аппроксимации – среднее относительное отклонение расчетных значений от фактических

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y - \hat{y}}{y} \right| \cdot 100\%.$$

Построенное уравнение регрессии считается удовлетворительным, если значение \bar{A} не превышает 10–12 %.

1.4. Оценка значимости уравнения регрессии, его коэффициентов, коэффициента детерминации

Оценка значимости всего уравнения регрессии в целом осуществляется с помощью F -критерия Фишера.

F -критерий Фишера заключается в проверке гипотезы H_0 о **статистической незначимости уравнения регрессии**. Для этого выполняется сравнение фактического $F_{\text{факт}}$ и критического (табличного) $F_{\text{табл}}$ значений F -критерия Фишера.

$F_{\text{факт}}$ определяется из соотношения значений факторной и остаточной дисперсий, рассчитанных на одну степень свободы

$$F_{\text{факт}} = \frac{\sum \frac{(\hat{y} - \bar{y})^2}{m}}{\sum \frac{(y - \hat{y})^2}{n - m - 1}} = \frac{r_{xy}^2}{1 - r_{xy}^2} \cdot \frac{n - m - 1}{m}, \quad (1.5)$$

где n – число единиц совокупности; m – число параметров при переменных. Для линейной регрессии $m = 1$.

Для нелинейной регрессии вместо r_{xy}^2 используется R^2 .

$F_{\text{табл}}$ – максимально возможное значение критерия под влиянием случайных факторов при степенях свободы $k_1 = m$, $k_2 = n - m - 1$ (для линейной регрессии $m = 1$) и уровне значимости α .

Уровень значимости α – вероятность отвергнуть правильную гипотезу при условии, что она верна. Обычно величина α принимается равной 0,05 или 0,01.

Если $F_{\text{табл}} < F_{\text{факт}}$, то H_0 -гипотеза о случайной природе оцениваемых характеристик отклоняется и признается их статистическая значимость и надежность. Если $F_{\text{табл}} > F_{\text{факт}}$, то гипотеза H_0 не отклоняется и признается статистическая незначимость, ненадежность уравнения регрессии.

Для **оценки статистической значимости коэффициентов линейной регрессии и линейного коэффициента парной корреляции r_{xy}** применяется **t -критерий Стьюдента** и рассчитываются доверительные интервалы каждого из показателей.

Согласно t -критерию выдвигается гипотеза H_0 о случайной природе показателей, т. е. о незначимом их отличии от нуля. Далее рассчитываются фактические значения критерия $t_{\text{факт}}$ для оцениваемых коэффициентов регрессии и коэффициента корреляции r_{xy} путем сопоставления их значений с величиной стандартной ошибки

$$t_b = \frac{b}{m_b}; \quad t_a = \frac{a}{m_a}; \quad t_r = \frac{r_{xy}}{m_{r_{xy}}}.$$

Стандартные ошибки параметров линейной регрессии и коэффициента корреляции определяются по формулам

$$m_b = \sqrt{\frac{\sum (y - \hat{y}_x)^2 / (n-2)}{\sum (x - \bar{x})^2}} = \sqrt{\frac{S_{ocm}^2}{\sum (x - \bar{x})^2}} = \frac{S_{ocm}}{\sigma_x \sqrt{n}} ;$$

$$m_a = \sqrt{\frac{\sum (y - \hat{y}_x)^2}{(n-2)} \cdot \frac{\sum x^2}{n \sum (x - \bar{x})^2}} = \sqrt{S_{ocm}^2 \frac{\sum x^2}{n^2 \sigma_x^2}} = S_{ocm} \frac{\sqrt{\sum x^2}}{n \sigma_x} ;$$

$$m_{r_{xy}} = \sqrt{\frac{1 - r_{xy}^2}{n-2}} .$$

Сравнивая фактическое и критическое (табличное) значения t -статистики $t_{табл}$ и $t_{факт}$ принимают или отвергают гипотезу H_0 .

$t_{табл}$ – максимально возможное значение критерия под влиянием случайных факторов при данной степени свободы $k = n-2$ и уровне значимости α .

Связь между F -критерием Фишера (при $k_1 = 1$; $m = 1$) и t -критерием Стьюдента выражается равенством

$$t_r^2 = t_b^2 = t_a^2 = \sqrt{F} .$$

Если $t_{табл} < t_{факт}$, то H_0 отклоняется, т. е. a , b и r_{xy} не случайно отличаются от нуля и сформировались под влиянием систематически действующего фактора x . Если $t_{табл} > t_{факт}$, то гипотеза H_0 не отклоняется и признается случайная природа формирования a , b или r_{xy} .

Значимость коэффициента детерминации R^2 (индекса корреляции) определяется с помощью F -критерия Фишера. Фактическое значение критерия $F_{факт}$ определяется по формуле

$$F_{факт} = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m} . \quad (1.6)$$

$F_{табл}$ определяется из таблицы при степенях свободы $k_1 = 1$, $k_2 = n-2$ и при заданном уровне значимости α . Если $F_{табл} < F_{факт}$, то признается статистическая значимость коэффициента детерминации. В формуле (1.6) величина m означает число параметров при переменных в соответствующем уравнении регрессии.

1.5. Расчет доверительных интервалов

Рассчитанные значения показателей (коэффициенты a , b , r_{xy}) являются приближенными, полученными на основе имеющихся выборочных данных. Для оценки того, насколько точные значения показателей могут отличаться от рассчитанных, осуществляется построение доверительных интервалов.

Доверительные интервалы определяют пределы, в которых лежат точные значения определяемых показателей с заданной степенью уверенности, соответствующей заданному уровню значимости α .

Для расчета доверительных интервалов для параметров a и b уравнения линейной регрессии определяем **предельную ошибку** Δ для каждого показателя:

$$\Delta_a = t_{\text{табл}} \cdot m_a, \quad \Delta b = t_{\text{табл}} \cdot m_b.$$

Величина $t_{\text{табл}}$ представляет собой табличное значение t -критерия Стьюдента под влиянием случайных факторов при степени свободы $k = n-2$ и заданном уровне значимости α .

Формулы для расчета *доверительных интервалов* имеют следующий вид:

$$\begin{aligned} \gamma_a &= a \pm \Delta_a; & \gamma_{a_{\min}} &= a - \Delta_a; & \gamma_{a_{\max}} &= a + \Delta_a; \\ \gamma_b &= b \pm \Delta_b; & \gamma_{b_{\min}} &= b - \Delta_b; & \gamma_{b_{\max}} &= b + \Delta_b. \end{aligned}$$

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.

Для статистически значимого линейного коэффициента корреляции можно построить интервальные оценки с помощью Z -распределения Фишера:

$$Z = \frac{1}{2} \cdot \ln \frac{1 + r_{xy}}{1 - r_{xy}}. \quad (1.7)$$

Первоначально определяется интервальная оценка для z по выражению

$$z \in \left[z' \pm t_\gamma \cdot \sqrt{\frac{1}{n-3}} \right],$$

где t_γ – значение случайной величины, подчиняющейся стандартному нормальному распределению, соответствующее вероятности $\gamma = 1 - \alpha/2$ (α – уровень значимости);

$z' = Z(r_{xy})$ – значение Z -распределения Фишера, соответствующее полученному значению линейного коэффициента корреляции r_{xy} .

Граничные значения доверительного интервала (r^- , r^+) для r_{xy} получаются из граничных значений доверительного интервала (z^- , z^+) для z с помощью функции, обратной Z -распределению Фишера

$$r_{xy}^\pm = Z^{-1}(z^\pm) = Z^{-1} \left(z' \pm t_\gamma \cdot \sqrt{\frac{1}{n-3}} \right).$$

1.6. Точечный и интервальный прогноз по уравнению линейной регрессии

Точечный прогноз заключается в получении прогнозного значения y_p , которое определяется путем подстановки в уравнение регрессии $\hat{y}_x = a + b \cdot x$ соответствующего (прогнозного) значения x_p

$$y_p = a + b \cdot x_p.$$

Интервальный прогноз заключается в построении доверительного интервала прогноза, т. е. нижней и верхней границ $y_{p\min}$, $y_{p\max}$ интервала, содержащего

го точную величину для прогнозного значения \hat{y}_p ($y_{p \min} < \hat{y}_p < y_{p \max}$). Доверительный интервал всегда определяется с заданной вероятностью (степенью уверенности), соответствующей принятому значению уровня значимости α .

Предварительно вычисляется **стандартная ошибка прогноза** $m_{y_p}^{\wedge}$

$$m_{y_p}^{\wedge} = \sigma_{ocm} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x - \bar{x})^2}},$$

$$\text{где } \sigma_{ocm} = \sqrt{\frac{\sum (y - \hat{y})^2}{n - m - 1}},$$

и затем строится **доверительный интервал прогноза**, т. е. определяются нижняя $\gamma_{y_{p \min}}^{\wedge}$ и верхняя $\gamma_{y_{p \max}}^{\wedge}$ границы интервала прогноза

$$\gamma_{y_{p \min}}^{\wedge} = \hat{y}_p - \Delta_{y_p}^{\wedge}; \quad \gamma_{y_{p \max}}^{\wedge} = \hat{y}_p + \Delta_{y_p}^{\wedge},$$

$$\text{где } \Delta_{y_p}^{\wedge} = t_{табл} \cdot m_{y_p}^{\wedge}.$$

1.7. Коэффициент эластичности

В экономических исследованиях широкое применение находит такой показатель, как **коэффициент эластичности**, вычисляемый по формуле

$$\mathcal{E} = f'(x) \frac{x}{y}. \quad (1.8)$$

Коэффициент эластичности показывает на сколько процентов изменится результат y при изменении фактора x на 1 % от своего номинального значения. Для линейной регрессии коэффициент эластичности равен

$$\mathcal{E} = b \frac{x}{y}$$

и зависит от x , поэтому рассчитывают средний коэффициент эластичности

$$\bar{\mathcal{E}} = f'(\bar{x}) \frac{\bar{x}}{\bar{y}} = b \frac{\bar{x}}{\bar{y}}.$$

Средний коэффициент эластичности $\bar{\mathcal{E}}$ показывает, на сколько процентов в среднем по совокупности изменится результат y от своей величины при изменении фактора x на 1 % от своего значения.

Контрольные вопросы:

1. Что понимается под парной регрессией?
2. Какие задачи решаются при построении уравнения регрессии?

3. Какие методы применяются для выбора вида модели регрессии?
4. Какие функции чаще всего используются для построения уравнения парной регрессии?
5. Какой вид имеет система нормальных уравнений метода наименьших квадратов в случае линейной регрессии?
6. Какой вид имеет система нормальных уравнений метода наименьших квадратов в случае гиперболической, показательной регрессии?
7. По какой формуле вычисляется линейный коэффициент парной корреляции r_{xy} ?
8. Как строится доверительный интервал для линейного коэффициента парной корреляции?
9. Как вычисляется индекс корреляции?
10. Как вычисляется и что показывает индекс детерминации?
11. Как проверяется значимость уравнения регрессии и отдельных коэффициентов?
12. Как строится доверительный интервал прогноза в случае линейной регрессии?
13. Как вычисляются и что показывают коэффициент эластичности $\bar{\varepsilon}$ и средний коэффициент эластичности $\bar{\varepsilon}$?

Лабораторная работа № 1

Задание. На основании данных табл. П1 для соответствующего варианта (табл. 1.1):

1. Вычислить линейный коэффициент парной корреляции.
2. Проверить значимость коэффициента парной корреляции.
3. Построить доверительный интервал для линейного коэффициента парной корреляции.

Лабораторная работа № 2

Задание. На основании данных табл. П1 для соответствующего варианта (табл. 1.1):

1. Построить предложенные уравнения регрессии, включая линейную регрессию.
2. Вычислить индексы парной корреляции для каждого уравнения.
3. Проверить значимость уравнений регрессии и отдельных коэффициентов линейного уравнения.
4. Определить лучшее уравнение регрессии на основе средней ошибки аппроксимации.
5. Построить интервальный прогноз для значения $x = x_{\max}$ для линейного уравнения регрессии.
6. Определить средний коэффициент эластичности.

Требования к оформлению результатов

Отчет о лабораторной работе должен содержать разделы:

1. Описание задания;
2. Описание решения лабораторной работы (по этапам);
3. Изложение полученных результатов.

Таблица 1. 1

Варианты кривых выравнивания к лабораторным работам № 1, 2

Вариант	Графы из табл. П1	Виды кривых выравнивания					
		Линейная	Параболическая	Экспоненциальная	Показательная	Логарифмическая	Гиперболическая
1	1, 2	*	*		*	*	
2	2, 3	*		*	*		
3	3, 4	*	*	*		*	
4	4, 5	*		*		*	
5	5, 6	*		*			*
6	6, 7	*			*	*	
7	7, 8	*	*			*	*
8	8, 9	*	*				*
9	9, 10	*			*	*	
10	1, 3	*		*		*	*
11	1, 4	*			*		*
12	1, 5	*			*	*	
13	1, 6	*	*		*	*	
14	1, 7	*			*		*
15	1, 8	*				*	*
16	1, 9	*	*	*			*
17	2, 4	*		*			*
18	2, 5	*			*	*	
19	2, 6	*				*	*
20	2, 7	*			*		*
21	2, 8	*	*		*	*	
22	2, 9	*	*			*	*
23	3, 6	*		*			*
24	3, 7	*			*	*	
25	3, 8	*	*		*		

2. Множественная регрессия и корреляция

2.1. Общие положения

Множественная регрессия – уравнение связи с несколькими независимыми переменными:

$$y = f(x_1, x_2, \dots, x_p),$$

где y – зависимая переменная (результативный признак);

x_1, x_2, \dots, x_p – независимые переменные (факторы).

Множественная регрессия применяется в ситуациях, когда из множества факторов, влияющих на результативный признак, нельзя выделить один доминирующий фактор и необходимо учитывать влияние нескольких факторов.

Основная цель множественной регрессии – построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также совокупное их воздействие на моделируемый показатель.

Постановка задачи множественной регрессии. По имеющимся данным n наблюдений за совместным изменением $n+1$ параметра y и x_j и $((y_i, x_{j,i}); j=1, 2, \dots, p; i=1, 2, \dots, n)$ необходимо определить аналитическую зависимость $\hat{y}=f(x_1, x_2, \dots, x_p)$, наилучшим образом описывающую данные наблюдений.

Как и в случае парной регрессии, построение уравнения множественной регрессии осуществляется в два этапа:

- спецификация модели;
- оценка параметров выбранной модели.

Спецификация модели включает в себя решение двух задач:

- отбор p факторов x_j , наиболее влияющих на величину y ;
- выбор вида уравнения регрессии $\hat{y}=f(x_1, x_2, \dots, x_p)$;

2.2. Отбор факторов при построении множественной регрессии

Включение в уравнение множественной регрессии того или иного набора факторов связано, прежде всего, с представлением исследователя о природе взаимосвязи моделируемого показателя с другими экономическими явлениями. Факторы, включаемые во множественную регрессию, должны отвечать следующим требованиям:

1. *Они должны быть количественно измеримы.* Если необходимо включить в модель качественный фактор, не имеющий количественного измерения, то ему нужно придать количественную определенность (например, в модели урожайности качество почвы задается в виде баллов; в модели стоимости объектов недвижимости районам присваиваются ранги);

2. *Факторы не должны быть взаимно коррелированы* и тем более находиться в точной функциональной связи. Если между факторами существует высокая корреляция, то нельзя определить их изолированное влияние на результа-

тивный показатель, и параметры уравнения регрессии оказываются неинтерпретируемыми.

Включаемые во множественную регрессию факторы должны объяснить вариацию независимой переменной. Если строится модель с набором p факторов, то для нее рассчитывается показатель детерминации R^2 , который фиксирует долю объясненной вариации результативного признака за счет рассматриваемых в регрессии p факторов. Влияние других, не учтенных в модели, факторов оценивается как $1 - R^2$ с соответствующей остаточной дисперсией S^2 .

При дополнительном включении в регрессию $(p + 1)$ -фактора x_{p+1} коэффициент детерминации должен возрасти, а остаточная дисперсия уменьшаться, т. е.

$$R_{p+1}^2 \geq R_p^2 \quad \text{и} \quad S_{p+1}^2 \leq S_p^2.$$

Если же этого не происходит и данные показатели практически мало отличаются друг от друга, то включаемый в анализ фактор x_{p+1} не улучшает модель и практически является лишним фактором.

Насыщение модели лишними факторами не только не снижает величину остаточной дисперсии и не увеличивает показатель детерминации, но и приводит к статистической незначимости параметров регрессии по t -критерию Стьюдента.

Отбор факторов производится на основе качественного теоретико-экономического анализа и обычно осуществляется в две стадии:

- на первой подбираются факторы исходя из сущности проблемы;
- на второй – на основе матрицы показателей корреляции определяют t -статистики для параметров регрессии.

Коэффициенты интеркорреляции (т. е. корреляции между объясняющими переменными) позволяют исключать из модели дублирующие факторы. Считается, что две переменные явно коллинеарные, т. е. находятся между собой в линейной зависимости, если $r_{x_i x_j} \geq 0,7$.

Если факторы явно коллинеарны, то они дублируют друг друга и один из них рекомендуется исключить из регрессии. Предпочтение при этом отдается тому фактору, который при достаточно тесной связи с результатом имеет наименьшую тесноту связи с другими факторами. В этом требовании проявляется специфика множественной регрессии как метода исследования комплексного воздействия факторов в условиях их независимости друг от друга.

Пусть, например, при изучении зависимости $y = f(x, z, v)$ матрица парных коэффициентов корреляции оказалась следующей:

	y	x	z	v
y	1			
x	0,85	1		
z	0,75	0,8	1	
v	0,5	0,4	0,3	1

Очевидно, что факторы x и z дублируют друг друга. В анализ целесообразно включить фактор z , а не x , хотя корреляция z с результатом y слабее, чем корреляция фактора x ($r_{yz} < r_{yx}$), но зато слабее межфакторная корреляция между z и v ($r_{zv} < r_{xv}$). Поэтому в данном случае в уравнение множественной регрессии включаются факторы z , v .

Наибольшие трудности в использовании аппарата множественной регрессии возникают при наличии мультиколлинеарности факторов, когда более чем два фактора связаны между собой линейной зависимостью, т. е. имеет место совокупное воздействие факторов друг на друга.

Для оценки мультиколлинеарности факторов может использоваться определитель матрицы парных коэффициентов корреляции $r_{x_i x_j}$ между факторами.

В случае трех факторов определитель имеет вид

$$Det|R| = \begin{vmatrix} r_{x_1 x_1} & r_{x_2 x_1} & r_{x_3 x_1} \\ r_{x_1 x_2} & r_{x_2 x_2} & r_{x_3 x_2} \\ r_{x_1 x_3} & r_{x_2 x_3} & r_{x_3 x_3} \end{vmatrix}.$$

Чем ближе к нулю определитель матрицы межфакторной корреляции, тем сильнее мультиколлинеарность факторов и ненадежнее результаты множественной регрессии. И, наоборот, чем ближе к единице определитель матрицы межфакторной корреляции, тем меньше мультиколлинеарность факторов.

Оценка значимости мультиколлинеарности факторов может быть проведена методом испытания гипотезы о независимости переменных $H_0 : Det|R| = 1$.

Доказано, что величина $\left[n - 1 - \frac{1}{6}(2m + 5) \lg DetR \right]$ имеет приближенное распределение χ^2 с $\frac{1}{2}n(n-1)$ степени свободы. Если фактическое значение χ^2 превосходит табличное (критическое) $\chi_{факт}^2 > \chi_{табл(df, \alpha)}^2$, то гипотеза H_0 отклоняется. Это означает, что $Det|R| \neq 1$, недиагональные ненулевые коэффициенты корреляции указывают на коллинеарность факторов. Мультиколлинеарность считается доказанной.

Через коэффициенты множественной детерминации можно найти переменные, ответственные за мультиколлинеарность факторов. Для этого в качестве зависимой переменной рассматривается каждый из факторов. Чем ближе значение коэффициента множественной детерминации к единице, тем сильнее проявляется мультиколлинеарность факторов. Сравнивая между собой коэффициенты множественной детерминации факторов

$$(R_{x_1|x_2 x_3 \dots x_p}^2; R_{x_2|x_1 x_3 \dots x_p}^2 \text{ и т. п.}),$$

можно выделить переменные, ответственные за мультиколлинеарность, следовательно, можно решать проблему отбора факторов, оставляя в уравнении факторы с минимальной величиной коэффициента множественной детерминации.

Существует ряд подходов **преодоления сильной межфакторной корреляции**:

- *исключение из модели одного или нескольких факторов*;
- *преобразование факторов*, при котором уменьшается корреляция между ними. Например, переходят от исходных переменных к их линейным комбинациям, не коррелированным друг с другом (метод главных компонент). При построении модели на основе рядов динамики переходят от первоначальных данных к первым разностям уровней $\Delta y_t = y_t - y_{t-1}$, чтобы исключить влияние тенденции;

- *переход к совмещенным уравнениям регрессии*, т. е. к уравнениям, которые отражают не только влияние факторов, но и их взаимодействие. Так, если $y = f(x_1, x_2, x_3)$, то возможно построение следующего совмещенного уравнения:

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 + b_{12} \cdot x_1 \cdot x_2 + b_{13} \cdot x_1 \cdot x_3 + b_{23} \cdot x_2 \cdot x_3 + \varepsilon.$$

Рассматриваемое уравнение включает взаимодействие первого порядка (взаимодействие двух факторов). Часть этих взаимодействий могут оказаться несущественными, поэтому нецелесообразно полное включение в модель взаимодействий всех факторов. Так, если анализ совмещенного уравнения показал значимость только взаимодействия факторов x_1 и x_3 , то уравнение будет иметь вид

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 + b_{13} \cdot x_1 \cdot x_3 + \varepsilon.$$

После исключения коллинеарных факторов осуществляется процедура отбора факторов, наиболее влияющих на изменение результативного признака (факторов, включаемых в регрессию). Подходы к отбору факторов на основе показателей корреляции могут быть разные.

