Учебно-методические материалы предназначены для студентов Института открытого образования Финансовой академии при Правительстве РФ, подготовка которых ведется по заочной форме с использованием элементов дистанционных технологий. Они представляют собой методическое обеспечение самостоятельной подготовки и выполнения контрольных работ по дисциплине «Эконометрика» для студентов, обучающихся по специальностям «Бухгалтерский учет и аудит», «Финансы и кредит».

В математической экономике изучается взаимосвязь между экономическими переменными на неколичественном, общем уровне. Если коэффициенты, представленные в общем виде в этих взаимосвязях заменить конкретными численными значениями, рассчитанными по соответствующим экономическим данным, то получим *эконометрическую модель*.

Отсюда - главное назначение *эконометрики* – модельное описание конкретных количественных взаимосвязей, существующих между экономическими показателями.

Эконометрические модели представляют собой либо отдельные алгебраические уравнения, либо системы алгебраических уравнений (чаще всего линейных) и, возможно, неравенств относительно коэффициентов этих уравнений. В эти уравнения, как правило, включают *остаточную случайную (случайный остаток, случайное возмущение)* составляющую, отражающую влияние на результирующий показатель всех неучтенных факторов.

1. Спецификация эконометрической модели

При составлении *спецификации* (т.е. подробного математического описания) эконометрической модели все используемые в ней экономические переменные разделяют на:

- **экзогенные** («внешние», их значения известны и задаются вне модели) – объясняющие переменные;
- **эндогенные** («внутренние», их значения неизвестны и определяются внутри модели под воздействием экзогенных переменных) – объясняемые переменные.

В динамических моделях оба вида переменных могут рассматриваться как в текущий момент времени (*текущие переменные*), так и в прошлый (*лаговые переменные*). При этом значения лаговых эндогенных переменных, включенные в модель, также считаются известными, поскольку их значения уже вычислены в пошлом по отношению к рассматриваемому текущему периоду времени. В связи с этим для объяснения эндогенных величин используют **предопределенные** переменные, к которым относятся *текущие экзогенные*, *лаговые