Наиболее широкое применение получили следующие методы построения уравнения множественной регрессии:

- метод исключения;
- метод включения;
- шаговый регрессионный анализ.

Каждый из этих методов по-своему решает проблему отбора факторов, давая в целом близкие результаты – отсев факторов из полного его набора (метод исключения), дополнительное введение фактора (метод включения), исключение ранее введенного фактора (шаговый регрессионный анализ).

В процедуре отсева факторов наиболее широко используется **матрица частных коэффициентов корреляции** (см. п. 2.7).

При отборе факторов рекомендуется, кроме всего прочего, пользоваться следующим правилом: число включаемых факторов должно быть в 6–7 раз меньше объема совокупности, по которой строится регрессия.

2.3. Выбор формы уравнения регрессии

Как и в парной зависимости, возможны разные виды уравнений множественной регрессии: линейные и нелинейные.

Ввиду четкой интерпретации параметров наиболее широко используются линейная и степенная функции.

В уравнении **линейной множественной регрессии**

$$\hat{y}_x = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p \quad (2.1)$$

параметры при x_i называются коэффициентами «чистой» регрессии. Они характеризуют среднее изменение результата с изменением соответствующего фактора на единицу при неизменном значении других факторов, закрепленных на среднем уровне.

Предположим, например, что зависимость расходов на продукты питания по совокупности семей характеризуется следующим уравнением:

$$\hat{y}_x = 0,5 + 0,35 \cdot x_1 + 0,73 \cdot x_2,$$

где y – расходы семьи за месяц на продукты питания, тыс. руб.;

x_1 – месячный доход на одного члена семьи, тыс. руб.;

x_2 – размер семьи, человек.

Анализ данного уравнения позволяет сделать выводы – с ростом дохода на одного члена семьи на 1 тыс. руб. расходы на питание возрастут в среднем на 350 руб. при том же среднем размере семьи. Иными словами, 35 % дополнительных семейных расходов тратится на питание. Увеличение размера семьи при тех же ее доходах предполагает дополнительный рост расходов на питание на 730 руб. Параметр a не подлежит экономической интерпретации.

В уравнении степенной функции

$$\hat{y}_x = a \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot \dots \cdot x_p^{b_p} \quad (2.2)$$

коэффициенты b_j являются коэффициентами эластичности. Они показывают, на сколько процентов изменяется в среднем результат с изменением соответствующего фактора на 1 % при неизменности действия других факторов. Этот вид уравнения регрессии получил наибольшее распространение в производственных функциях, в исследованиях спроса и потребления.

Предположим, что при исследовании спроса на мясо получено уравнение

$$\hat{y}_x = 0,82 \cdot x_1^{-2,63} x_2^{1,11} \quad \text{или} \quad \hat{y}_x = 0,82 \cdot \frac{x_2^{1,11}}{x_1^{2,63}},$$

где y – количество спрашиваемого мяса; x_1 – цена; x_2 – доход.

Следовательно, рост цен на 1 % при том же доходе вызывает снижение спроса в среднем на 2,63 %. Увеличение дохода на 1 % обуславливает при неизменных ценах рост спроса на 1,11 %.

В производственных функциях вида

$$P = a \cdot F_1^{b_1} \cdot F_2^{b_2} \cdot \dots \cdot F_m^{b_m} \cdot \varepsilon,$$

где P – количество продукта, изготавливаемого с помощью m производственных факторов (F_1, F_2, \dots, F_m), параметры b_i характеризуют эластичность количества продукции по отношению к количеству соответствующего производственного фактора.

Экономический смысл имеют не только коэффициенты b_i каждого фактора, но и их сумма, т. е. сумма эластичностей: $B = b_1 + b_2 + \dots + b_m$. Эта величина фиксирует обобщенную характеристику эластичности производства.

Для построения уравнения множественной регрессии чаще всего используются следующие функции:

- линейная – $y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p + \varepsilon$;
- степенная – $y = a \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot \dots \cdot x_p^{b_p} \cdot \varepsilon$;
- экспонента – $y = e^{a+b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p + \varepsilon}$;
- гипербола – $y = \frac{1}{a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p + \varepsilon}$.

Если исследователя не устраивает предлагаемый набор функций регрессии, то можно использовать любые другие функции, приводимые путем соответствующих преобразований к линейному виду, например:

$$\hat{y}_x = a + b_1 \cdot x_1 + b_2 \cdot \frac{1}{x_2} + b_3 \cdot x_3^{\frac{1}{2}} + b_4 \cdot \ln x_4.$$

Обозначив

$$z_1 = x_1, \quad z_2 = \frac{1}{x_2}, \quad z_3 = x_3^{\frac{1}{2}}, \quad z_4 = \ln x_4,$$

получим линейное уравнение множественной регрессии

$$y = a + b_1 \cdot z_1 + b_2 \cdot z_2 + b_3 \cdot z_3 + b_4 \cdot z_4 + \varepsilon.$$

Однако чем сложнее функция, тем менее интерпретируемы ее параметры.

Если один и тот же фактор вводится в регрессию в разных степенях, то каждая степень рассматривается как самостоятельный фактор. Так, если модель имеет вид полинома второго порядка

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + b_{11} \cdot x_1^2 + b_{22} \cdot x_2^2 + b_{12} \cdot x_1 \cdot x_2 + \varepsilon,$$

то после замены переменных $z_1 = x_1, z_2 = x_2, z_3 = x_1^2, z_4 = x_2^2, z_5 = x_1 x_2$ получим линейное уравнение регрессии с пятью факторами:

$$y = a + b_1 \cdot z_1 + b_2 \cdot z_2 + b_3 \cdot z_3 + b_4 \cdot z_4 + b_5 \cdot z_5 + \varepsilon.$$

Поскольку, как отмечалось, должно выполняться соотношение между числом параметров и числом наблюдений, для полинома второй степени требуется не менее 30-35 наблюдений.

2.4. Оценка параметров уравнения множественной регрессии

Для оценки параметров уравнения множественной регрессии применяют **метод наименьших квадратов** (МНК). Для линейных уравнений регрессии (и нелинейных уравнений, приводимых к линейным) строится система нормальных уравнений, решение которой позволяет получить оценки параметров регрессии. В случае линейной множественной регрессии

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p$$

система нормальных уравнений имеет следующий вид:

$$\begin{aligned} \sum y &= n \cdot a + b_1 \sum x_1 + b_2 \sum x_2 + \dots + b_p \sum x_p; \\ \sum yx_1 &= a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_2x_1 + \dots + b_p \sum x_px_1; \\ &..... \\ \sum yx_p &= a \sum x_p + b_1 \sum x_1x_p + b_2 \sum x_2x_p + \dots + b_p \sum x_p^2. \end{aligned}$$

Для определения значимости факторов и повышения точности результата используется **уравнение множественной регрессии в стандартизованном масштабе**

$$t_y = \beta_1 \cdot t_{x_1} + \beta_2 \cdot t_{x_2} + \dots + \beta_p \cdot t_{x_p} + \varepsilon, \quad (2.3)$$

где $t_y, t_{x_1}, \dots, t_{x_p}$ – **стандартизованные переменные**

$$t_y = \frac{y - \bar{y}}{\sigma_y}, \quad t_{x_i} = \frac{x_i - \bar{x}_i}{\sigma_{x_i}}, \quad (2.4)$$

для которых среднее значение равно нулю $\bar{t}_y = \bar{t}_{x_i} = 0$, а среднее квадратическое отклонение равно единице $\sigma_{t_y} = \sigma_{t_{x_i}} = 1$.

Величины β_i называются **стандартизованными коэффициентами регрессии**.

К уравнению множественной регрессии в стандартизованном масштабе применим МНК. Стандартизованные коэффициенты регрессии (β -коэффициенты) определяются из следующей системы уравнений:

$$\begin{aligned} \sum t_y t_{x_1} &= \beta_1 \sum t_{x_1}^2 + \beta_2 \sum t_{x_1} t_{x_2} + \beta_3 \sum t_{x_1} t_{x_3} + \dots + \beta_p \sum t_{x_1} t_{x_p}; \\ \sum t_y t_{x_2} &= \beta_1 \sum t_{x_1} t_{x_2} + \beta_2 \sum t_{x_2}^2 + \beta_3 \sum t_{x_2} t_{x_3} + \dots + \beta_p \sum t_{x_2} t_{x_p}; \\ \sum t_y t_{x_p} &= \beta_1 \sum t_{x_1} t_{x_p} + \beta_2 \sum t_{x_2} t_{x_p} + \beta_3 \sum t_{x_3} t_{x_p} + \dots + \beta_p \sum t_{x_p}^2 \end{aligned}$$

либо из системы уравнений

$$\begin{aligned} r_{yx_1} &= \beta_1 + \beta_2 r_{x_2x_1} + \beta_3 r_{x_3x_1} + \dots + \beta_p r_{x_px_1}; \\ r_{yx_2} &= \beta_1 r_{x_1x_2} + \beta_2 + \beta_3 r_{x_3x_2} + \dots + \beta_p r_{x_px_2}; \\ &..... \\ r_{yx_p} &= \beta_1 r_{x_1x_p} + \beta_2 r_{x_2x_p} + \beta_3 r_{x_3x_p} + \dots + \beta_p. \end{aligned}$$

Стандартизованные коэффициенты регрессии показывают, на сколько сигм (средних квадратических отклонений) изменится в среднем результат, если соответствующий фактор x_i изменится на одну сигму при неизменном среднем уровне других факторов. В силу того, что все переменные заданы как центрированные и нормированные, стандартизованные коэффициенты регрессии β_i сравнимы между собой. Сравнивая их друг с другом, **можно ранжировать факторы по силе их воздействия на результат**. В этом основное достоинство стандартизованных коэффициентов регрессии в отличие от коэффициентов «чистой» регрессии, которые несравнимы между собой.

В парной зависимости стандартизованный коэффициент регрессии β есть не что иное, как линейный коэффициент корреляции r_{yx} .

Связь коэффициентов множественной регрессии b_i со стандартизованными коэффициентами β_i описывается соотношением

$$b_i = \beta_i \frac{\delta_y}{\delta_{x_i}}.$$

Параметр a определяется из соотношения $a = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2 - \dots - b_p \bar{x}_p$.

Средние коэффициенты эластичности для линейной множественной регрессии рассчитываются по формуле

$$\bar{\varepsilon}_{yx_j} = b_j \frac{\bar{x}_j}{\bar{y}} \quad (2.5)$$

и показывают, на сколько процентов в среднем по совокупности изменится результат y от своей величины при изменении фактора x на 1 % от своего значения при неизменных значениях других факторов.

2.5. Частные уравнения регрессии

На основе линейного уравнения множественной регрессии

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p + \varepsilon$$

могут быть найдены частные уравнения регрессии, т. е. уравнения регрессии, которые связывают результативный признак с соответствующими факторами x_i при закреплении других, учитываемых во множественной регрессии, факторов на среднем уровне. Частные уравнения регрессии имеют следующий вид:

$$\begin{aligned} y_{x_1, x_2, x_3, \dots, x_p} &= a + b_1 \cdot x_1 + b_2 \cdot \bar{x}_2 + b_3 \cdot \bar{x}_3 + \dots + b_p \cdot \bar{x}_p + \varepsilon; \\ y_{x_2, x_1, x_3, \dots, x_p} &= a + b_1 \cdot \bar{x}_1 + b_2 \cdot x_2 + b_3 \cdot \bar{x}_3 + \dots + b_p \cdot \bar{x}_p + \varepsilon; \\ &\dots \\ y_{x_p, x_1, x_2, \dots, x_{p-1}} &= a + b_1 \cdot \bar{x}_1 + b_2 \cdot \bar{x}_2 + \dots + b_{p-1} \cdot \bar{x}_{p-1} + b_p \cdot x_p + \varepsilon. \end{aligned} \quad (2.6)$$

При подстановке в эти уравнения средних значений соответствующих факторов они принимают вид парных уравнений линейной регрессии, т. е. имеем

$$\begin{aligned}\hat{y}_{x_1, x_2, x_3, \dots, x_p} &= A_1 + b_1 \cdot x_1 \\ \hat{y}_{x_2, x_1, x_3, \dots, x_p} &= A_2 + b_2 \cdot x_2 \\ &\dots \dots \dots \\ \hat{y}_{x_p, x_1, x_2, \dots, x_{p-1}} &= A_p + b_p \cdot x_p\end{aligned}\quad (2.7)$$

где

$$\begin{aligned}A_1 &= a + b_2 \cdot \bar{x}_2 + b_3 \cdot \bar{x}_3 + \dots + b_p \cdot \bar{x}_p; \\ A_2 &= a + b_1 \cdot \bar{x}_1 + b_3 \cdot \bar{x}_3 + \dots + b_p \cdot \bar{x}_p; \\ &\dots \dots \dots \\ A_p &= a + b_1 \bar{x}_1 + b_2 \cdot \bar{x}_2 + \dots + b_{p-1} \cdot \bar{x}_{p-1}.\end{aligned}\quad (2.8)$$

В отличие от парной регрессии, частные уравнения регрессии характеризуют изолированное влияние фактора на результат, ибо другие факторы закреплены на неизменном уровне. Эффекты влияния других факторов присоединены в них к свободному члену уравнения множественной регрессии. Это позволяет на основе частных уравнений регрессии определять **частные коэффициенты эластичности**

$$\mathcal{E}_{y x_i} = b_i \frac{x_i}{\hat{y}_{x_i, x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_p}}, \quad (2.9)$$

где b_i – коэффициенты регрессии для фактора x_i в уравнении множественной регрессии; $\hat{y}_{x_i, x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_p}$ – значение результативного фактора, полученное из частного уравнения регрессии при данном значении фактора x_i .

Предположим, что по ряду регионов множественная регрессия величины импорта на определенный товар y относительно отечественного его производства x_1 , изменения запасов x_2 и потребления на внутреннем рынке x_3 оказалась следующей

$$\hat{y} = -66,028 + 0,135 \cdot x_1 + 0,476 \cdot x_2 + 0,343 \cdot x_3.$$

При этом средние значения для рассматриваемых признаков составили:

$$\bar{y} = 31,5, \quad \bar{x}_1 = 245,7, \quad \bar{x}_2 = 3,7, \quad \bar{x}_3 = 182,5.$$

На основе данной информации могут быть найдены средние по совокупности показатели эластичности

$$\bar{\mathcal{E}}_{y x_i} = b_i \cdot \frac{\bar{x}_i}{\bar{y}_{x_i}}.$$

Для данного примера они окажутся равными:

$$\begin{aligned}\bar{\mathcal{E}}_{y x_1} &= 0,135 \cdot \frac{245,7}{31,5} = 1,053\%, & \bar{\mathcal{E}}_{y x_2} &= 0,476 \cdot \frac{3,7}{31,5} = 0,056\%, \\ \bar{\mathcal{E}}_{y x_3} &= 0,343 \cdot \frac{182,5}{31,5} = 1,987\%,\end{aligned}$$

т. е.: 1) с ростом величины отечественного производства на 1 % размер импорта в среднем по совокупности регионов возрастет на 1,053 % при неизменных запасах и потреблении семей; 2) с ростом изменения запасов на 1 % при неизменном производстве и внутреннем потреблении величина импорта увеличивается в среднем на 0,056 %; 3) при неизменном объеме производства и величины запасов с увеличением внутреннего потребления на 1 % импорт товара возрастает в среднем по совокупности регионов на 1,987 %.

Средние показатели эластичности можно сравнивать друг с другом и соответственно ранжировать факторы по силе их воздействия на результат. В рассматриваемом примере наибольшее воздействие на величину импорта оказывает размер внутреннего потребления товара x_3 , а наименьшее – изменение запасов x_2 .

2.6. Множественная корреляция

Практическая значимость уравнения множественной регрессии оценивается с помощью показателя множественной корреляции и его квадрата – коэффициента детерминации.

Показатель множественной корреляции характеризует тесноту связи рассматриваемого набора факторов с исследуемым признаком, или, иначе, оценивает тесноту совместного влияния факторов на результат.

Независимо от формы связи показатель множественной корреляции может быть найден как индекс множественной корреляции

$$R_{yx_1x_2\dots x_p} = \sqrt{1 - \frac{\sigma_{ост}^2}{\sigma_y^2}}, \quad (2.10)$$

где σ_y^2 – общая дисперсия результативного признака; $\sigma_{ост}^2$ – остаточная дисперсия для уравнения $\hat{y} = f(x_1, x_2, \dots, x_p)$

$$\sigma_{ост}^2 = \frac{\sum (y - \hat{y}_{x_1x_2\dots x_p})^2}{n}.$$

Индекс множественной корреляции изменяется от 0 до 1. Чем ближе его значение к 1, тем теснее связь результативного признака со всем набором исследуемых факторов. Величина индекса множественной корреляции больше или равна максимального парного индекса корреляции

$$R_{yx_1x_2\dots x_p} \geq R_{yx_i(\max)} \quad (i = \overline{1, p}).$$

При правильном включении факторов в регрессионный анализ величина индекса множественной корреляции будет существенно отличаться от индекса корреляции парной зависимости. Отсюда ясно, что, сравнивая индексы множественной и парной корреляции, можно сделать вывод о целесообразности включения в уравнение регрессии того или иного фактора. Так, если у рассматривается как функция от x и z и получен индекс множественной корреляции $R_{yzx} = 0,85$, а индексы парной корреляции при этом были $R_{yx} = 0,82$ и $R_{yz} = 0,75$,

то совершенно ясно, что уравнение парной регрессии $y = f(x)$ охватывало 67,2 % ($0,82^2 = 0,672$) колеблемости результативного признака под влиянием фактора x , а дополнительное включение в анализ фактора z увеличило долю объясненной вариации до 72,3 % ($0,85^2 = 0,723$) т. е. уменьшилась доля остаточной вариации на 5,1 процентного пункта (с 32,8 до 27,7%).

Можно пользоваться следующей формулой индекса множественной корреляции:

$$R_{yx_1x_2\dots x_p} = \sqrt{1 - \frac{\sum (y - \hat{y}_{yx_1x_2\dots x_p})^2}{\sum (y - \bar{y})^2}}. \quad (2.11)$$

При линейной зависимости признаков формула индекса корреляции может быть представлена следующим выражением:

$$R_{yx_1x_2\dots x_p} = \sqrt{\sum_i \beta_{x_i} * r_{yx_i}},$$

где β_{x_i} – стандартизованные коэффициенты регрессии;

r_{yx_i} – парные коэффициенты корреляции результата с каждым фактором.

Формула индекса множественной корреляции для линейной регрессии получила название **линейного коэффициента множественной корреляции**, или, что то же самое, **совокупного коэффициента корреляции**.

Индекс множественной корреляции равен совокупному коэффициенту корреляции не только при линейной зависимости рассматриваемых признаков. Тождественность этих показателей, как и в парной регрессии, имеет место и для криволинейной зависимости, нелинейной по переменным. Так, если для фирмы модель прибыли y имеет вид

$$y = a + b_1 \cdot x_1 + b_2 \cdot \ln x_2 + b_1 \cdot x_1 + b_3 \cdot \ln x_3 + b_4 \cdot \ln x_4 + \varepsilon,$$

где x_1 – удельные расходы на рекламу; x_2 – капитал фирмы; x_3 – доля продукции фирмы в общем объеме продаж данной группы товаров по региону; x_4 – процент увеличения объема продаж фирмы по сравнению с предыдущим годом.

Тогда независимо от того, что фактор x_1 задан линейно, а факторы x_2, x_3, x_4 – в логарифмах, оценка тесноты связи может быть произведена с помощью линейного коэффициента множественной корреляции.

В рассмотренных показателях множественной корреляции (индекс и коэффициент) используется остаточная дисперсия, которая имеет систематическую ошибку в сторону преуменьшения, тем более значительную, чем больше параметров определяется в уравнении регрессии при заданном объеме наблюдений n . Если число параметров при x_j равно m и приближается к объему наблюдений, то остаточная дисперсия будет близка к нулю и коэффициент (индекс) корреляции приблизится к единице даже при слабой связи факторов с результатом.

Для того чтобы не допустить возможного преувеличения тесноты связи, используется скорректированный индекс (коэффициент) множественной корреляции. Скорректированный индекс множественной корреляции содержит по-

правку на число степеней свободы, а именно, остаточная сумма квадратов $\sum (y - \hat{y}_{x_1 x_2 \dots x_p})^2$ делится на число степеней свободы остаточной вариации $(n - m - 1)$, а общая сумма квадратов отклонений $\sum (y - \bar{y})^2$ – на число степеней свободы в целом по совокупности $(n - 1)$.

Формула скорректированного индекса множественной детерминации имеет вид

$$\bar{R}^2 = 1 - \frac{\sum (y - \hat{y})^2 : (n - m - 1)}{\sum (y - \bar{y})^2 : (n - 1)}, \quad (2.12)$$

где m – число параметров при переменных x_i , n – число наблюдений.

Поскольку

$$\sum (y - \hat{y})^2 / \sum (y - \bar{y})^2 = 1 - R^2,$$

то величину скорректированного индекса детерминации можно представить в виде

$$\bar{R}^2 = 1 - (1 - R^2) * \frac{(n - 1)}{(n - m - 1)}. \quad (2.13)$$

Чем больше величина m , тем сильнее различия \bar{R}^2 и R^2 .

Для линейной зависимости признаков скорректированный коэффициент множественной корреляции определяется по той же формуле, что и индекс множественной корреляции, т. е. как корень квадратный из \bar{R}^2 . Отличие состоит лишь в том, что в линейной зависимости под m подразумевается число факторов, включенных в регрессионную модель, а в криволинейной зависимости m – число параметров при x и их преобразованиях (x^2 , $\ln x$ и др.), которое может быть больше числа факторов как экономических переменных. Так, если $y = f(x_1, x_2)$, то для линейной регрессии $m = 2$, а для регрессии вида

$$y = a + b_1 \cdot x_1 + b_{12} \cdot x_1^2 + b_2 \cdot x_2 + b_{22} \cdot x_2^2 + \varepsilon$$

число параметров при x равно 4, т. е. $m = 4$. При заданном объеме наблюдений, при прочих равных условиях, с увеличением числа независимых переменных (параметров) скорректированный коэффициент множественной детерминации убывает. Его величина может стать и отрицательной при слабых связях результата с факторами. В этом случае он должен считаться равным нулю. При небольшом числе наблюдений скорректированная величина коэффициента множественной детерминации R^2 имеет тенденцию переоценивать долю вариации результативного признака, связанную с влиянием факторов, включенных в регрессионную модель.

В статистических пакетах прикладных программ в процедуре множественной регрессии обычно приводится скорректированный коэффициент (индекс) множественной корреляции (детерминации). Величина коэффициента множественной детерминации используется для оценки качества регрессионной модели.

Низкое значение коэффициента (индекса) множественной корреляции означает, что в регрессионную модель не включены существенные факторы – с одной стороны, а с другой стороны – рассматриваемая форма связи не отражает реальные соотношения между переменными, включенными в модель.

В этом случае требуются дальнейшие исследования по улучшению качества модели и увеличению ее практической значимости.

2.7. Частная корреляция

Как было показано выше, ранжирование факторов, участвующих во множественной линейной регрессии, может быть проведено через стандартизованные коэффициенты регрессии (β -коэффициенты). Эта же цель может быть достигнута с помощью частных коэффициентов корреляции – для линейных связей. При нелинейной взаимосвязи исследуемых признаков эту функцию выполняют частные индексы детерминации. Кроме того, частные показатели корреляции широко используются при решении проблемы отбора факторов: целесообразность включения того или иного фактора в модель доказывается величиной показателя частной корреляции.

Частные коэффициенты (или индексы) **корреляции** характеризуют тесноту связи между результатом и соответствующим фактором при устранении влияния других факторов, включенных в уравнение регрессии.

Показатели частной корреляции представляют собой отношение сокращения остаточной дисперсии, произошедшего в результате дополнительного включения в анализ нового фактора, к остаточной дисперсии, имевшей место до введения в модель нового фактора.

В общем виде при наличии p факторов для уравнения

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p + \varepsilon$$

коэффициент частной корреляции, измеряющий влияние на y фактора x_i при неизменном уровне других факторов, можно определить по формуле

$$r_{yx_i * x_1 x_2 \dots x_{i-1} x_{i+1} \dots x_p} = \sqrt{1 - \frac{1 - R_{yx_1 x_2 \dots x_i \dots x_p}^2}{1 - R_{yx_1 x_2 \dots x_{i-1} x_{i+1} \dots x_p}^2}}, \quad (2.14)$$

где $R_{yx_1 x_2 \dots x_i \dots x_p}^2$ – множественный коэффициент детерминации всего комплекса p факторов с результатом;

$R_{yx_1 x_2 \dots x_{i-1} x_{i+1} \dots x_p}^2$ – показатель детерминации, но без введения в модель фактора x_i .

При $i = 1$ формула коэффициента частной корреляции примет вид

$$r_{yx_1 * x_2 \dots x_p} = \sqrt{1 - \frac{1 - R_{yx_1 x_2 \dots x_p}^2}{1 - R_{yx_2 \dots x_p}^2}}.$$

Данный коэффициент частной корреляции позволяет измерить тесноту связи между y и x_1 при неизменном уровне всех других факторов, включенных в уравнение регрессии.

Пример. Предположим, что зависимость объема продукции y от затрат труда x_1 характеризуется уравнением

$$\bar{y}_{x_1} = 27,5 + 3,5 \cdot x_1; \quad r_{yx_1} = 0,58.$$

Подставив в это уравнение фактические значения x_1 , найдем теоретические величины объема продукции y_{x_1} и соответствующую величину остаточной дисперсии S^2 :

$$S_{yx_1}^2 = \frac{\sum (y_i - \bar{y}_{x_1})^2}{n}.$$

Включив в уравнение регрессии дополнительный фактор x_2 – техническую оснащенность производства, получим уравнение регрессии вида

$$\hat{y}_{x_1 x_2} = 20,2 + 2,8 \cdot x_1 + 0,2 \cdot x_2.$$

Для этого уравнения остаточная дисперсия, естественно, меньше. Предположим, что $S_{x_1 x_2}^2 = 3,7$; $S_{yx_1}^2 = 6$. Чем большее число факторов включено в модель, тем меньше величина остаточной дисперсии.

Сокращение остаточной дисперсии за счет дополнительного включения фактора x_2 составит

$$S_{yx_1}^2 - S_{yx_1 x_2}^2 = 2,3.$$

Чем больше доля этого сокращения в остаточной вариации до введения дополнительного фактора, т. е. в $S_{yx_1}^2$, тем теснее связь между y и x_2 при постоянном действии фактора x_1 . Корень квадратный из этой величины и есть индекс частной корреляции, показывающий в «чистом» виде тесноту связи y с x_2 .

Следовательно, чистое влияние фактора x_2 на результат y можно определить как

$$r_{yx_2 * x_1} = \sqrt{\frac{S_{yx_1}^2 - S_{yx_1 x_2}^2}{S_{yx_1}^2}}.$$

Аналогично определяется и чистое влияние на результат фактора x_1

$$r_{yx_1 * x_2} = \sqrt{\frac{S_{yx_2}^2 - S_{yx_1 x_2}^2}{S_{yx_2}^2}}.$$

Если предположить, что $S_{yx_2}^2 = 5$, то частные показатели корреляции для уравнения $\hat{y}_{x_1 x_2} = 20,2 + 2,8 \cdot x_1 + 0,2 \cdot x_2$ составят

$$r_{yx_1 * x_2} = \sqrt{\frac{5 - 3,7}{5}} = 0,51 \quad \text{и} \quad r_{yx_2 * x_1} = \sqrt{\frac{6 - 3,7}{6}} = 0,619.$$

Сравнивая полученные результаты, видим, что более сильное воздействие на объем продукции оказывает техническая оснащенность предприятий.

Если выразить остаточную дисперсию через показатель детерминации $S_{ост}^2 = \sigma_y^2(1 - r^2)$, то формула коэффициента частной корреляции примет вид

$$r_{yx_1 * x_2} = \sqrt{\frac{S_{yx_2}^2 - S_{yx_1x_2}^2}{S_{yx_1}^2}} = \sqrt{1 - \frac{S_{yx_1x_2}^2}{S_{yx_2}^2}} = \sqrt{1 - \frac{1 - R_{yx_1x_2}^2}{1 - r_{yx_2}^2}}.$$

Соответственно

$$r_{yx_2 * x_1} = \sqrt{1 - \frac{1 - R_{yx_1x_2}^2}{1 - r_{yx_1}^2}}.$$

Рассмотренные показатели частной корреляции принято называть коэффициентами (индексами) частной корреляции первого порядка, ибо они фиксируют тесноту связи двух переменных при закреплении (элиминировании влияния) одного фактора.

Рассчитанные частные коэффициенты корреляции изменяются в пределах от 0 до +1. Сравнение их друг с другом позволяет ранжировать факторы по тесноте их связи с результатом.

Частные коэффициенты корреляции, подтверждая ранжировку факторов по их воздействию на результат, полученную на основе стандартизованных коэффициентов регрессии (β -коэффициентов), в отличие от последних, дают конкретную меру тесноты связи каждого фактора с результатом в чистом виде. Если из стандартизованного уравнения регрессии $t_y = \beta_{x_1} \cdot t_{x_1} + \beta_{x_2} \cdot t_{x_2} + \beta_{x_3} \cdot t_{x_3}$ по степени влияния на результат порядок факторов таков: x_1, x_2, x_3 , то тот же порядок факторов определяется и по соотношению частных коэффициентов корреляции, $r_{yx_1 * x_2x_3} > r_{yx_2 * x_1x_3} > r_{yx_3 * x_1x_2}$.

В эконометрике частные коэффициенты корреляции обычно не имеют самостоятельного значения. В основном их используют на стадии формирования модели, в частности в процедуре отсева факторов.

Так, строя многофакторную модель, например, методом исключения переменных, на первом шаге определяется уравнение регрессии с полным набором факторов и рассчитывается матрица частных коэффициентов корреляции.

На втором шаге отбирается фактор с наименьшей и несущественной по t -критерию Стьюдента величиной показателя частной корреляции (п. 2.8). Исключив его из модели, строится новое уравнение регрессии. Процедура продолжается до тех пор, пока не окажется, что все частные коэффициенты корреляции существенно отличаются от нуля. Если исключен несущественный фактор, то множественные коэффициенты детерминации на двух смежных шагах построения регрессионной модели почти не отличаются друг от друга, т. е.

$R_{p+1}^2 \approx R_p^2$, где p – число факторов.

2.8. Оценка надежности результатов множественной регрессии и корреляции

Качество построенной модели в целом оценивает коэффициент (индекс) детерминации.

Коэффициент (индекс) множественной детерминации рассчитывается как квадрат индекса множественной корреляции

$$R^2_{yx_1x_2\dots x_p}.$$

Скорректированный коэффициент (индекс) множественной детерминации содержит поправку на число степеней свободы и рассчитывается по формуле

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-m-1},$$

где n – число наблюдений;

m – число факторов.

Значимость уравнения множественной регрессии в целом оценивается с помощью F -критерия Фишера

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m}.$$

$F_{табл}$ определяется при заданном α (0,05; 0,01) и степенях свободы $k_1 = m$, $k_2 = n - m - 1$ (m – число параметров при факторных переменных в уравнении множественной регрессии).

Частный F -критерий оценивает статистическую значимость присутствия каждого из факторов в уравнении. В общем виде, для факторов частный F -критерий определится как

$$F_{част\ x_i} = \frac{R^2_{yx_1\dots x_i\dots x_p} - R^2_{yx_1\dots x_{i-1}x_{i+1}\dots x_p}}{1 - R^2_{yx_1\dots x_i\dots x_p}} \cdot \frac{n-m-1}{1}.$$

$F_{табл}$ для частного F -критерия определяется при заданном α и степенях свободы $k_1 = 1$, $k_2 = n - m - 1$.

Оценка значимости коэффициентов чистой регрессии с помощью t -критерия Стьюдента сводится к вычислению значения

$$t_{b_i} = \frac{b_i}{m_{b_i}} = \sqrt{F_{x_i}},$$

где m_{b_i} – средняя квадратическая ошибка коэффициента регрессии b_i , она может быть определена по следующей формуле:

$$m_{b_i} = \frac{\sigma_y \cdot \sqrt{1 - R^2_{yx_1\dots x_p}}}{\sigma_{x_i} \cdot \sqrt{1 - R^2_{x_i x_1\dots x_p}}} \cdot \frac{1}{\sqrt{n-m-1}}.$$

$t_{табл}$ определяется при заданном α и степени свободы $k = n - m - 1$.

2.9. Проверка остатков регрессии на гомоскедастичность

Для применения МНК требуется, чтобы дисперсия остатков была *гомоскедастичной*. Это значит, что для каждого значения фактора x_i остатки ε_i имеют одинаковую дисперсию. Если это условие не соблюдается, то имеет место *гетероскедастичность*.

При нарушении гомоскедастичности мы имеем неравенства

$$\sigma_{\varepsilon_i}^2 \neq \sigma_{\varepsilon_j}^2 \neq \sigma^2, \quad j \neq i.$$

При малом объеме выборки для оценки гетероскедастичности может использоваться метод Гольдфельда–Квандта. Основная идея теста Гольдфельда–Квандта состоит в следующем:

- 1) упорядочение n наблюдений по мере возрастания переменной x ;
- 2) исключение из рассмотрения C центральных наблюдений, при этом $(n - C) : 2 > p$, где p – число оцениваемых параметров;
- 3) разделение совокупности из $(n - C)$ наблюдений на две группы (соответственно с малыми и с большими значениями фактора x) и определение по каждой из групп уравнений регрессии;
- 4) определение остаточной суммы квадратов для первой (S_1) и второй (S_2) групп и нахождение их отношения: $R = S_2 : S_1$.

При выполнении нулевой гипотезы о гомоскедастичности отношение R будет удовлетворять F-критерию со степенями свободы $k_1 = (n - C - 2p) : 2$, $k_2 = (n - C - 2p) : 2$. Чем больше величина R превышает табличное значение F-критерия, тем более нарушена предпосылка о равенстве дисперсий остаточных величин.

Авторами метода рекомендовано для случая одного фактора $n=20$ принимать $C=4$, при $n=30$ принимать $C=8$, при $n=60$ принимать $C=16$.

Контрольные вопросы:

1. Что понимается под множественной регрессией?
2. Какие задачи решаются при построении уравнения регрессии?
3. Какие задачи решаются при спецификации модели?
4. Какие требования предъявляются к факторам, включаемым в уравнение регрессии?
5. Что понимается под коллинеарностью и мультиколлинеарностью факторов?
6. Как проверяется наличие коллинеарности и мультиколлинеарности?
7. Какие подходы применяются для преодоления межфакторной корреляции?
8. Какие функции чаще используются для построения уравнения множественной регрессии?
9. Какой вид имеет система нормальных уравнений метода наименьших квадратов в случае линейной регрессии?
10. По какой формуле вычисляется индекс множественной корреляции?
11. Как вычисляются индекс множественной детерминации и скорректированный индекс множественной детерминации?

12. Что означает низкое значение коэффициента (индекса) множественной корреляции?
13. Как проверяется значимость уравнения регрессии и отдельных коэффициентов?
14. Как строятся частные уравнения регрессии?
15. Как вычисляются средние частные коэффициенты эластичности?
16. Что такое стандартизированные переменные?
17. Какой вид имеет уравнение линейной регрессии в стандартизированном масштабе?
18. Как оценивается информативность (значимость) факторов?
19. Как вычисляются частные коэффициенты корреляции?
20. Опишите процедуру метода исключения переменных с использованием частных коэффициентов корреляции.
21. Что понимается под гомоскедастичностью?
22. Как проверяется гипотеза о гомоскедастичности ряда остатков?

Лабораторная работа № 3

Задание. На основании данных табл. П1 для соответствующего варианта (табл. 2.3):

1. Проверить наличие коллинеарности и мультиколлинеарности. Отобрать неколлинеарные факторы.
2. Построить уравнение линейной регрессии.
3. Определить коэффициент множественной корреляции.
4. Проверить значимость уравнения при уровнях значимости 0,05 и 0,01.
5. Построить частные уравнения регрессии.
6. Определить средние частные коэффициенты эластичности.

Лабораторная работа № 4

Задание. На основании данных табл. П1 для соответствующего варианта (табл. 2.3):

1. Построить уравнение линейной регрессии в стандартизированном масштабе.
2. Оценить информативность факторов на основе уравнения линейной регрессии в стандартизированном масштабе.
3. Вычислить частные коэффициенты корреляции.
4. Оценить их значимость при уровнях значимости 0,05 и 0,01.
5. Оценить информативность факторов на основе частных коэффициентов корреляции.
6. Построить уравнение регрессии с учетом только информативных факторов.
7. Проверить гипотезу о гомоскедастичности ряда остатков с уровнем значимости $\alpha = 0,05$.

Указания к решению. При выполнении лабораторной работы использовать возможности надстройки «Анализ данных» табличного процессора MS Excel (для расчета корреляционной матрицы, нахождения уравнений регрессии, нахождения коэффициентов координации и др.).

Требования к оформлению результатов

Отчет о лабораторной работе должен содержать разделы:

1. Описание задания;
2. Описание решения лабораторной работы (по этапам);
3. Изложение полученных результатов.

Пример выполнения лабораторной работы № 3

Исходные данные для выполнения лабораторной работы даны в таблице 2.1.

Табл. 2.1

	у	x1	x2	x3	у _{расч}	остатки
1	113	10	1	77	124,915	-11,915
2	124	5	2	64	121,515	2,485
3	124	10	2	77	126,676	-2,676
4	122	13	2	66	122,309	-0,309
5	128	9	1	71	122,533	5,467
6	140	14	6	81	135,308	4,692
7	117	12	1	58	117,372	-0,372
8	113	15	3	66	124,070	-11,070
9	122	13	2	73	125,088	-3,088
10	139	27	14	81	149,396	-10,396
11	126	8	6	73	132,132	-6,132
12	120	8	3	65	123,673	-3,673
13	125	24	6	66	129,353	-4,353
14	118	8	1	74	123,724	-5,724
15	122	8	4	64	125,037	-3,037
16	133	15	5	79	132,753	0,247
17	136	12	4	71	127,816	8,184
18	146	16	9	68	135,430	10,570
19	148	23	5	78	132,356	15,644
20	136	16	8	74	136,051	-0,051
21	138	10	3	64	123,276	14,724
22	124	12	7	74	134,290	-10,290
23	123	8	3	71	126,055	-3,055
24	149	29	8	87	141,212	7,788
25	130	9	4	56	121,861	8,139
26	117	91	3	65	123,673	-6,673
27	126	12	1	61	118,563	7,437
28	110	7	1	35	108,241	1,759
29	98	6	0	26	102,907	-4,907

1) Проверка наличия коллинеарности или мультиколлинеарности. Отбор неколлинеарных факторов.

Построим корреляционную матрицу, используя функцию «Сервис.Анализ данных.Корреляция» табличного процессора MS Excel.

	y	x ₁	x ₂	x ₃
y	1			
X ₁	0,638	1		
X ₂	0,680	0,710	1	
X ₃	0,661	0,513	0,506	1

Из матрицы следует, что наблюдается коллинеарность между факторами x₁ и x₂, так как $r_{x_1 x_2} = 0,710$. Для дальнейшего рассмотрения оставляем фактор x₂, так как он меньше коррелирует с фактором x₃ ($r_{x_2 x_3} = 0,506 < r_{x_1 x_3} = 0,513$).

Таким образом, далее будет строиться регрессия y на факторы x₂ и x₃.

2) Для построения уравнения линейной регрессии используем функцию «Сервис.Анализ данных.Регрессия» (рис 2.1). Задав соответствующие диапазоны данных в окне,

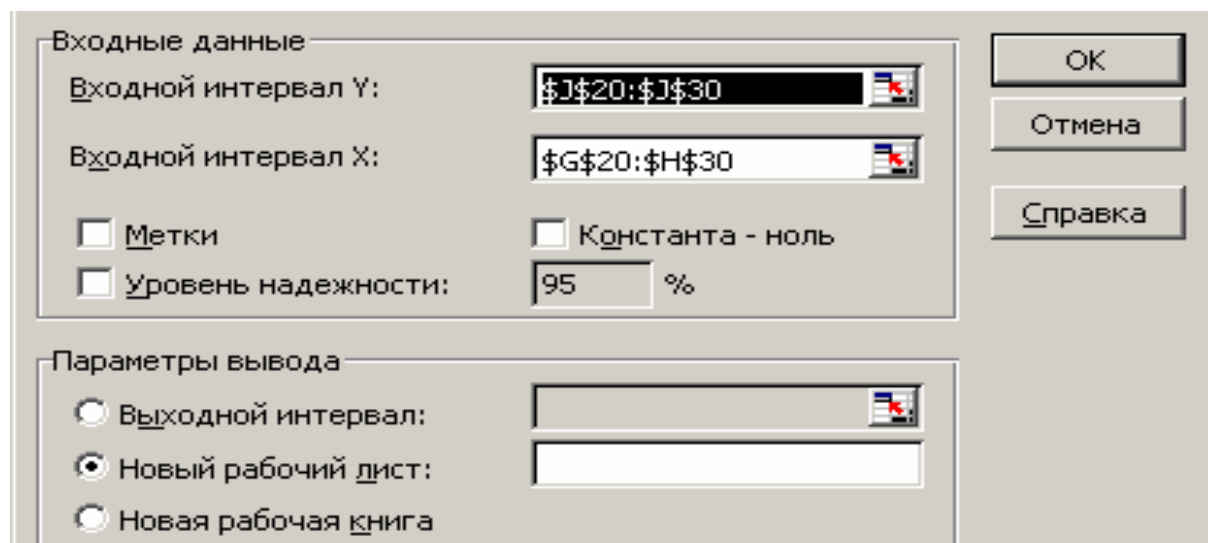


Рис. 2.1. Окно параметров регрессии

получим набор таблиц А, Б, В.

Табл. А

Показатель	Значение	Комментарий автора
Множественный R	0,773	Множественный коэффициент корреляции R
R-квадрат	0,597	Коэффициент координации R
Нормированный R-квадрат	0,566	
Стандартная ошибка	7,768	Стандартная ошибка определения R
Наблюдения	29	Число наблюдений

Дисперсионный анализ

Табл. Б

	Число степеней свободы	Дисперсия	Дисперсия на 1 степень свободы	Статистика Фишера	
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
Регрессия	2	2326,1	1163,1	19,3	7,35E-06
Остаток	26	1569,1	60,3		
Итого	28	3895,2			

Табл. В

	Коэффициенты уравнения регрессии	Стандартная ошибка определения коэффициентов	t-статистика	Вероятность ошибки	Нижние 95%–пределы	Верхние 95%–пределы
	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
У-пересечение	92,585	8,351	11,087	0,0000	75,420	109,750
Переменная X 1	1,761	0,547	3,219	0,0030	0,637	2,886
Переменная X 2	0,397	0,134	2,952	0,0070	0,120	0,673

Из табл. В следует, что уравнение регрессии имеет вид

$$y = 92,585 + 1,761 \cdot x_2 + 0,397 \cdot x_3.$$

3) Коэффициент множественной корреляции определяется из табл. А

$$R_{yx_1x_2\dots x_p} = 0,773.$$

4) Проверка значимости уравнения регрессии основана на использовании *F*-критерия Фишера. Фактическое значение критерия берется из табл. Б, т. е. $F_{\text{факт}} = 19,3$.

Для определения табличных значений используем встроенную функцию MS Excel «ФРАСПОБР» (рис. 2.2), задавая параметры $k_1 = 2$, $k_2 = 29 - 2 - 1 = 26$, $\alpha = 0,05$ и $\alpha = 0,01$.

ФРАСПОБР

Вероятность = ЧИСЛО

Степени_свободы1 = ЧИСЛО

Степени_свободы2 = ЧИСЛО

=

Возвращает обратное значение для F-распределения вероятностей: если $p = \text{ФРАСП}(x, \dots)$, то $\text{ФРАСПОБР}(p, \dots) = x$.

Рис. 2.1. Окно параметров регрессии

В результате получаем $F_{факт,0,05} = 3,369$, $F_{факт,0,01} = 5,526$. Откуда следует, что уравнение регрессии значимо и при $\alpha = 0,05$, и $\alpha = 0,01$.

5) Частные уравнения регрессии. Предварительно определим средние значения переменных

$$\bar{y} = 126,448, \quad \bar{x}_2 = 3,966, \quad \bar{x}_3 = 67,76.$$

С учетом средних значений построим частные уравнения регрессии

$$\hat{y}_{x_2, x_3} = 92,585 + 1,761 \cdot x_2 + 0,397 \cdot 67,76 = 119,486 + 1,761 \cdot x_2;$$

$$\hat{y}_{x_3, x_2} = 92,585 + 1,761 \cdot 3,966 + 0,397 \cdot x_3 = 99,569 + 0,397 \cdot x_3.$$

6) Средние частные коэффициенты эластичности

$$\bar{\varepsilon}_{yx_2} = b_2 \frac{\bar{x}_2}{\bar{y}} = 1,761 \cdot \frac{3,966}{126,448} = 0,0552;$$

$$\bar{\varepsilon}_{yx_3} = b_3 \frac{\bar{x}_3}{\bar{y}} = 0,397 \cdot \frac{67,76}{126,448} = 0,213.$$

Пример выполнения лабораторной работы № 4

Исходные данные возьмем из лабораторной работы № 3.

1) Построим уравнение линейной регрессии в стандартизированном масштабе. Его коэффициенты связаны с коэффициентами обычного уравнения регрессии соотношениями

$$b_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}} \quad \text{или} \quad \beta_i = b_i \frac{\sigma_{x_i}}{\sigma_y}.$$

Определим средние квадратические отклонения σ_y , σ_{x_i} , используя функцию MS Excel «СТАНДОТКЛОНП».

$$\sigma_y = 11,59, \quad \sigma_{x_2} = 3,057, \quad \sigma_{x_3} = 12,44.$$

Следовательно,

$$\beta_2 = 1,761 \cdot \frac{3,057}{11,59} = 0,464;$$

$$\beta_3 = 0,397 \cdot \frac{12,44}{11,59} = 0,426$$

и уравнение линейной регрессии в стандартизированном масштабе имеет вид

$$t_y = 0,464 \cdot t_{x_2} + 0,426 \cdot t_{x_3}.$$

2) Информативность факторов. Так как $\beta_1 = 0,464$ и $\beta_2 = 0,426$, то делаем вывод, что факторы практически одинаково информативны.

3) Частные коэффициенты корреляции. Для их вычисления воспользуемся формулой (2.14), в соответствии с которой необходимо вычислить величины

$R^2_{yx_2x_3}$, $R^2_{yx_2}$, $R^2_{yx_3}$. В данном примере величину $R^2_{yx_2x_3}$ можно взять из табл. А, а величины $R^2_{yx_2}$, $R^2_{yx_3}$ вычислить, используя соответствующие коэффициенты ли-

нейной корреляции r_{y, x_2} , r_{y, x_3} из корреляционной матрицы в примере лабораторной работы № 3

$$R_{yx_2x_3}^2 = 0,597, R_{yx_2}^2 = 0,4624, R_{yx_3}^2 = 0,4369.$$

В результате получим частные коэффициенты корреляции

$$r_{yx_2*x_3} = \sqrt{1 - \frac{1 - 0,597}{1 - 0,4369}} = 0,533;$$

$$r_{yx_3*x_2} = \sqrt{1 - \frac{1 - 0,597}{1 - 0,4624}} = 0,500.$$

4) Оценим их значимость. Вычислим фактические значения частного F -критерия Фишера

$$F_{\text{част } x_2} = \frac{R_{yx_2x_3}^2 - R_{yx_3}^2}{1 - R_{yx_2x_3}^2} \cdot \frac{n - 2 - 1}{1} = \frac{0,597 - 0,4369}{1 - 0,597} \cdot \frac{29 - 2 - 1}{1} = 10,33;$$

$$F_{\text{част } x_3} = \frac{R_{yx_2x_3}^2 - R_{yx_2}^2}{1 - R_{yx_2x_3}^2} \cdot \frac{n - 2 - 1}{1} = \frac{0,597 - 0,4624}{1 - 0,597} \cdot \frac{29 - 2 - 1}{1} = 8,68.$$

Для определения табличных значений используем встроенную функцию MS Excel «ФРАСПОБР» (рис. 2.2), задавая параметры $k_1 = 1$, $k_2 = 29 - 2 - 1 = 26$, $\alpha = 0,05$ и $\alpha = 0,01$. В результате получаем $F_{\text{факт}, 0,05} = 4,225$, $F_{\text{факт}, 0,01} = 7,721$. Откуда следует, что оба частных коэффициента корреляции значимы и при $\alpha = 0,05$, и при $\alpha = 0,01$.

5) Информативность факторов. Так как оба частных коэффициента значимы, то оба фактора x_2 и x_3 информативны и должны быть включены в уравнение регрессии.

6) Уравнение регрессии (из лабораторной работы № 3)

$$y = 92,585 + 1,761 \cdot x_2 + 0,397 \cdot x_3.$$

7) Проверка гомоскедастичности. Вычислим расчетные значения результативного признака по уравнению регрессии и определим остатки (результаты приведены в таблице 2.1). Согласно методу Гольдфельда–Квандта, упорядочим ряд остатков отдельно по фактору x_2 и по фактору x_3 . Результаты приведены в следующей таблице 2.2. Цветом отмечены данные, не участвующие в рассмотрении. Согласно рекомендациям, их число равно $C = 7$.

Таблица. 2.2

Остатки, упорядоченные по x_2				Остатки, упорядоченные по x_3			
	x_2	x_3	Остатки		x_2	x_3	остатки
1	0	26	-4,907	1	0	26	-4,907
2	1	77	-11,915	2	1	35	1,759
3	1	71	5,467	3	4	56	8,139
4	1	58	-0,372	4	1	58	-0,372
5	1	74	-5,724	5	1	61	7,437
6	1	61	7,437	6	2	64	2,485
7	1	35	1,759	7	4	64	-3,037
8	2	64	2,485	8	3	64	14,724
9	2	77	-2,676	9	3	65	-3,673
10	2	66	-0,309	10	3	65	-6,673
11	2	73	-3,088	11	2	66	-0,309
12	3	66	-11,070	12	3	66	-11,070
13	3	65	-3,673	13	6	66	-4,353
14	3	64	14,724	14	9	68	10,570
15	3	71	-3,055	15	1	71	5,467
16	3	65	-6,673	16	4	71	8,184
17	4	64	-3,037	17	3	71	-3,055
18	4	71	8,184	18	2	73	-3,088
19	4	56	8,139	19	6	73	-6,132
20	5	79	0,247	20	1	74	-5,724
21	5	78	15,644	21	8	74	-0,051
22	6	81	4,692	22	7	74	-10,290
23	6	73	-6,132	23	1	77	-11,915
24	6	66	-4,353	24	2	77	-2,676
25	7	74	-10,290	25	5	78	15,644
26	8	74	-0,051	26	5	79	0,247
27	8	87	7,788	27	6	81	4,692
28	9	68	10,570	28	14	81	-10,396
29	14	81	-10,396	29	8	87	7,788

а) Проверим гомоскедастичность по фактору x_2 . Построим уравнение регрессии на основе данных верхней части левой таблицы 2.2, используя функцию «Сервис.Анализ данных.Регрессия» (рис 2.1). Получим три таблицы

Табл. Г

Множественный R	0,33006
R-квадрат	0,10894
Нормированный R-квадрат	-0,11382
Стандартная ошибка	5,75599
Наблюдения	11

Табл. Д

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
Регрессия	2	32,4050	16,2025	0,489037	0,630416
Остаток	8	265,0514	33,1314		
Итого	10	297,4564			

Табл.Е

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
Y-пересечение	2,70928	6,935642	0,39063	0,706266	-13,2844	18,7029
Переменная X 1	3,12327	3,594754	0,86884	0,410241	-5,1663	11,4128
Переменная X 2	-0,12518	0,137506	-0,91034	0,389246	-0,4423	0,1919

Из табл. Д находим остаточную дисперсию (графа «SS») $S_1 = 265,05$. Аналогично определим остаточную дисперсию для уравнения регрессии для нижней части левой таблицы 2.2. упорядоченных остатков $S_2 = 624,21$. Найдем отношение

$$S_2 : S_1 = 624,21 : 265,05 = 2,36.$$

Определим критическое значение для теста Гольдфельда–Квандта как значение F -критерия со степенями свободы $k_1 = (n - C - 2p) : 2$, $k_2 = (n - C - 2p) : 2$. В нашем случае $C = 7$ и $p = 2$ (переменные x_2 и x_3)

$$k_1 = k_2 = 29 - 7 - 2 \cdot 2 = 18.$$

Соответствующее значение критерия при $\alpha = 0,05$ равно $F_{\text{факт},0,05} = 2,217$.

Так как $S_2 : S_1 = 2,36 > F_{\text{факт},0,05} = 2,217$, то нарушается предпосылка о равенстве дисперсий, т. е. о гомоскедастичности остатков по переменной x_2 .

б) Проверим гомоскедастичность по фактору x_3 . Действуя аналогично и используя правую часть таблицы 2.2, получим следующие величины остаточных дисперсий $S_1 = 403,39$, $S_2 = 563,0$.

Так как отношение $S_2 : S_1 = 563,0 : 403,39 = 1,40 < F_{\text{факт},0,05} = 2,217$, то предпосылка о равенстве дисперсий, т. е. о гомоскедастичности остатков по переменной x_3 , не нарушается.

Таблица 2.1

Варианты выполнения лабораторных работ № 3, 4

Варианты	Номер графы для результирующей переменной у (табл. 1.2)	Номера граф для переменных-факторов (табл. П1)								
		2	3	4	5	6	7	8	9	10
1	1	*	*	*						
2	1		*	*	*					
3	1			*	*	*				
4	1				*	*	*			
5	1					*	*	*		
6	1						*	*	*	
7	1							*	*	*
8	2		*	*	*					
9	2			*	*	*				
10	2				*	*	*			
11	2					*	*	*		
12	2						*	*	*	
13	2							*	*	*
14	3			*	*	*				
15	3				*	*	*			
16	3					*	*	*		
17	3						*	*	*	
18	3							*	*	*
19	4				*	*	*			
20	4					*	*	*		
21	4						*	*	*	
22	4							*	*	*
23	5					*	*	*		
24	5						*	*	*	
25	5							*	*	*

3. Системы эконометрических уравнений

3.1. Виды систем эконометрических уравнений

Сложные экономические процессы описывают с помощью системы взаимосвязанных (одновременных) уравнений.

Различают несколько видов систем уравнений, применяемых в эконометрике:

– *система независимых уравнений* – когда каждая зависимая переменная y рассматривается как функция одного и того же набора факторов x_i

$$\begin{aligned} y_1 &= a_{11} \cdot x_1 + a_{12} \cdot x_2 + \dots + a_{1m} \cdot x_m + \varepsilon_1; \\ y_2 &= a_{21} \cdot x_1 + a_{22} \cdot x_2 + \dots + a_{2m} \cdot x_m + \varepsilon_2; \\ &\dots\dots\dots \\ y_n &= a_{n1} \cdot x_1 + a_{n2} \cdot x_2 + \dots + a_{nm} \cdot x_m + \varepsilon_n. \end{aligned}$$

Для построения такой системы и нахождения ее параметров используется метод наименьших квадратов, применяемый к каждому уравнению в отдельности;

– *система рекурсивных уравнений* – когда зависимая переменная y одного уравнения выступает в виде фактора x в другом уравнении

$$\begin{aligned} y_1 &= a_{11} \cdot x_1 + a_{12} \cdot x_2 + \dots + a_{1m} \cdot x_m + \varepsilon_1; \\ y_2 &= b_{21} \cdot y_1 + a_{21} \cdot x_1 + a_{22} \cdot x_2 + \dots + a_{2m} \cdot x_m + \varepsilon_2; \\ y_3 &= b_{31} \cdot y_1 + b_{32} \cdot y_2 + a_{31} \cdot x_1 + a_{32} \cdot x_2 + \dots + a_{3m} \cdot x_m + \varepsilon_3; \\ &\dots\dots\dots \\ y_n &= b_{n1} \cdot y_1 + b_{n2} \cdot y_2 + \dots + b_{nn-1} \cdot y_{n-1} + a_{n1} \cdot x_1 + a_{n2} \cdot x_2 + \dots + a_{nm} \cdot x_m + \varepsilon_n. \end{aligned}$$

Для построения такой системы и нахождения ее параметров используется метод наименьших квадратов, применяемый последовательно к каждому уравнению в отдельности;

– *система взаимосвязанных (совместных) уравнений* – когда одни и те же зависимые переменные в одних уравнениях входят в левую часть, а в других – в правую

$$\begin{aligned} y_1 &= b_{12} \cdot y_2 + b_{13} \cdot y_3 + \dots + b_{1n} \cdot y_n + a_{11} \cdot x_1 + a_{12} \cdot x_2 + \dots + a_{1m} \cdot x_m + \varepsilon_1; \\ y_2 &= b_{21} \cdot y_1 + b_{23} \cdot y_3 + \dots + b_{2n} \cdot y_n + a_{21} \cdot x_1 + a_{22} \cdot x_2 + \dots + a_{2m} \cdot x_m + \varepsilon_2; \\ &\dots\dots\dots \\ y_n &= b_{n1} \cdot y_1 + b_{n2} \cdot y_2 + \dots + b_{nn-1} \cdot y_{n-1} + a_{n1} \cdot x_1 + a_{n2} \cdot x_2 + \dots + a_{nm} \cdot x_m + \varepsilon_n. \end{aligned}$$

Такая система уравнений называется **структурной формой модели**. Для построения таких систем и нахождения их параметров используются косвенный и двухшаговый методы наименьших квадратов.

Введем следующие определения.

Эндогенными переменными называются взаимозависимые переменные, которые определяются внутри модели (системы) – переменные y .

Экзогенными переменными называются независимые переменные, которые определяются вне системы – переменные x .

Предопределенными переменными называются экзогенные и лаговые (за предыдущие моменты времени y_{-1}, y_{-2}, \dots) эндогенные переменные системы.

Коэффициенты a и b при переменных носят название **структурных коэффициентов модели**.

Система линейных функций эндогенных переменных от всех предопределенных переменных системы называется **приведенной формой модели**

$$\hat{y}_1 = \delta_{11} \cdot x_1 + \delta_{12} \cdot x_2 + \dots + \delta_{1m} \cdot x_m;$$

$$\hat{y}_2 = \delta_{21} \cdot x_1 + \delta_{22} \cdot x_2 + \dots + \delta_{2m} \cdot x_m;$$

.....

$$\hat{y}_n = \delta_{n1} \cdot x_1 + \delta_{n2} \cdot x_2 + \dots + \delta_{nm} \cdot x_m,$$

где δ_{ij} – коэффициенты приведенной формы модели.

3.2. Проблема идентификации

При переходе от приведенной формы модели к структурной исследователь сталкивается с проблемой идентификации. **Идентификация** – это единственность соответствия между приведенной и структурной формами модели.

С позиции идентифицируемости структурные модели можно подразделить на три вида:

- идентифицируемые;
- неидентифицируемые;
- сверхидентифицируемые.

Модель идентифицируема, если все структурные ее коэффициенты определяются однозначно, единственным образом по коэффициентам приведенной формы модели, т. е. если число параметров структурной модели равно числу параметров приведенной формы модели. В этом случае структурные коэффициенты модели оцениваются через параметры приведенной формы модели и модель идентифицируема.

Модель неидентифицируема, если число приведенных коэффициентов меньше числа структурных коэффициентов, и в результате структурные коэффициенты не могут быть оценены через коэффициенты приведенной формы модели.

Модель сверхидентифицируема, если число приведенных коэффициентов больше числа структурных коэффициентов. В этом случае на основе коэффициентов приведенной формы можно получить два или более значений одного структурного коэффициента. В этой модели число структурных коэффициентов меньше числа коэффициентов приведенной формы.

Сверхидентифицируемая модель, в отличие от неидентифицируемой, модели практически решается, но требует для этого специальных методов исчисления параметров.

Структурная модель всегда представляет собой систему совместных уравнений, каждое из которых требуется проверять на идентификацию. Модель считается идентифицируемой, если каждое уравнение системы идентифицируемо. Если хотя бы одно из уравнений системы неидентифицируемо, то и вся модель считается неидентифицируемой. Сверхидентифицируемая модель содержит хотя бы одно сверхидентифицируемое уравнение.

Выполнение условия идентифицируемости модели проверяется для каждого уравнения системы. Чтобы уравнение было идентифицируемо, необходимо, чтобы число predetermined переменных, отсутствующих в данном уравнении, но присутствующих в системе, было равно числу эндогенных переменных в данном уравнении без одного.

Обозначим через H – число эндогенных переменных в уравнении, а через D – число predetermined переменных, отсутствующих в уравнении, но присутствующих в системе. Тогда **необходимое условие идентификации** отдельного уравнения принимает вид:

- уравнение идентифицируемо, если $D + 1 = H$;
- уравнение неидентифицируемо, если $D + 1 < H$;
- уравнение сверхидентифицируемо, если $D + 1 > H$.

Если необходимое условие выполнено, то далее проверяется достаточное условие идентификации.

Достаточное условие идентификации. Уравнение идентифицируемо, если 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.

3.3. Оценивание параметров структурной модели

Для решения идентифицируемых уравнений применяется косвенный метод наименьших квадратов, для решения сверхидентифицированных – двухшаговый метод наименьших квадратов.

Косвенный МНК состоит в следующем:

1) составляют приведенную форму модели и определяют численные значения параметров для каждого ее уравнения в отдельности с помощью обычного МНК;

2) путем алгебраических преобразований переходят от приведенной формы к уравнениям структурной формы модели, получая тем самым численные оценки структурных параметров.

Двухшаговый МНК заключается в следующем:

1) составляют приведенную форму модели и определяют численные значения параметров каждого ее уравнения в отдельности с помощью обычного МНК;

2) выявляют эндогенные переменные, находящиеся в правой части структурного уравнения (параметры которого определяют двухшаговым МНК) и находят расчетные значения по полученным на первом этапе соответствующим уравнениям приведенной формы модели;

3) с помощью обычного МНК определяют параметры каждого структурного уравнения в отдельности, используя в качестве исходных данных фактические значения predetermined переменных и расчетные значения эндогенных переменных, стоящих в правой части данного структурного уравнения, полученные на втором этапе.

Контрольные вопросы:

1. Какие виды систем уравнений применяются в эконометрике? Охарактеризуйте их.
2. Какие методы применяются для нахождения структурных коэффициентов модели для различных видов систем уравнений?
3. Какие переменные называются эндогенными, экзогенными, predetermined?
4. Что представляют собой структурная и приведенная форма модели?
5. Что понимается под идентификацией модели?
6. На какие виды подразделяются структурные модели с позиции идентифицируемости?
7. Что представляют собой необходимые и достаточные условия идентификации уравнения?
8. В каком случае применяется и что представляет собой косвенный МНК?
9. В каком случае применяется и что представляет собой двухшаговый МНК?

3.4. Решение типовых задач

Пример 1. Требуется:

1. Оценить следующую структурную модель на идентификацию:

$$y_1 = b_{13}y_3 + a_{11}x_1 + a_{13}x_3;$$

$$y_2 = b_{21}y_1 + b_{23}y_3 + a_{22}x_2;$$

$$y_3 = b_{32}y_2 + a_{31}x_1 + a_{33}x_3.$$

2. Исходя из приведенной формы модели уравнений

$$y_1 = 2 \cdot x_1 + 4 \cdot x_2 + 10 \cdot x_3;$$

$$y_2 = 3 \cdot x_1 - 6 \cdot x_2 + 2 \cdot x_3;$$

$$y_3 = -5 \cdot x_1 + 8 \cdot x_2 + 5 \cdot x_3,$$

найти структурные коэффициенты модели.

Решение:

1. Исследование модели на идентифицируемость. Модель имеет три эндогенные (y_1, y_2, y_3) и три экзогенные (x_1, x_2, x_3) переменные.

Проверим каждое уравнение системы на необходимое (Н) и достаточное (Д) условия идентификации.

Первое уравнение.

Необходимое условие (Н): эндогенных переменных – 2 (y_1, y_3), отсутствующих экзогенных – 1 (x_2).

Выполняется необходимое равенство: $2 = 1+1$, следовательно, уравнение точно идентифицируемо.

Достаточное условие (Д): в первом уравнении отсутствуют y_2 и x_2 . Построим матрицу из коэффициентов при них в других уравнениях системы

Уравнение	Отсутствующие переменные	
	Y_2	X_2
Второе	-1	a_{22}
Третье	b_{32}	0

Определитель матрицы $Det A = -1 \cdot 0 - b_{32} \cdot a_{22} \neq 0$.

Определитель матрицы не равен 0, ранг матрицы равен 2; следовательно, выполняется достаточное условие идентификации, и первое уравнение точно идентифицируемо.

Второе уравнение.

Н: эндогенных переменных – 3 (y_1, y_2, y_3), отсутствующих экзогенных – 2 (x_1, x_3).

Выполняется необходимое равенство: $3 = 2+1$, следовательно, уравнение точно идентифицируемо.

Д: во втором уравнении отсутствуют x_1 и x_3 . Построим матрицу из коэффициентов при них в других уравнениях системы:

Уравнение	Отсутствующие переменные	
	X_1	X_3
Первое	a_{11}	a_{13}
Третье	a_{31}	a_{33}

Определитель матрицы $Det A = a_{11} \cdot a_{33} - a_{31} \cdot a_{13} \neq 0$.

Определитель матрицы не равен 0, ранг матрицы равен 2, следовательно, выполняется достаточное условие идентификации, и второе уравнение точно идентифицируемо.

Аналогично доказывается, что и третье уравнение точно идентифицируемо.

Следовательно, исследуемая система точно идентифицируема и может быть решена косвенным методом наименьших квадратов.

2. Вычисление структурных коэффициентов модели:

1) из третьего уравнения приведенной формы выразим x_2 (так как его нет в первом уравнении структурной формы)

$$x_2 = \frac{y_3 + 5 \cdot x_1 - 5 \cdot x_3}{8}.$$

Данное выражение содержит переменные y_3, x_1 и x_3 , которые входят в правую часть первого уравнения структурной формы модели (СФМ). Подставим полученное выражение x_2 в первое уравнение приведенной формы модели (ПФМ)

$$y_1 = 2 \cdot x_1 + 4 \cdot \frac{y_3 + 5 \cdot x_1 - 5 \cdot x_3}{8} + 10 \cdot x_3.$$

Откуда получим первое уравнение СФМ в виде

$$y_1 = 0,5 \cdot y_3 + 4,5 \cdot x_1 + 7,5 \cdot x_3.$$

2) во втором уравнении СФМ нет переменных x_1 и x_3 . Структурные параметры второго уравнения СФМ можно будет определить в два этапа.

Первый этап: выразим x_1 в данном случае из первого или третьего уравнения ПФМ. Например, из первого уравнения

$$x_1 = \frac{y_1 - 4 \cdot x_2 - 10 \cdot x_3}{2} = 0,5 \cdot y_1 - 2 \cdot x_2 - 5 \cdot x_3.$$

Подстановка данного выражения во второе уравнение ПФМ не решило бы задачу до конца, так как в выражении присутствует x_3 , которого нет в СФМ.

Выразим x_3 из третьего уравнения ПФМ

$$x_3 = \frac{y_3 - 5 \cdot x_1 - 8 \cdot x_2}{5}.$$

Подставим его в выражение для x_1

$$x_1 = 0,5 \cdot y_1 - 2 \cdot x_2 - 5 \cdot \frac{y_3 + 5 \cdot x_1 - 8 \cdot x_2}{5} = 0,5 \cdot y_1 - y_3 + 6 \cdot x_2 - 5 \cdot x_1;$$

$$x_1 = \frac{0,5 \cdot y_1 - y_3 + 6 \cdot x_2}{6}.$$

Второй этап: аналогично, чтобы выразить x_3 через искомые y_1 , y_3 и x_2 , заменим в выражении x_3 значение x_1 на полученное из первого уравнения ПФМ

$$x_3 = \frac{y_3 + 5 \cdot (0,5 \cdot y_1 - 2 \cdot x_2 - 5 \cdot x_3) - 8 \cdot x_2}{5} =$$

$$= 0,2 \cdot y_3 + 0,5 \cdot y_1 - 3,6 \cdot x_2 - 5 \cdot x_3.$$

Следовательно,

$$x_3 = 0,033 \cdot y_3 + 0,083 \cdot y_1 - 0,6 \cdot x_2.$$

Подставим полученные x_1 и x_3 во второе уравнение ПФМ

$$y_2 = 3 \cdot \frac{0,5 \cdot y_1 - y_3 + 6 \cdot x_2}{6} - 6 \cdot x_2 + 2 \cdot (0,033 \cdot y_3 + 0,083 \cdot y_1 - 0,6 \cdot x_2).$$

В результате получаем второе уравнение СФМ

$$y_2 = 0,416 \cdot y_1 - 0,434 \cdot y_3 - 4,2 \cdot x_2.$$

3) из второго уравнения ПФМ выразим x_2 , так как его нет в третьем уравнении СФМ

$$x_2 = \frac{-y_2 + 3 \cdot x_1 + 2 \cdot x_3}{6} = -0,167 \cdot y_2 + 0,5 \cdot x_1 + 0,333 \cdot x_3.$$

Подставим полученное выражение в третье уравнение ПФМ

$$y_3 = -5 \cdot x_1 + 8 \cdot (-0,167 \cdot y_2 + 0,5 \cdot x_1 + 0,333 \cdot x_3) + 5 \cdot x_3.$$

В результате получаем третье уравнение СФМ

$$y_3 = -1,336 \cdot y_2 - x_1 + 7,664 \cdot x_3.$$

Таким образом, СФМ примет вид

$$y_1 = 0,5 \cdot y_3 + 4,5 \cdot x_1 + 7,5 \cdot x_3;$$

$$y_2 = 0,416 \cdot y_1 - 0,434 \cdot y_3 - 4,2 \cdot x_2;$$

$$y_3 = -1,336 \cdot y_2 - x_1 + 7,664 \cdot x_3.$$

Пример 2. Изучается модель вида

$$y = a_1 + b_1(C + D) + \varepsilon_1;$$

$$C = a_2 + b_2 \cdot y + b_3 \cdot y_{-1} + \varepsilon_2,$$

где y – валовой национальный доход;

y_{-1} – валовой национальный доход предшествующего года;

C – личное потребление;

D – конечный спрос (помимо личного потребления);

ε_1 и ε_2 – случайные составляющие.

Информация за девять лет о приростах всех показателей дана в табл. 3.1.

Таблица 3.1

Год	D	y_{-1}	y	C	Год	D	y_{-1}	y	C
1	-6,8	46,7	3,1	7,4	6	44,7	17,8	37,2	8,6
2	22,4	3,1	22,8	30,4	7	23,1	37,2	35,7	30,0
3	-17,3	22,8	7,8	1,3	8	51,2	35,7	46,6	31,4
4	12,0	7,8	21,4	8,7	9	32,3	46,6	56,0	39,1
5	5,9	21,4	17,8	25,8	Σ	167,5	239,1	248,4	182,7

Для данной модели была получена система приведенных уравнений

$$y = 8,219 + 0,6688 \cdot D + 0,261 \cdot y_{-1};$$

$$C = 8,636 + 0,3384 \cdot D + 0,202 \cdot y_{-1}.$$

Требуется:

1. Провести идентификацию модели.
2. Рассчитать параметры первого уравнения структурной модели.

Решение:

1. В данной модели две эндогенные переменные (y и C) и две экзогенные переменные (D и y_{-1}). Второе уравнение точно идентифицировано, так как содержит две эндогенные переменные и не содержит одну экзогенную переменную из системы. Иными словами, для второго уравнения имеем по счетному правилу идентификации равенство: $2=1+1$.

Первое уравнение сверхидентифицировано, так как в нем на параметры при C и D наложено ограничение: они должны быть равны. В этом уравнении содержится одна эндогенная переменная y . Переменная C в данном уравнении не рассматривается как эндогенная, так как она участвует в уравнении не само-

стоятельно, а вместе с переменной D . В данном уравнении отсутствует одна экзогенная переменная, имеющаяся в системе. По счетному правилу идентификации получаем: $1 + 1 = 2$; $D + 1 > H$. Это больше, чем число эндогенных переменных в данном уравнении, следовательно, система сверхидентифицирована.

2. Для определения параметров сверхидентифицированной модели используется двухшаговый метод наименьших квадратов.

Шаг 1. На основе системы приведенных уравнений по точно идентифицированному второму уравнению определим теоретические значения эндогенной переменной C . Для этого в приведенное уравнение

$$C = 8,636 + 0,3384 \cdot D + 0,202 \cdot y_{-1}$$

подставим значения D и y_{-1} , имеющиеся в условии задачи. Полученные значения обозначим \hat{C}_i ($i = 1, \dots, 9$) (табл. 3.2).

Шаг 2. По сверхидентифицированному уравнению структурной формы модели заменяем фактические значения C на теоретические \hat{C} и рассчитываем новую переменную $\hat{C} + D$ (табл. 3.2).

Таблица 3.2

Год	D	\hat{C}	$\hat{C} + D$	Год	D	\hat{C}	$\hat{C} + D$
1	-6,8	15,8	9,0	6	44,7	27,4	72,1
2	22,4	16,8	39,2	7	23,1	24,0	47,1
3	-17,3	7,4	-9,9	8	51,2	33,2	84,4
4	12,0	14,3	26,3	9	32,3	29,0	61,3
5	5,9	15,0	20,9	Σ	167,5	182,9	350,4

Далее к сверхидентифицированному уравнению применяется метод наименьших квадратов. Обозначим новую переменную $\hat{C} + D$ через Z . Решаем уравнение

$$y = a_1 + b_1 \cdot Z.$$

С помощью МНК получим $a_1 = 7,678$; $b_1 = 0,542$.

Запишем первое уравнение структурной модели

$$y = 7,678 + 0,542 \cdot (C + D).$$

Пример 3. Рассматривается следующая модель:

$$C_t = a_1 + b_{11} \cdot Y_t + b_{12} \cdot C_{t-1} + u_1 \quad (\text{функция потребления});$$

$$I_t = a_2 + b_{21} \cdot r_t + b_{22} \cdot I_{t-1} + u_2 \quad (\text{функция инвестиций});$$

$$r_t = a_3 + b_{31} \cdot Y_t + b_{32} \cdot M_{t-1} + u_3 \quad (\text{функция денежного рынка});$$

$$Y_t = C_t + I_t + G_t \quad (\text{тождество дохода}),$$

где C_t – расходы на потребление в период t ;

Y_t – совокупный доход в период t ;

I_t – инвестиции в период t ;

r_t – процентная ставка в период t ;

M_t – денежная масса в период t ;
 G_t – государственные расходы в период t ;
 C_{t-1} – расходы на потребление в период $t-1$;
 I_{t-1} – инвестиции в период $t-1$;
 u_1, u_2, u_3 – случайные ошибки.

Требуется:

1. В предположении, что имеются временные ряды данных по всем переменным модели, предложить способ оценки ее параметров.
2. Как изменится ваш ответ на вопрос п. 1, если из модели исключить тождество дохода?

Решение:

1. Модель представляет собой систему одновременных уравнений. Для ответа на вопрос о способе оценки параметров модели проверим каждое ее уравнение на идентификацию.

Модель включает четыре эндогенные переменные (C_t , I_t , Y_t , и r_t) и четыре predetermined переменные (две экзогенные переменные – M_t и G_t (и две лаговые эндогенные переменные – C_{t-1} и I_{t-1}).

Проверим необходимое условие идентификации для уравнений модели.

I уравнение.

Это уравнение включает две эндогенные переменные (C_t и Y_t) и одну predetermined переменную (C_{t-1}). Следовательно, число predetermined переменных, не входящих в это уравнение, плюс 1, больше числа эндогенных переменных, входящих в уравнение:

$3 + 1 > 2$. Уравнение сверхидентифицировано.

II уравнение.

Уравнение II включает две эндогенные переменные (I_t и r_t) и не включает три predetermined переменные. Как и I уравнение, оно сверхидентифицировано.

III уравнение.

Уравнение III тоже включает две эндогенные переменные (Y_t , и r_t) и не включает три predetermined переменные. Это уравнение сверхидентифицировано.

IV уравнение.

Уравнение IV представляет собой тождество, параметры которого известны. Необходимости в его идентификации нет.

Проверим для каждого из уравнений достаточное условие идентификации. Для этого составим матрицу коэффициентов при переменных модели

	C_t	Y_t	C_{t-1}	I_t	r_t	I_{t-1}	M_t	G_t
I уравнение	-1	b_{11}	b_{12}	0	0	0	0	0
II уравнение	0	0	0	-1	b_{21}	b_{22}	0	0
III уравнение	0	b_{31}	0	0	-1	0	b_{32}	0
Тождество	1	-1	0	1	0	0	0	1

В соответствии с достаточным условием идентификации определитель матрицы коэффициентов при переменных, не входящих в исследуемое уравнение, не должен быть равен нулю, а ранг матрицы должен быть равен числу эндогенных переменных модели минус 1, т. е. $4-1=3$.

I уравнение.

Матрица коэффициентов при переменных, не входящих в уравнение, имеет вид

$$A = \begin{pmatrix} -1 & b_{21} & b_{22} & 0 & 0 \\ 0 & -1 & 0 & b_{32} & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Ее ранг равен 3, так как определитель квадратной подматрицы 3x3 этой матрицы не равен нулю

$$\text{Det}A^* = \begin{vmatrix} -1 & b_{21} & 0 \\ 0 & -1 & 0 \\ 1 & 0 & 1 \end{vmatrix} \neq 0.$$

Достаточное условие идентификации для I уравнения выполняется.

II уравнение.

Выпишем матрицу коэффициентов при переменных, не входящих в уравнение

$$A = \begin{pmatrix} -1 & b_{11} & b_{12} & 0 & 0 \\ 0 & b_{31} & 0 & b_{32} & 0 \\ 1 & -1 & 0 & 0 & 1 \end{pmatrix}.$$

Ее ранг равен трем, так как имеется квадратная подматрица 3 x 3 этой матрицы, определитель которой не равен нулю.

Достаточное условие идентификации для II уравнения выполняется.

III уравнение.

Выпишем матрицу коэффициентов при переменных, не входящих в уравнение

$$A = \begin{pmatrix} -1 & b_{12} & 0 & 0 & 0 \\ 0 & 0 & -1 & b_{22} & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

Ее ранг также равен трем.

Достаточное условие идентификации для III уравнения выполняется.

Таким образом, все уравнения модели сверхидентифицированы. Для оценки параметров каждого из уравнений будем применять двухшаговый МНК.

Шаг 1. Запишем приведенную форму модели в общем виде

$$C_t = A_1 + A_2 \cdot C_{t-1} + A_3 \cdot I_{t-1} + A_4 \cdot M_t + A_5 \cdot G_t + v_1;$$

$$I_t = B_1 + B_2 \cdot C_{t-1} + B_3 \cdot I_{t-1} + B_4 \cdot M_t + B_5 \cdot G_t + v_2;$$

$$Y_t = D_1 + D_2 \cdot C_{t-1} + D_3 \cdot I_{t-1} + D_4 \cdot M_t + D_5 \cdot G_t + v_3;$$

$$r_t = E_1 + E_2 \cdot C_{t-1} + E_3 \cdot I_{t-1} + E_4 \cdot M_t + E_5 \cdot G_t + v_4,$$

где v_1, v_2, v_3, v_4 – случайные ошибки.

Определим параметры каждого из приведенных выше уравнений в отдельности обычным МНК. Затем найдем расчетные значения эндогенных переменных Y_t, r_t используемых в правой части структурной модели, подставляя в каждое уравнение приведенной формы соответствующее значение предопределенных переменных.

Шаг 2. В исходных структурных уравнениях заменим эндогенные переменные, выступающие в качестве факторных признаков, их расчетными значениями

$$C_t = a_1 + b_{11} \cdot \hat{Y}_t + b_{12} \cdot C_{t-1} + u_1^*, \quad \text{где} \quad u_1^* = u_1 + b_{11} \cdot v_1;$$

$$I_t = a_2 + b_{21} \cdot \hat{r}_t + b_{22} \cdot I_{t-1} + u_2^*, \quad \text{где} \quad u_2^* = u_2 + b_{21} \cdot v_2;$$

$$r_t = a_3 + b_{31} \cdot \hat{Y}_t + b_{32} \cdot M_{t-1} + u_3^*, \quad \text{где} \quad u_3^* = u_3 + b_{31} \cdot v_3.$$

Применяя к каждому из полученных уравнений в отдельности обычный МНК, определим структурные параметры $a_1, b_{11}, b_{12}, a_2, b_{21}, b_{22}, a_3, b_{31},$ и b_{32} .

2. Если из модели исключить тождество дохода, число предопределенных переменных модели уменьшится на 1 (из модели будет исключена переменная G_t). Число эндогенных переменных модели также снизится на единицу – переменная Y_t станет экзогенной. В правых частях функции потребления и функции денежного рынка будут находиться только предопределенные переменные. Функция инвестиций постулирует зависимость эндогенной переменной I_t , от эндогенной переменной r_t (которая зависит только от предопределенных переменных) и предопределенной переменной I_{t-1} . Таким образом, мы получим рекурсивную систему. Ее параметры можно оценивать обычным МНК, и нет необходимости исследования системы уравнения на идентификацию.

Лабораторная работа № 5

Задание. По заданным исходным данным для заданной модели (в соответствии с вариантом):

- 1) выделить эндогенные и экзогенные переменные;
- 2) применив необходимое и достаточное условие идентификации, определить, идентифицировано ли каждое из уравнений модели;
- 3) определить метод оценки параметров модели;
- 4) записать приведенную форму модели;
- 5) определить коэффициенты приведенной формы модели;
- 6) определить коэффициенты структурной формы модели;

7) проверить значимость полученных уравнений и их коэффициентов.

Указания к решению. Для нахождения приведенных уравнений (а также коэффициентов структурных уравнений при применении ДМНК) рекомендуется использовать табличный процессор Excel (надстройка «Анализ данных», функция – расчет уравнения регрессии):

1) вызов модуля для нахождения регрессии – пункты меню: Сервис – Анализ данных – Регрессия.

2) указать ячейки, содержащие исходные значения y и x .

3) если отсутствует свободный член в уравнении регрессии – установить флажок «Константа–ноль».

Искомые значения коэффициентов линейного уравнения регрессии (a , b_i) берутся из столбца «Коэффициенты» таблицы результатов регрессии.

Требования к оформлению результатов

Отчет о лабораторной работе должен содержать разделы:

1. Описание задания;
2. Описание решения лабораторной работы (по этапам);
3. Изложение полученных результатов.

Варианты заданий к лабораторным работам № 5

Если иное не оговорено, то исходные данные берутся из табл. П2.

Вариант 1

Модель денежного рынка:

$$R_t = a_1 + b_{11} \cdot M_t + b_{12} \cdot Y_t + \varepsilon_1;$$

$$Y_t = a_2 + b_{21} \cdot R_t + b_{22} \cdot I_t + \varepsilon_2,$$

где R – процентная ставка;

Y – ВВП;

M – денежная масса;

I – внутренние инвестиции;

t – текущий период.

Вариант 2

Модель Менгеса:

$$Y_t = a_1 + b_{11} \cdot Y_{t-1} + b_{12} \cdot I_t + \varepsilon_1;$$

$$I_t = a_2 + b_{21} \cdot Y_t + b_{22} \cdot Q_t + \varepsilon_2;$$

$$C_t = a_3 + b_{31} \cdot Y_t + b_{32} \cdot C_{t-1} + b_{33} \cdot P_t + \varepsilon_3;$$

$$Q_t = a_4 + b_{41} \cdot Q_{t-1} + b_{42} \cdot R_t + \varepsilon_4,$$

где Y – национальный доход;

C – расходы на личное потребление;

I – чистые инвестиции;

Q – валовая прибыль экономики;

P – индекс стоимости жизни;
 R – объем продукции промышленности;
 t – текущий период;
 $t-1$ – предыдущий период.

Вариант 3

Одна из версий модифицированной модели Кейнса имеет вид

$$C_t = a_1 + b_{11} \cdot Y_t + b_{12} \cdot Y_{t-1} + \varepsilon_1;$$

$$I_t = a_2 + b_{21} \cdot Y_t + b_{22} \cdot Y_{t-1} + \varepsilon_2;$$

$$Y_t = C_t + I_t + G_t,$$

где C – расходы на потребление;
 Y – доход;
 I – инвестиции;
 G – государственные расходы;
 t – текущий период;
 $t-1$ – предыдущий период.

Вариант 4

Модель мультипликатора-акселератора:

$$C_t = a_1 + b_{11} \cdot R_t + b_{12} \cdot C_{t-1} + \varepsilon_1;$$

$$I_t = a_2 + b_{21} \cdot (R_t - R_{t-1}) + \varepsilon_2;$$

$$R_t = C_t + I_t,$$

где C – расходы на потребление;
 R – доход;
 I – инвестиции;
 t – текущий период;
 $t-1$ – предыдущий период.

Вариант 5

Конъюнктурная модель имеет вид

$$C_t = a_1 + b_{11} \cdot Y_t + b_{12} \cdot C_{t-1} + \varepsilon_1;$$

$$I_t = a_2 + b_{21} \cdot r_t + b_{22} \cdot I_{t-1} + \varepsilon_2;$$

$$r_t = a_3 + b_{31} \cdot Y_t + b_{32} \cdot M_t + \varepsilon_3;$$

$$Y_t = C_t + I_t + G_t,$$

где C – расходы на потребление;
 Y – ВВП;
 I – инвестиции;
 r – процентная ставка;
 M – денежная масса;
 G – государственные расходы;
 t – текущий период;
 $t-1$ – предыдущий период.

Вариант 6

Модель протекционизма Сальватора (упрощенная версия):

$$M_t = a_1 + b_{12} \cdot N_t + b_{13} \cdot S_t + b_{14} \cdot E_{t-1} + b_{15} \cdot M_{t-1} + \varepsilon_1;$$

$$N_t = a_2 + b_{21} \cdot M_t + b_{23} \cdot S_t + b_{26} \cdot Y_t + \varepsilon_2;$$

$$S_t = a_3 + b_{31} \cdot M_t + b_{32} \cdot N_t + b_{37} \cdot X_t + \varepsilon_3,$$

где M – доля импорта в ВВП;

N – общее число прошений об освобождении от таможенных пошлин;

S – число удовлетворенных прошений об освобождении от таможенных пошлин;

E – фиктивная переменная, равная 1 для тех лет, в которые курс рубля на международных валютных рынках был искусственно завышен, и 0 – для всех остальных лет;

Y – реальный ВВП;

X – реальный объем чистого экспорта;

t – текущий период;

$t-1$ – предыдущий период.

Исходные данные

Текущий период t	Реальный ВВП, Y	Доля импорта в ВВП, M	Общее число прошений об освобождении от таможенных пошлин N	число удовлетворенных прошений об освобождении от таможенных пошлин, S	Фиктивная переменная, E	Реальный объем чистого экспорта, X
1	1 398,5	0,129471	900	800	1	185,6
2	19 005,5	0,482632	2 200	1 500	1	11 847
3	171 509,5	0,304956	5 500	4 000	1	65 524
4	610 745,2	0,232131	10 500	6 000	1	169 534
5	1 524 049,0	0,242857	20 350	18 000	1	426 735
6	2 145 655,5	0,205965	20 000	15 000	0	532 239
7	2 478 594,1	0,209359	30 000	20 000	0	592 332
8	2 741 051,2	0,234951	45 000	35 000	0	840 596
9	4 757 233,7	0,268908	50 000	45 000	0	2 090 687
10	7 063 392,8	0,249292	45 000	40 000	0	3 232 388

Вариант 7

Макроэкономическая модель (упрощенная версия модели Клейна):

$$C_t = a_1 + b_{11} \cdot Y_t + b_{13} \cdot T_t + \varepsilon_1;$$

$$I_t = a_2 + b_{21} \cdot Y_t + b_{24} \cdot K_{t-1} + \varepsilon_2;$$

$$Y_t = C_t + I_t,$$

где C – потребление;

I – инвестиции;

Y – доход;
 T – налоги;
 K – запас капитала;
 t – текущий период;
 $t-1$ – предыдущий период.

Вариант 8

Макроэкономическая модель экономики России (одна из версий):

$$C_t = a_1 + b_{11} \cdot Y_t + b_{12} \cdot C_{t-1} + \varepsilon_1 \quad (\text{функция потребления});$$

$$I_t = a_2 + b_{21} \cdot Y_t + b_{23} \cdot r_t + \varepsilon_2 \quad (\text{функция инвестиций});$$

$$r_t = a_3 + b_{31} \cdot Y_t + b_{34} \cdot M_t + b_{35} \cdot r_{t-1} + \varepsilon_3 \quad (\text{функция денежного рынка});$$

$$Y_t = C_t + I_t + G_t \quad (\text{тождество дохода}),$$

где C – потребление;
 Y – ВВП;
 I – инвестиции;
 r – процентная ставка;
 M – денежная масса;
 G – государственные расходы;
 t – текущий период;
 $t-1$ – предыдущий период.

Вариант 9

Модель Кейнса (одна из версий):

$$C_t = a_1 + b_{11} \cdot Y_t + b_{12} \cdot Y_{t-1} + \varepsilon_1 \quad (\text{функция потребления});$$

$$I_t = a_2 + b_{21} \cdot Y_t + \varepsilon_2 \quad (\text{функция инвестиций});$$

$$Y_t = C_t + I_t + G_t \quad (\text{тождество дохода}),$$

где C – потребление;
 Y – ВВП;
 I – валовые инвестиции;
 G – государственные расходы;
 t – текущий период;
 $t-1$ – предыдущий период.

Вариант 10

Модель денежного и товарного рынков:

$$R_t = a_1 + b_{11} \cdot Y_t + b_{13} \cdot r_t + \varepsilon_1 \quad (\text{функция товарного рынка});$$

$$Y_t = a_2 + b_{21} \cdot R_t + b_{23} \cdot I_t + b_{25} \cdot G_t + \varepsilon_2 \quad (\text{функция денежного рынка});$$

$$I_t = a_3 + b_{31} \cdot R_t + \varepsilon_2 \quad (\text{функция инвестиций}),$$

где R – процентные ставки;
 Y – реальный ВВП;
 M – денежная масса;
 I – внутренние инвестиции;

G – реальные государственные расходы.

Вариант 11

Для прогнозирования спроса на свою продукцию предприятия используется следующая модель, характеризующая общую экономическую ситуацию:

$$Q_t = a_1 + b_{11} \cdot Y_t + \varepsilon_1;$$

$$C_t = a_2 + b_{21} \cdot Y_t + \varepsilon_2;$$

$$I_t = a_3 + b_{32} \cdot (Y_{t-1} - K_{t-1}) + \varepsilon_3;$$

$$Y_t = C_t + I_t,$$

где Q – реализованная продукция в период t ;

Y – ВДС;

C – конечное потребление;

I – инвестиции;

K – запас капитала;

t – текущий период;

$t-1$ – предыдущий период.

Исходные данные

Текущий период t	ВДС региона, Y	Инвестиции, I	Конечное потребление, C	Реализованная продукция в период t , Q	Запас капитала, K
1	560,5	210,5	450	52	325
2	12 170	2 670	7 500	250	4 550
3	77 725	271 25	40 600	790	34 965
4	292 810	108 810	124 000	1390,4	133 209
5	476 974	266 974	310 000	7318,2	327 941
6	735 998	375 998	260 000	7524,4	454 369
7	698 797	408 797	390 000	7323,6	482 451
8	797 086	407 086	490 000	8804,5	485 452
9	2 160 439	970 439	990 000	13130,3	766 672
10	2 415 181	1 165 181	1 650 000	14874,2	1 293 750

Вариант 12

Модифицированная модель Кейнса:

$$C_t = a_1 + b_{11} \cdot Y_t + \varepsilon_1;$$

$$I_t = a_2 + b_{21} \cdot Y_t + b_{22} \cdot Y_{t-1} + \varepsilon_2;$$

$$Y_t = C_t + I_t + G_t,$$

где C – расходы на потребление;

Y – доход;

I – инвестиции;

G – государственные расходы;
 t – текущий период;
 $t-1$ – предыдущий период.

Вариант 13

Макроэкономическая модель:

$$\begin{aligned} C_t &= a_1 + b_{11} \cdot D_t + \varepsilon_1; \\ I_t &= a_2 + b_{22} \cdot Y_t + b_{23} \cdot Y_{t-1} + \varepsilon_2; \\ Y_t &= D_t + T_t; \\ D_t &= C_t + I_t + G_t, \end{aligned}$$

где C – расходы на потребление;
 Y – чистый национальный продукт;
 D – чистый национальный доход;
 I – инвестиции;
 T – налоги;
 G – государственные расходы;
 t – текущий период;
 $t-1$ – предыдущий период.

Вариант 14

Дана следующая структурная форма модели:

$$\begin{aligned} C_t &= b_1 + b_2 \cdot S_t + b_3 \cdot P_t + \varepsilon_1; \\ S_t &= a_1 + a_2 \cdot R_t + a_3 \cdot R_{t-1} + a_4 \cdot t + \varepsilon_2; \\ R_t &= S_t + P_t, \end{aligned}$$

где C_t – личное потребление в период t ;
 S_t – зарплата в период t ;
 P_t – прибыль в период t ;
 R_t – национальный доход в период t ;
 R_{t-1} – национальный доход в период $t-1$;
 $t-1$ – предыдущий период.

Вариант 15

Эконометрическая модель имеет вид

$$\begin{aligned} C_t &= a + b \cdot Y_t + \varepsilon; \\ Y_t &= C_t + I_t, \end{aligned}$$

где C – расходы на потребление;
 Y – доход;
 I – инвестиции;
 t – текущий период.

Вариант 16

Гипотетическая модель экономики:

$$C_t = a_1 + b_{11} \cdot Y_t + b_{12} \cdot J_t + \varepsilon_1;$$

$$I_t = a_2 + b_{21} \cdot Y_{t-1} + \varepsilon_2;$$

$$T_t = a_3 + b_{31} \cdot Y_t + \varepsilon_3;$$

$$Y_t = C_t + J_t + G_t,$$

где C – совокупное потребление в период t ;
 Y – совокупный доход в период t ;
 J – инвестиции в период t ;
 T – налоги в период t ;
 G – государственные расходы в период t .

Вариант 17

Эконометрическая модель имеет вид

$$C_t = a + b \cdot Y_t + \varepsilon;$$

$$Y_t = C_t + I_t + G_t,$$

где C – расходы на потребление;
 Y – доход;
 I – инвестиции;
 G – государственные расходы;
 t – текущий период.

Вариант 18

Модель спроса и предложения на деньги:

$$R_t = a_1 + b_{11} \cdot M_t + b_{12} \cdot Y_t + \varepsilon_1;$$

$$Y_t = a_2 + b_{21} \cdot R_t + \varepsilon_2,$$

где R – процентные ставки в период t ;
 Y – ВВП в период t ;
 M – денежная масса в период t ;
 t – текущий период.

Вариант 19

Модель денежного рынка:

$$R_t = a_1 + b_{11} \cdot M_t + b_{12} \cdot Y_t + \varepsilon_1;$$

$$Y_t = a_2 + b_{21} \cdot R_t + b_{22} \cdot I_t + \varepsilon_2;$$

$$I_t = a_3 + b_{31} \cdot R_t + \varepsilon_3,$$

где R – процентные ставки;
 Y – ВВП;
 M – денежная масса;

I – внутренние инвестиции;
 t – текущий период.

Вариант 20

Имеется следующая модель кейнсианского типа:

$$C_t = a_1 + b_{11} \cdot Y_t + b_{12} \cdot T_{t-1} + \varepsilon_1 \quad (\text{функция потребления});$$

$$I_t = a_2 + b_{21} \cdot Y_{t-1} + \varepsilon_2 \quad (\text{функция инвестиций});$$

$$T_t = a_3 + b_{31} \cdot Y_t + \varepsilon_3 \quad (\text{функция денежного рынка});$$

$$Y_t = C_t + I_t + G_t \quad (\text{тождество дохода}),$$

где C – совокупное потребление в период t ;
 Y – совокупный доход в период t ;
 I – инвестиции в период времени t ;
 T – налоги в период времени t ;
 G – государственные расходы в период времени t ;
 Y_{t-1} – совокупный доход в период $t-1$;
 t – текущий период;
 $t-1$ – предыдущий период.

В этой модели переменные C , I , T и Y являются эндогенными.

4. Изучение взаимосвязей по временным рядам

4.1. Специфика статистической оценки взаимосвязи двух временных рядов

Модели, построенные по данным, характеризующим один объект за ряд последовательных моментов (периодов), называются **моделями временных рядов**.

Временной ряд – это совокупность значений какого-либо показателя за несколько последовательных моментов или периодов.

Применение традиционных методов корреляционно-регрессионного анализа для изучения причинно-следственных зависимостей переменных, представленных в форме временных рядов, может привести к ряду серьезных проблем, возникающих как на этапе построения, так и на этапе анализа эконометрических моделей. В первую очередь эти проблемы связаны со спецификой временных рядов как источника данных в эконометрическом моделировании. Предполагается, что в общем случае каждый уровень временного ряда содержит три основные компоненты: тенденцию (Т), циклические или сезонные колебания (S) и случайную компоненту (Е).

Если временные ряды содержат сезонные или циклические колебания, то перед проведением дальнейшего исследования взаимосвязи необходимо устранить сезонную или циклическую компоненту из уровней каждого ряда, поскольку ее наличие приведет к завышению истинных показателей силы и связи изучаемых временных рядов в случае, если оба ряда содержат циклические колебания одинаковой периодичности, либо к занижению этих показателей в случае, если сезонные или циклические колебания содержит только один из рядов или периодичность колебаний в рассматриваемых временных рядах различна.

Устранение сезонной компоненты из уровней временных рядов можно проводить в соответствии с методикой построения аддитивной и мультипликативной моделей.

Если рассматриваемые временные ряды имеют тенденцию, коэффициент корреляции по абсолютной величине будет высоким, что в данном случае есть результат того, что x и y зависят от времени, или содержат тенденцию. Для того чтобы получить коэффициенты корреляции, характеризующие причинно-следственную связь между изучаемыми рядами, следует избавиться от так называемой ложной корреляции, вызванной наличием тенденции в каждом ряде.

Влияние фактора времени будет выражено в корреляционной зависимости между значениями остатков ε_t за текущий и предыдущие моменты времени, которая получила название **«автокорреляция в остатках»**.

Автокорреляция в остатках есть нарушение одной из основных предпосылок МНК – предпосылки о случайности остатков, полученных по уравнению регрессии. Один из возможных путей решения этой проблемы состоит в применении к оценке параметров модели обобщенного МНК. При построении уравнения множественной регрессии по временным рядам данных, помимо двух вышеназванных проблем, возникает также проблема мультиколлинеарности

факторов, входящих в уравнение регрессии, в случае если эти факторы содержат тенденцию.

4.2. Методы исключения тенденции

Сущность всех методов исключения тенденции заключается в том, чтобы устранить или зафиксировать воздействие фактора времени на формирование уровней ряда. Основные методы исключения тенденции можно разделить на две группы:

- методы, основанные на **преобразовании уровней исходного ряда** в новые переменные, не содержащие тенденции. Полученные переменные используются далее для анализа взаимосвязи изучаемых временных рядов. Эти методы предполагают непосредственное устранение трендовой компоненты T из каждого уровня временного ряда. Два основных метода в данной группе – это **метод последовательных разностей** и **метод отклонений от трендов**;

- методы, основанные на изучении взаимосвязи исходных уровней временных рядов при **элиминировании воздействия фактора времени** на зависимую и независимые переменные модели. В первую очередь, это метод включения в модель регрессии по временным рядам фактора времени. Рассмотрим методику применения, преимущества и недостатки каждого из перечисленных выше методов.

4.2.1 Метод отклонений от тренда

Пусть имеются два временных ряда x_t и y_t , каждый из которых содержит трендовую компоненту T и случайную компоненту ε . Проведение аналитического выравнивания по каждому из этих рядов позволяет найти параметры соответствующих уравнений трендов и определить расчетные по тренду уровни \hat{x}_t и \hat{y}_t соответственно. Эти расчетные значения можно принять за оценку трендовой компоненты T каждого ряда. Поэтому влияние тенденции можно устранить путем вычитания расчетных значений уровней ряда из фактических. Эту процедуру проделывают для каждого временного ряда в модели. Дальнейший анализ взаимосвязи рядов проводят с использованием не исходных уровней, а отклонений от тренда $x_t - \hat{x}_t$ и $y_t - \hat{y}_t$ при условии, что последние не содержат тенденции. Т. е. уравнение регрессии строится в виде

$$y_t - \hat{y}_t = a + b \cdot (x_t - \hat{x}_t). \quad (4.1)$$

4.2.2. Метод последовательных разностей

Если временной ряд содержит ярко выраженную полиномиальную тенденцию (в виде полинома от t), то с целью устранения тенденции можно применить метод последовательных разностей, заключающийся в замене исходных уровней ряда последовательными разностями соответствующих порядков (порядок разности равен порядку полинома).

Если временной ряд содержит ярко выраженную линейную тенденцию, ее можно устранить путем замены исходных уровней ряда цепными абсолютными приростами – первыми последовательными разностями

$$\Delta y_t = y_t - y_{t-1}.$$

Если временной ряд содержит тенденцию в форме параболы второго порядка, то для ее устранения можно заменить исходные уровни ряда на вторые разности

$$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}.$$

Если тенденции временного ряда соответствует экспоненциальный или степенной тренд, метод последовательных разностей следует применять не к исходным уровням ряда, а к их логарифмам.

При всей своей простоте метод последовательных разностей имеет два существенных недостатка. Во-первых, его применение связано с сокращением числа пар наблюдений, по которым строится уравнение регрессии, и, следовательно, с потерей числа степеней свободы. Во-вторых, использование вместо исходных уровней временных рядов их приростов или ускорений приводит к потере информации, содержащейся в исходных данных.

4.2.3. Включение в модель регрессии фактора времени

В корреляционно-регрессионном анализе устранить воздействие какого-либо фактора можно, если зафиксировать воздействие этого фактора на результат и другие включенные в модель факторы. Этот прием широко используется в анализе временных рядов, когда тенденция фиксируется через включение фактора времени в модель в качестве независимой переменной.

Модель вида

$$y_t = a + b_1 \cdot x_t + b_2 \cdot t + \varepsilon_t, \quad (4.2)$$

относится к группе моделей, включающих фактор времени. Очевидно, что число независимых переменных в такой модели может быть больше единицы. Кроме того, это могут быть не только текущие, но и лаговые значения независимой переменной, а также лаговые значения результативной переменной.

Преимущество данной модели по сравнению с методами отклонений от трендов и последовательных разностей в том, что она позволяет учесть всю информацию, содержащуюся в исходных данных, поскольку значения y_t и x_t есть уровни исходных временных рядов. Кроме того, модель строится по всей совокупности данных за рассматриваемый период в отличие от метода последовательных разностей, который приводит к потере числа наблюдений. Параметры a и b модели с включением фактора времени определяются обычным МНК.

Интерпретация параметров уравнения регрессии:

– параметр b_1 показывает, насколько в среднем изменится значение результативного признака y_t при увеличении фактора x_t на единицу при неизменной величине других факторов.

– параметр b_2 показывает, насколько в среднем за год изменится значение результативного признака y_t за счет воздействия всех факторов, кроме фактора x_t .

4.3. Автокорреляция в остатках. Критерий Дарбина–Уотсона

Рассмотрим уравнение регрессии вида

$$\hat{y}_t = a + \sum_j^k b_j \cdot x_{jt} + \varepsilon_t, \quad (4.3)$$

где k – число независимых переменных модели.

Для каждого момента (периода) времени $t = 1, \dots, n$ значение компоненты ε определяется из соотношения

$$\varepsilon_t = y_t - \hat{y}_t = y_t - (a + \sum_j^k b_j \cdot x_{jt}). \quad (4.4)$$

Рассматривая последовательность остатков как временной ряд, можно построить график их зависимости от времени. В соответствии с предпосылками МНК остатки ε_t должны быть случайными. Однако при моделировании временных рядов нередко встречается ситуация, когда остатки содержат тенденцию или циклические колебания. Что свидетельствует о том, что каждое следующее значение остатков зависит от предшествующих. В этом случае говорят о наличии автокорреляции остатков.

Автокорреляция остатков может быть вызвана несколькими причинами, имеющими различную природу:

- 1) наличие ошибок измерения в значениях результативного признака;
- 2) модель может не включать фактор, оказывающий существенное воздействие на результат, влияние которого отражается в остатках, вследствие чего последние могут оказаться автокоррелированными. Очень часто этим фактором является фактор времени t . Кроме того, в качестве таких существенных факторов могут выступать лаговые значения переменных, включенных в модель;
- 3) модель не учитывает несколько второстепенных факторов, совместное влияние которых на результат существенно ввиду совпадения тенденций их изменения или фаз циклических колебаний;
- 4) неправильная спецификация функциональной формы модели. В этом случае следует изменить форму связи факторных и результативного признаков, а не использовать специальные методы расчета параметров уравнения регрессии при наличии автокорреляции остатков.

Существуют два наиболее распространенных метода определения автокорреляции остатков.

Первый метод — это построение графика зависимости остатков от времени и визуальное определение наличия или отсутствия автокорреляции.

Второй метод – использование критерия Дарбина — Уотсона и расчет величины

$$d = \frac{\sum_{i=2}^n (\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^n \varepsilon_i^2}. \quad (4.5)$$

Согласно (4.5) величина d есть отношение суммы квадратов разностей последовательных значений остатков к остаточной сумме квадратов по модели регрессии. Практически во всех статистических ППП значение критерия Дарбина – Уотсона указывается наряду с коэффициентом детерминации, значениями t - и F -критериев.

Коэффициент автокорреляции остатков первого порядка определяется как

$$r_1^\varepsilon = \frac{\sum_{i=2}^n (\varepsilon_i - \bar{\varepsilon}_1)(\varepsilon_{i-1} - \bar{\varepsilon}_2)}{\sqrt{\sum_{i=2}^n (\varepsilon_i - \bar{\varepsilon}_1)^2 \cdot \sum_{i=2}^n (\varepsilon_{i-1} - \bar{\varepsilon}_2)^2}}, \quad (4.6)$$

где

$$\bar{\varepsilon}_1 = \frac{\sum_{i=2}^n \varepsilon_i}{n-1}; \quad \bar{\varepsilon}_2 = \frac{\sum_{i=2}^n \varepsilon_{i-1}}{n-1}.$$

Между критерием Дарбина–Уотсона и коэффициентом автокорреляции остатков первого порядка имеет место следующее соотношение:

$$d \approx 2 \cdot (1 - r_1^\varepsilon).$$

Таким образом, если в остатках существует полная положительная автокорреляция и $r_1^\varepsilon = 1$, то $d = 0$. Если в остатках полная отрицательная автокорреляция, то $r_1^\varepsilon = -1$ и, следовательно, $d = 4$. Если автокорреляция остатков отсутствует, то $r_1^\varepsilon = 0$ и $d = 2$. Следовательно,

$$0 \leq d \leq 4.$$

Алгоритм выявления автокорреляции остатков на основе критерия Дарбина–Уотсона следующий. Выдвигается гипотеза H_0 об отсутствии автокорреляции остатков. Альтернативные гипотезы H_1 и H_1^* состоят, соответственно, в наличии положительной или отрицательной автокорреляции в остатках. Далее по специальным таблицам (см. приложение) определяются критические значения критерия Дарбина–Уотсона d_L и d_U для заданного числа наблюдений n , числа независимых переменных модели k и уровня значимости α . По этим значениям числовой промежуток $[0;4]$ разбивают на пять отрезков. Принятие или отклонение каждой из гипотез с вероятностью $(1-\alpha)$ рассматривается на рис. 4.1.

Есть положительная автокорреляция остатков. H_0 отклоняется. С вероятностью $P = (1-\alpha)$ принимается H_1 .	Зона неопределенности	Нет оснований отклонять H_0 (автокорреляция остатков отсутствует)	Зона неопределенности	Есть отрицательная автокорреляция остатков. H_0 отклоняется. С вероятностью $P = (1-\alpha)$ принимается H_1^* .
0	d_L	d_U	2	$4-d_U$
4				$4-d_L$

Рис. 4.1. Механизм проверки гипотезы о наличии автокорреляции остатков

Если фактическое значение критерия Дарбина – Уотсона попадает в зону неопределенности, то на практике предполагают существование автокорреляции остатков и отклоняют гипотезу H_0 .

Пример 4.1. Проверка гипотезы о наличии автокорреляции в остатках.

Исходные данные, значения ε_t и результаты промежуточных расчетов представлены в табл. 4.1.

Таблица 4.1.

Расчет критерия Дарбина–Уотсона для модели зависимости потребления от дохода

t	y_t	x_t	\hat{y}_t	$\varepsilon_t = y_t - \hat{y}_t$	$\varepsilon_t - \varepsilon_{t-1}$	$(\varepsilon_t - \varepsilon_{t-1})^2$	ε_t^2
1	-	-	-	-	-	-	-
2	1	2	1,54	-0,54	-	-	0,2916
3	0	2	1,54	-0,54	-	-	0,2916
4	2	1	1,11	0,89	1,14	1,2996	0,7921
5	1	2	1,54	-0,54	-1,43	2,0449	0,2916
6	1	1	1,11	-0,11	0,43	0,1849	0,0121
7	2	2	1,54	0,46	0,57	0,3249	0,2116
8	2	3	1,97	0,03	-0,43	0,1849	0,0009
Сумма	9	10	9,06	-0,06*	0,57	4,1233	1,6624

*Сумма не равна нулю ввиду наличия ошибок округления.

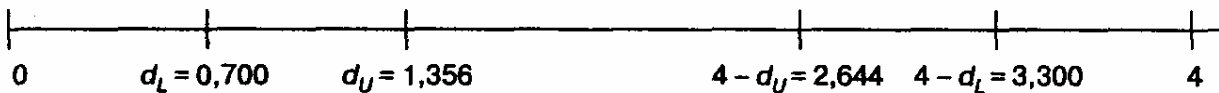
Фактическое значение критерия Дарбина–Уотсона для этой модели составляет $d = 4,1233/1,6624 = 2,48$. Сформулируем гипотезы:

H_0 – в остатках нет автокорреляции;

H_1 – в остатках есть положительная автокорреляция;

H_1^* – в остатках есть отрицательная автокорреляция.

Зададим уровень значимости $\alpha = 0,05$. По таблицам значений критерия Дарбина–Уотсона определим для числа наблюдений $n = 7$ и числа независимых переменных модели $k' = 1$ критические значения $d_L = 0,700$ и $d_U = 1,356$. Получим следующие промежутки внутри интервала $[0;4]$ (рис. 4.2):

Рис. 4.2. Промежутки внутри интервала $[0; 4]$

Фактическое значение $d = 2,48$ попадает в промежуток от d_U до $4 - d_U$. Следовательно, нет оснований отклонять гипотезу H_0 об отсутствии автокорреляции в остатках.

Есть несколько существенных ограничений на применение критерия Дарбина–Уотсона.

Во-первых, он неприменим к моделям, включающим в качестве независимых переменных лаговые значения результативного признака, т. е. к моделям авторегрессии.

Во-вторых, методика расчета и использования критерия Дарбина – Уотсона направлена только на выявление автокорреляции остатков первого порядка. При проверке остатков на автокорреляцию более высоких порядков следует применять другие методы, рассмотрение которых выходит за рамки данного учебника.

В-третьих, критерий Дарбина–Уотсона дает достоверные результаты только для больших выборок.

4.4. Оценивание параметров уравнения регрессии при наличии автокорреляции в остатках

Обратимся к уравнению регрессии

$$y_t = a + b \cdot x_t + \varepsilon_t. \quad (4.7)$$

Примем некоторые допущения относительно этого уравнения:

- пусть y_t и x_t не содержат тенденции, например, представляют собой отклонения выравненных по трендам значений от исходных уровней временных рядов;
- пусть оценки a и b параметров уравнения регрессии найдены обычным МНК;
- пусть критерий Дарбина – Уотсона показал наличие автокорреляции в остатках первого порядка.

Основной подход к оценке параметров модели регрессии в случае, когда имеет место автокорреляция остатков, заключается в следующем: исходная модель регрессии (6.1) с помощью замены переменных приводится к виду

$$y'_t = a' + b \cdot x'_t + u_t, \quad (4.8)$$

где

$$\begin{aligned} y'_t &= y_t - r_1^\varepsilon \cdot y_{t-1}; & x'_t &= x_t - r_1^\varepsilon \cdot x_{t-1}; \\ u_t &= \varepsilon_t - r_1^\varepsilon \cdot \varepsilon_{t-1}; & a' &= a(1 - r_1^\varepsilon). \end{aligned} \quad (4.9)$$

Здесь r_1^ε – коэффициент автокорреляции первого порядка.

Поскольку u_t – случайная ошибка, то для оценки параметров преобразованного уравнения можно применять обычный МНК.

Итак, если остатки по исходному уравнению регрессии содержат автокорреляцию, то для оценки параметров уравнения используют обобщенный МНК. Его реализация разбивается на следующие этапы:

1. Перейти от исходных переменных y_t и x_t к переменным y'_t и x'_t по формулам (4.9).
2. Применив обычный МНК к уравнению (4.8), определить оценки параметров a' и b .
3. Рассчитать параметр a исходного уравнения из соотношения (4.9) как

$$a = a' / (1 - r_1^\varepsilon). \quad (4.10)$$

4. Выписать исходное уравнение (4.7).

Обобщенный метод наименьших квадратов аналогичен методу последовательных разностей. Однако мы вычитаем из y_t (или x_t) не все значение преды-

дущего уровня y_{t-1} (или x_{t-1}), а некоторую его долю $r^{\varepsilon}_1 \cdot y_{t-1}$ (или $r^{\varepsilon}_1 \cdot x_{t-1}$). Если $r^{\varepsilon}_1 = 1$, то данный метод есть просто метод первых разностей, так как

$$\begin{aligned} y'_t &= y_t - y_{t-1}; \\ x'_t &= x_t - x_{t-1}. \end{aligned}$$

Поэтому в случае, если значение критерия Дарбина – Уотсона близко к нулю, применение метода первых разностей вполне обоснованно.

Основная проблема, связанная с применением данного метода, заключается в том, как получить оценку r^{ε}_1 . Основными способами являются оценка этого коэффициента непосредственно по остаткам, полученным по исходному уравнению регрессии, и получение его приближенного значения из соотношения между коэффициентом автокорреляции остатков первого порядка и критерием Дарбина–Уотсона

$$r^{\varepsilon}_1 = 1 - d/2.$$

4.5. Динамические эконометрические модели

Эконометрическая модель называется **динамической**, если в данный момент времени t она учитывает значения входящих в нее переменных, относящиеся как к текущему, так и к предыдущим моментам времени.

При исследовании экономических процессов нередко приходится моделировать ситуации, когда значение результативного признака в текущий момент времени t формируется под воздействием ряда факторов, действовавших в прошлые моменты времени $t - 1, t - 2, \dots, t - l$. Например, на выручку от реализации или прибыль компании текущего периода могут оказывать влияние расходы на рекламу или проведение маркетинговых исследований, сделанные компанией в предшествующие моменты времени.

Величину l , характеризующую запаздывание в воздействии фактора на результат, называют в эконометрике **лагом**, а временные ряды самих факторных переменных, сдвинутые на один или более моментов времени, – **лаговыми переменными**.

Эконометрическое моделирование охарактеризованных выше процессов осуществляется с применением моделей, содержащих не только текущие, но и лаговые значения факторных переменных. Эти модели **называются моделями с распределенным лагом**. Модель вида

$$y_t = a + b_0 \cdot x_t + b_1 \cdot x_{t-1} + b_2 \cdot x_{t-2} + \varepsilon_t \quad (4.11)$$

является примером модели с распределенным лагом.

Решение ряда задач макроэкономики требует ответа на вопрос – какое воздействие окажут значения управляемых переменных текущего периода на будущие значения экономических показателей. Например, как повлияют инвестиции в промышленность на валовую добавленную стоимость этой отрасли экономики будущих периодов? Т. е. исследуются ситуации, когда на величину зависимой переменной текущего периода могут оказывать влияние ее значения в прошлые моменты или периоды времени. Эти процессы обычно описывают с помощью моделей регрессии, содержащих в качестве факторов лаговые значе-

ния зависимой переменной, которые называются **моделями авторегрессии**. Модель вида

$$y_t = a + b_0 \cdot x_t + c_1 \cdot y_{t-1} + \varepsilon_t \quad (4.12)$$

относится к моделям авторегрессии.

Таким образом, выделяют два основных типа динамических эконометрических моделей:

- модели авторегрессии;
- модели с распределенным лагом, в которых значения факторной переменной за прошлые периоды времени (лаговые переменные) непосредственно включены в модель.

Построение моделей с распределенным лагом и моделей авторегрессии имеет свою специфику. Во-первых, оценка параметров моделей авторегрессии, а в большинстве случаев и моделей с распределенным лагом не может быть произведена с помощью обычного МНК ввиду нарушения его предпосылок и требует специальных статистических методов. Во-вторых, исследователям приходится решать проблемы выбора оптимальной величины лага и определения его структуры. Наконец, в-третьих, между моделями с распределенным лагом и моделями авторегрессии существует определенная взаимосвязь, и в некоторых случаях необходимо осуществлять переход от одного типа моделей к другому.

4.6. Интерпретация параметров моделей с распределенным лагом

Рассмотрим модель с распределенным лагом в ее общем виде в предположении, что максимальная величина лага конечна и равна p

$$y_t = a + b_0 \cdot x_t + b_1 \cdot x_{t-1} + \dots + b_p \cdot x_{t-p} + \varepsilon_t. \quad (4.13)$$

Эта модель говорит о том, что если в некоторый момент времени t происходит изменение независимой переменной x , то это изменение будет влиять на значения переменной y в течение t следующих моментов времени.

Коэффициент регрессии b_0 при переменной x_t характеризует среднее абсолютное изменение y_t при изменении x_t на одну единицу своего измерения в некоторый фиксированный момент времени t , без учета воздействия лаговых значений фактора x . Этот коэффициент называют краткосрочным мультипликатором.

В момент $(t + 1)$ совокупное воздействие факторной переменной x_t на результат y_t составит $(b_0 + b_1)$ усл. ед., в момент $(t + 2)$ это воздействие можно охарактеризовать суммой $(b_0 + b_1 + b_2)$ и т. д. Полученные таким образом суммы называют промежуточными мультипликаторами.

С учетом конечной величины лага можно сказать, что изменение переменной x_t в момент t на 1 усл. ед. приведет к общему изменению результата через l моментов времени на $(b_0 + b_1 + \dots + b_l)$ абсолютных единиц.

Введем следующее обозначение:

$$b = b_0 + b_1 + \dots + b_l. \quad (4.14)$$

Величину b называют **долгосрочным мультипликатором**. Он показывает абсолютное изменение в долгосрочном периоде $t + l$ результата под влиянием изменения на 1 ед. фактора x .

Применение обычного МНК к таким моделям в большинстве случаев затруднительно по следующим причинам.

Во-первых, текущие и лаговые значения независимой переменной, как правило, тесно связаны друг с другом. Тем самым оценка параметров модели проводится в условиях высокой мультиколлинеарности факторов.

Во-вторых, при большой величине лага снижается число наблюдений, по которому строится модель, и увеличивается число ее факторных признаков. Это ведет к потере числа степеней свободы в модели.

В-третьих, в моделях с распределенным лагом часто возникнет проблема автокорреляции остатков. Вышеуказанные обстоятельства приводят к значительной неопределенности относительно оценок параметров модели, снижению их точности и получению неэффективных оценок. Чистое влияние факторов на результат в таких условиях выявить невозможно. Поэтому на практике параметры моделей с распределенным лагом проводят в предположении определенных ограничений на коэффициенты регрессии и в условиях выбранной структуры лага.

Рассмотрим теперь следующую модель авторегрессии:

$$y_t = a + b_0 \cdot x_t + c_1 \cdot y_{t-1} + \varepsilon_t.$$

Как и в модели с распределенным лагом, b_0 в этой модели характеризует краткосрочное изменение y_t под воздействием изменения x_t на 1 ед. Однако промежуточные и долгосрочный мультипликаторы в моделях авторегрессии несколько иные. К моменту времени $(t + 1)$ результат y_t изменился под воздействием изменения изучаемого фактора в момент времени t на b_0 ед., а y_{t+1} под воздействием своего изменения в непосредственно предшествующий момент времени – на c_1 ед. Таким образом, общее абсолютное изменение результата в момент $(t + 1)$ составит $b_0 \cdot c_1$ ед. Аналогично в момент времени $(t + 2)$ абсолютное изменение результата составит $b_0 \cdot c_1^2$ ед. и т. д. Следовательно, долгосрочный мультипликатор в модели авторегрессии можно рассчитать как сумму краткосрочного и промежуточных мультипликаторов

$$b = b_0 + b_0 c_1 + b_0 c_1^2 + b_0 c_1^3 \dots$$

С учетом предположения $|c_1| < 1$ (называемое условие стабильности) последнее соотношение преобразуется к виду

$$b = b_0 \cdot (1 + c_1 + c_1^2 + c_1^3 \dots) = \frac{b_0}{1 - c_1}, \quad (4.15)$$

где $|c_1| < 1$.

Отметим, что такая интерпретация коэффициентов модели авторегрессии и расчет долгосрочного мультипликатора основаны на предпосылке о наличии бесконечного лага в воздействии текущего значения зависимой переменной на ее будущие значения.

4.7. Оценка параметров моделей авторегрессии

При построении моделей авторегрессии возникают две серьезные проблемы.

Первая проблема связана с выбором метода оценки параметров уравнения авторегрессии. Наличие лаговых значений результативного признака в правой части уравнения приводит к нарушению предпосылки МНК о делении переменных на результативную (стохастическую) и факторные (нестохастические).

Вторая проблема состоит в том, что поскольку в модели авторегрессии в явном виде постулируется зависимость между текущими значениями результата y_t и текущими значениями остатков u_t , очевидно, что между временными рядами y_{t-1} и u_{t-1} также существует взаимозависимость. Тем самым нарушается еще одна предпосылка МНК, а именно, предпосылка об отсутствии связи между факторным признаком и остатками в уравнении регрессии. Поэтому применение обычного МНК для оценки параметров уравнения авторегрессии приводит к получению смещенной оценки параметра при переменной y_{t-1} .

Одним из методов расчета параметров уравнения авторегрессии является **метод инструментальных переменных**. Сущность этого метода состоит в том, чтобы заменить переменную y_{t-1} из правой части модели, для которой нарушаются предпосылки МНК, на новую переменную \hat{y}_{t-1} , включение которой в модель регрессии не приводит к нарушению его предпосылок.

Искомая новая переменная, которая будет введена в модель вместо y_{t-1} , должна иметь два свойства. Во-первых, она должна тесно коррелировать с y_{t-1} , во-вторых, она не должна коррелировать с остатками u_t .

Существует несколько способов получения такой инструментальной переменной.

1 способ. Поскольку в модели (4.12) переменная y_t зависит не только от y_{t-1} , но и от x_t , можно предположить, что имеет место зависимость y_{t-1} от x_{t-1} , т. е.

$$y_{t-1} = d_0 + d_1 \cdot x_{t-1} + u_t. \quad (4.16)$$

Таким образом, переменную y_{t-1} можно выразить следующим образом:

$$y_{t-1} = \hat{y}_{t-1} + u_t,$$

где

$$\hat{y}_{t-1} = d_0 + d_1 \cdot x_{t-1}. \quad (4.17)$$

Найденная с помощью уравнения (4.17) (его параметры можно искать обычным МНК) оценка \hat{y}_{t-1} может служить в качестве инструментальной переменной для фактора y_t . Эта переменная, во-первых, тесно коррелирует с y_{t-1} , во-вторых, как показывает соотношение (4.17), она представляет собой линейную комбинацию переменной x_{t-1} , для которой не нарушается предпосылка МНК об отсутствии зависимости между факторным признаком и остатками в модели регрессии. Следовательно, переменная \hat{y}_{t-1} также не будет коррелировать с ошибкой u_t .

Таким образом, оценки параметров уравнения (4.12) можно найти из соотношения

$$y_t = a + b_0 * x_t + c_1 * y_{t-1} + v_t, \quad (4.18)$$

предварительно определив по уравнению (4.17) расчетные значения \hat{y}_{t-1} .

2 способ. Подставим в модель (4.12) вместо y_{t-1} его выражение из уравнения (4.16)

$$y_t = a + b_0 \cdot x_t + c_1 \cdot (d_0 + d_1 \cdot x_{t-1} + u_t) + \varepsilon_t.$$

Получим следующую модель:

$$y_t = (a + c_1 \cdot d_0) + b_0 \cdot x_t + c_1 \cdot d_1 \cdot x_{t-1} + (c_1 \cdot u_t + \varepsilon_t). \quad (4.19)$$

Уравнение (4.19) представляет собой модель с распределенным лагом, для которой не нарушаются предпосылки обычного МНК, приводящие к несостоятельности и смещенности оценок параметров. Определив параметры моделей (4.16) и (4.19), можно рассчитать параметры исходной модели (4.12) a , b_0 и c_1 . Модель (4.19) демонстрирует еще одно важное свойство изложенного выше метода инструментальных переменных для оценки параметров моделей авторегрессии: этот метод приводит к замене модели авторегрессии на модель с распределенным лагом.

Отметим, что практическая реализация метода инструментальных переменных осложняется появлением проблемы мультиколлинеарности факторов в модели (4.18): функциональная связь между переменными \hat{y}_{t-1} и x_{t-1} приводит к появлению высокой корреляционной связи между переменными \hat{y}_{t-1} и x_t . В некоторых случаях эту проблему можно решить включением в модель (4.18) и соответственно в модель (4.12) фактора времени в качестве независимой переменной.

Для проверки гипотезы об автокорреляции остатков в моделях авторегрессии Дарбин предложил использовать критерий, который называется критерием h Дарбина. Его расчет производится по следующей формуле:

$$h = \left(1 - \frac{d}{2}\right) \cdot \sqrt{\frac{n}{1 - n \cdot V}},$$

где d – фактическое значение критерия Дарбина–Уотсона для модели авторегрессии;

n – число наблюдений в модели;

V – квадрат стандартной ошибки при лаговой результативной переменной.

Распределение этой величины приблизительно можно аппроксимировать стандартизованным нормальным распределением. Поэтому для проверки гипотезы о наличии автокорреляции остатков можно либо сравнивать полученное фактическое значение критерия h с табличным, воспользовавшись таблицами стандартизованного нормального распределения, либо действовать в соответствии со следующим правилом принятия решения.

1. Если $h > 1,96$, нуль–гипотеза об отсутствии положительной автокорреляции остатков отклоняется.

2. Если $h < -1,96$, нуль–гипотеза об отсутствии отрицательной автокорреляции остатков отклоняется.

3. Если $-1,96 < h < 1,96$, нет оснований отклонять нуль–гипотезу об отсутствии автокорреляции остатков.

Контрольные вопросы:

1. В чем состоит специфика построения моделей регрессии по временным рядам данных?
2. Перечислите основные методы исключения тенденции. Сравните их преимущества и недостатки.
3. Изложите суть метода отклонений от тренда.
4. В чем сущность метода последовательных разностей?
5. Какова интерпретация параметра при факторе времени в моделях регрессии с включением фактора времени?
6. Охарактеризуйте понятие автокорреляции в остатках. Какими причинами может быть вызвана автокорреляция в остатках?
7. Что такое критерий Дарбина – Уотсона? Изложите алгоритм его применения для тестирования модели регрессии на автокорреляцию в остатках.
8. Перечислите основные этапы обобщенного МНК.
9. Приведите примеры экономических задач, эконометрическое моделирование которых требует применения моделей с распределенным лагом и моделей авторегрессии.
10. Какова интерпретация параметров модели с распределенным лагом?
11. Какова интерпретация параметров модели авторегрессии?
12. Изложите методику применения метода инструментальных переменных для оценки параметров модели авторегрессии.
13. Изложите методику тестирования модели авторегрессии на автокорреляцию в остатках.

Лабораторная работа № 6

Задание. На основании данных табл. П2 для соответствующего варианта (табл. 4.2):

1. Построить уравнение авторегрессии.

$$y_t = a + b_0 \cdot x_t + c_1 \cdot y_{t-1} + \varepsilon_t.$$

2. Проверить значимость уравнения регрессии и отдельных коэффициентов.
3. Проверить наличие автокорреляции в остатках.
4. Построить уравнение авторегрессии с учетом фактора времени

$$y_t = a + b_0 \cdot x_t + c_1 \cdot y_{t-1} + c_2 \cdot t + \varepsilon_t.$$

5. Проверить значимость уравнения регрессии и коэффициента при t и оценить целесообразность включения в модель фактора времени.

Указания к решению. Для нахождения уравнений регрессии использовать табличный процессор MS Excel (функция – расчет уравнения регрессии):

Таблица 4.1

Варианты выполнения лабораторных работ № 6

Варианты	Номер графы табл. П2 для результативной переменной	Номер графы табл. П2 для факторной переменной
	y	x
1	3	4
2	3	5
3	3	11
4	3	15
5	10	4
6	10	5
7	10	11
8	10	15
9	16	4
10	16	5
11	16	11
12	16	15
13	7	4
14	7	5
15	7	11
16	7	15
17	14	4
18	14	5
19	14	11
20	14	15
21	6	4
22	6	5
23	6	11
24	6	15
25	11	8

Библиографический список

1. Елисеева, И. И. Эконометрика: учебное пособие /И. И. Елисеева, С. В. Курышева, Д. М. Гордиенко и др. – М.: Финансы и статистика, 2001.
2. Практикум по эконометрике: учебное пособие / под ред. Елисеевой И. И. – М.: Финансы и статистика, 2002.
3. Магнус, Я. Р. Эконометрика. Начальный курс / Я. Р. Магнус, П. К. Катъшев, А. А. Пересецкий. – М.: Дело, 1997. – С. 142 – 163.
4. Айвазян, С. А. Прикладная статистика и основы эконометрики: учебник для вузов / С. А. Айвазян, В. С. Мхитарян. – М.: ЮНИТИ, 1998. – С. 907 – 956.
5. Доугерти, К. Введение в эконометрику. – М.: ИНФРА-М, 1997. – С. 322 – 347.
6. Джонстон, Дж. Эконометрические методы. – М.: Статистика, 1980. – С. 375 – 408.

ПРИЛОЖЕНИЕ

Таблица П1

Исходные данные к лабораторным работам № 1, 2, 3, 4

Наличие предметов длительного пользования в домашних хозяйствах по регионам Российской Федерации (европейская часть территории без республик Северного Кавказа)

(по материалам выборочного обследования бюджетов домашних хозяйств; на 100 домохозяйств; штук)

Области и республики	Телевизоры	Видеомагнитофоны, видеокамеры	Магнитофоны, плееры	Музыкальные центры	Персональные компьютеры	Холодильники. Морозильники	Стиральные машины	Электропылесосы	Швейные, вязальные машины	Легковые автомобили
	1	2	3	4	5	6	7	8	9	10
Белгородская область	113	39	65	10	1	103	93	77	58	26
Брянская область	124	37	47	5	2	99	72	64	50	18
Владимирская область	124	36	61	10	2	105	90	77	68	24
Воронежская область	122	36	44	13	2	102	96	66	60	25
Ивановская область	128	26	47	9	1	106	92	71	71	9
Калужская область	140	43	61	14	6	106	88	81	66	28
Костромская область	117	31	45	12	1	100	85	58	58	14
Курская область	113	40	47	15	3	100	78	66	51	28
Липецкая область	122	48	58	13	2	113	95	73	66	34
Московская область	139	64	57	27	14	106	87	81	62	22
Орловская область	126	39	69	8	6	111	93	73	72	27
Рязанская область	120	34	46	8	3	106	80	65	51	21
Смоленская область	125	39	55	24	6	115	93	66	49	21
Тамбовская область	118	37	59	8	1	108	99	74	65	23
Тверская область	122	35	52	8	4	102	87	64	65	12
Тульская область	133	54	58	15	5	102	93	79	36	16
Ярославская область	136	36	47	12	4	110	88	71	69	14
Республика Карелия	146	49	65	16	9	106	87	68	55	32
Республика Коми	148	58	59	23	5	111	92	78	69	34

Продолжение таблицы П1

	1	2	3	4	5	6	7	8	9	10
Архангельская область	136	35	58	16	8	103	95	74	71	15
Вологодская область	138	34	51	10	3	104	95	64	60	19
Калининградская область	124	48	53	12	7	105	85	74	38	29
Ленинградская область	123	30	65	8	3	102	84	71	52	10
Мурманская область	149	59	63	29	8	107	92	87	74	21
Новгородская область	130	26	63	9	4	96	76	56	45	14
Псковская область	117	26	44	91	3	99	82	65	60	20
Краснодарский край	114	44	60	14	4	109	90	74	67	28
Ставропольский край	114	40	63	12	2	104	91	78	49	29
Астраханская область	126	54	55	11	5	116	87	76	66	29
Волгоградская область	109	41	55	8	1	106	93	74	70	23
Ростовская область	120	43	57	20	8	109	91	73	60	26
Республика Башкортостан	115	40	52	16	4	116	94	75	67	29
Республика Марий Эл	134	28	64	8	1	108	87	72	73	22
Республика Мордовия	130	33	52	10	1	109	89	77	58	18
Республика Татарстан	120	52	61	19	5	119	90	76	55	22
Удмуртская Республика	123	32	53	7	4	111	97	69	74	26
Чувашская Республика	128	31	60	11	0	105	85	76	78	19
Кировская область	144	27	60	7	2	120	109	74	86	22
Нижегородская область	125	36	52	6	3	114	101	81	75	21
Оренбургская область	124	47	57	17	6	119	105	82	66	36
Пензенская область	121	36	47	7	2	109	94	70	69	15
Пермская область	123	40	50	15	6	113	98	73	75	27
Самарская область	128	62	56	24	14	121	100	76	68	35
Саратовская область	118	38	51	12	4	124	87	65	62	27
Ульяновская область	116	37	51	9	4	109	96	77	67	25

Таблица П2

Исходные данные к лабораторной работе № 6

Текущий период	Процентная ставка	ВВП	Денежная масса	Внутренние инвестиции	Национальный доход	Расходы на личное потребление	Валовая прибыль экономики
t	R (%)	Y (млн руб.)	M (млн руб.)	I (млн руб.)	Y (млн руб.)	C* (млн руб.)	Q (млн руб.)
1	2	3	4	5	6	7	8
1	10,01	1 398,5	0,12	211	310	450	725,6
2	6,25	19 005,5	0,95	2 670	5 328	7 500	11 390,5
3	6,00	171 509,5	9,20	27 125	49 730	40 600	76 961,7
4	7,14	610 745,2	33,20	108 810	172 380	124 000	251 944,4
5	8,83	152 404,9	98,70	266 974	437 007	310 000	662 374,4
6	8,27	2 145 655,5	220,80	375 998	558 500	260 000	790 819,2
7	8,44	2 478 594,1	288,30	408 797	711 600	390 000	881 001,1
8	8,35	2 741 051,2	374,10	407 086	686 000	490 000	1 032 768,6
9	7,99	4 757 233,7	448,30	970 439	1 213 600	990 000	2 050 276,8
10	7,83	7 063 392,8	704,70	1 165 181	2 097 700	1 650 000	3 033 247,2

Окончание таблицы П2

Текущий период	Индекс стоимости жизни	Объем продукции промышленности	Государственные расходы	Доля импорта в ВВП	Реальный объем чистого экспорта	Налоги	Запас капитала	Зарплата
t	P (%)	R (млн руб.)	G (млн руб.)	M	X (млн руб.)	T (млн руб.)	K (млн руб.)	S (тыс.руб.)
1	9	10	11	12	13	14	15	16
1	200	600	348	0,1295	186	152	325	0,6
2	210	1 300	5 970	0,4826	1 1847	3 893	4 550	6,0
3	220	18 500	57 674	0,3050	6 5524	28 672	34 965	58,7
4	238	129 000	230 385	0,2321	169 534	85 044	133 209	220,4
5	195	384 000	486 112	0,2429	426 735	253 326	327 941	472,4
6	208	1 108 000	652 700	0,2060	532 239	380 685	454 369	790,2
7	229	1 469 000	839 000	0,2094	592 332	471 657	482 451	950,2
8	204	1 626 000	842 100	0,2350	840 596	520 534	485 452	1051,5
9	180	1 707 000	1258 000	0,2689	2 090 687	875 751	766 672	1522,6
10	181	3 150 000	1960 100	0,2493	3 232 388	1 348 178	1 293 750	2223,4

Удвоенная нормированная функция Лапласа

$$\Phi_{(e)} = \frac{2}{\sqrt{2\pi}} \int_0^e e^{-\frac{t^2}{2}} \cdot dt = \frac{2}{\sqrt{2\pi}} \int_{-t}^e e^{-\frac{t^2}{2}} \cdot dt$$

t	0	1	2	3	4	5	6	7	8	9
0,0	0 000	0 080	0 160	0 239	0 319	0 399	0 478	0 558	0 638	0 717
0,1	0 797	0 876	0 955	1 034	1 113	1 192	1 271	1 350	1 428	1 507
0,2	1 585	1 663	1 741	1 819	1 897	1 974	2 051	2 128	2 205	2 282
0,3	2 358	2 434	2 510	2 586	2 661	2 737	2 812	2 886	2 961	3 035
0,4	3 108	3 182	3 255	3 328	3 401	3 473	3 545	3 616	3 683	3 759
0,5	3 829	3 899	3 969	4 039	4 108	4 177	4 245	4 313	4 381	4 448
0,6	4 515	4 581	4 647	4 713	4 778	4 843	4 907	4 971	5 035	5 098
0,7	5 161	5 223	5 385	5 346	5 407	5 467	5 527	5 587	5 646	5 705
0,8	5 763	5 821	5 878	5 935	5 991	6 047	6 102	6 157	6 211	6 265
0,9	6 319	6 372	6 424	6 176	6 528	6 579	6 629	6 680	6 729	6 778
1,0	6 827	6 875	6 923	6 970	7 017	7 063	7 109	7 154	7 199	7 243
1,1	7 287	7 339	7 373	7 415	7 457	7 499	7 540	7 580	7 620	7 660
1,2	7 699	7 737	7 775	7 813	7 850	7 887	7 923	7 959	7 995	8 029
1,3	8 064	8 098	8 132	8 165	8 198	8 230	8 262	8 293	8 324	8 355
1,4	8 385	8 415	8 444	8 473	8 501	8 529	8 557	8 584	8 611	8 638
1,5	8 664	8 690	8 715	8 740	8 764	8 789	8 812	8 836	8 859	8 882
1,6	8 904	8 926	8 948	8 969	8 990	9 011	9 031	9 051	9 070	9 090
1,7	9 109	9 127	9 146	9 164	9 181	9 199	9 216	9 233	9 249	9 265
1,8	9 281	9 297	9 312	9 327	9 342	9 357	9 371	9 385	9 399	9 412
1,9	9 426	9 439	9 451	9 464	9 476	9 488	9 500	9 512	9 523	9 534
2,0	9 545	9 556	9 566	9 576	9 586	9 596	9 606	9 616	9 625	9 634
2,1	9 643	9 651	9 660	9 668	9 674	9 684	9 692	9 700	9 707	9 715
2,2	9 722	9 729	9 736	9 743	9 749	9 756	9 762	9 768	9 774	9 780
2,3	9 786	9 791	9 797	9 802	9 807	9 812	9 817	9 822	9 827	9 832
2,4	9 836	9 840	9 845	9 849	9 853	9 857	9 861	9 865	9 869	9 872
2,5	9 876	9 879	9 883	9 886	9 889	9 892	9 895	9 898	9 901	9 904
2,6	9 907	9 909	9 912	9 915	9 917	9 920	9 922	9 924	9 926	9 929
2,7	9 931	9 933	9 935	9 937	9 939	9 940	9 942	9 944	9 946	9 947
2,8	9 949	9 950	9 952	9 953	9 955	9 956	9 958	9 959	9 960	9 961
2,9	9 963	9 964	9 965	9 966	9 967	9 968	9 969	9 970	9 971	9 972
3,0	9 973	9 981	9 986	9 990	9 993	9 995	9 997	9 998	9 999	9 999

*Значения ординат увеличены в 10 000 раз

Значения α -процентных пределов $t_{\alpha,k}$ в зависимости от k степеней свободы и заданного уровня значимости α для распределения Стьюдента

$k \backslash \alpha$ (%)	10,0	5,0	2,5	2,0	1,0	0,5	0,3	0,2	0,1
1	6,314	12,706	25,452	31,821	63,657	127,30 0	212,20 0	318,30 0	636,60 0
2	2,920	4,303	6,205	6,965	9,925	14,089	18,216	22,327	31,600
3	2,353	3,182	4,177	4,541	5,841	7,453	8,891	10,214	12,922
4	2,132	2,776	3,495	3,747	4,604	5,597	6,435	7,173	8,610
5	2,015	2,571	3,163	3,365	4,032	4,773	5,376	5,893	6,869
6	1,943	2,447	2,969	3,143	3,707	4,317	4,800	5,208	5,959
7	1,895	2,365	2,841	2,998	3,499	4,029	4,442	4,785	5,408
8	1,860	2,306	2,752	2,696	3,355	3,833	4,199	4,501	5,041
9	1,833	2,262	2,685	2,821	3,250	3,690	4,024	4,297	4,781
10	1,812	2,228	2,634	2,764	3,169	3,581	3,892	4,144	4,587
12	1,782	2,179	2,560	2,681	3,055	3,428	3,706	3,930	4,318
14	1,761	2,145	2,510	2,624	2,977	3,326	3,583	3,787	4,140
16	1,746	2,120	2,473	2,583	2,921	3,252	3,494	3,686	4,015
18	1,734	2,101	2,445	2,552	2,878	3,193	3,428	3,610	3,922
20	1,725	2,086	2,423	2,528	2,845	3,153	3,376	3,552	3,849
22	1,717	2,074	2,405	2,508	2,819	3,119	3,335	3,505	3,792
24	1,711	2,064	2,391	2,492	2,797	3,092	3,302	3,467	3,745
26	1,706	2,056	2,379	2,479	2,779	3,067	3,274	3,435	3,704
28	1,701	2,048	2,369	2,467	2,763	3,047	3,250	3,408	3,674
30	1,697	2,042	2,360	2,457	2,750	3,030	3,230	3,386	3,646
∞	1,645	1,960	2,241	2,326	2,576	2,807	2,968	3,090	3,291

**Таблица значения функции F для доверительной вероятности
 $P = (1 - 0,05) = 0,95$**

$K_2 \backslash K_1$	1	2	3	4	5	6	8	12	24
1	161,45	199,50	215,72	224,57	230,17	233,97	238,89	243,91	249,04
2	18,54	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45
3	10,18	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84
7	6,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74

Значения статистик Дарбина – Уотсона $d_L d_U$ при 5%-ном уровне значимости

n	$K^1=1$		$K^1=2$		$K^1=3$		$K^1=4$		$K^1=5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
6	0,61	1,40	—	—	—	—				
7	0,70	1,36	0,47	1,90	—	—				
8	0,76	1,33	0,56	1,78	0,37	2,29				
9	0,82	1,32	0,63	1,70	0,46	2,13				
10	0,88	1,32	0,70	1,64	0,53	2,02				
11	0,93	1,32	0,66	1,60	0,60	1,93				
12	0,97	1,33	0,81	1,58	0,66	1,86				
13	1,01	1,34	0,86	1,56	0,72	1,82				
14	1,05	1,35	0,91	1,55	0,77	1,78				
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83

Оглавление

1. Парная регрессия и корреляция.....	3
1.1. Понятие регрессии	3
1.2. Построение уравнения регрессии.....	3
1.2.1. Спецификация модели.....	4
1.2.2. Оценка параметров модели.....	4
1.3. Оценка тесноты связи	6
1.4. Оценка значимости уравнения регрессии, его коэффициентов, коэффициента детерминации.....	7
1.5. Расчет доверительных интервалов	8
1.6. Точечный и интервальный прогноз по уравнению линейной регрессии	9
1.7. Коэффициент эластичности	10
Лабораторная работа № 1	11
Лабораторная работа № 2.....	11
2. Множественная регрессия и корреляция.....	13
2.1. Общие положения	13
2.2. Отбор факторов при построении множественной регрессии.....	13
2.3. Выбор формы уравнения регрессии.....	17
2.4. Оценка параметров уравнения множественной регрессии.....	19
2.5. Частные уравнения регрессии.....	20
2.6. Множественная корреляция.....	22
2.7. Частная корреляция.....	25
2.8. Оценка надежности результатов	28
множественной регрессии и корреляции.....	28
2.9. Проверка остатков регрессии на гомоскедастичность.....	29
Лабораторная работа № 3.....	30
Лабораторная работа № 4.....	30
Пример выполнения лабораторной работы № 3.....	31
Пример выполнения лабораторной работы № 4.....	34
3. Система эконометрических уравнений.....	39
3.1. Системы эконометрических уравнений.....	39
3.2. Проблема идентификации	40
3.3. Оценивание параметров структурной модели	41
3.4. Решение типовых задач	42
Лабораторная работа № 5.....	49
4. Изучение взаимосвязей по временным рядам.....	58
4.1. Специфика статистической оценки взаимосвязи двух временных рядов	58
4.2. Методы исключения тенденции	59
4.2.1 Метод отклонений от тренда	59

4.2.2. Метод последовательных разностей	59
4.2.3. Включение в модель регрессии фактора времени	60
4.3. Автокорреляция в остатках. Критерий Дарбина–Уотсона	61
4.4. Оценивание параметров уравнения регрессии при наличии автокорреляции в остатках	64
4.5. Динамические эконометрические модели	65
4.6. Интерпретация параметров моделей с распределенным лагом	66
4.7. Оценка параметров моделей авторегрессии	68
Лабораторная работа № 6	70
Библиографический список	71
ПРИЛОЖЕНИЕ	72
Исходные данные к лабораторным работам № 1, 2, 3, 4	72
Исходные данные к лабораторной работе № 6	74
Удвоенная нормированная функция Лапласа	75
Значения α -процентных пределов $t_{\alpha,k}$ в зависимости от k степеней свободы и заданного уровня значимости α для распределения Стьюдента ...	76
Таблица значения функции F для доверительной вероятности $P = (1 - 0,05) = 0,95$	77
Значения статистик Дарбина – Уотсона $d_L d_U$ при 5%-ном уровне значимости	78
Оглавление	79

Учебное издание

Эконометрика. Лабораторный практикум
для студентов специальности 351400

Составитель ШАНЧЕНКО Николай Иванович
Корректор Л. В. Рыжова
Компьютерная верстка Н. И. Шанченко

Подписано в печать 30.12.2004 Формат 60×84/16.

Бумага писчая. Усл. печ. л. 4,56. Уч.–изд.л. 4,10.

Тираж 100 экз. Заказ

Ульяновский государственный технический университет
432027, г. Ульяновск, Северный Венец, 32

Типография УлГТУ, 432027, Ульяновск, Северный Венец, 32.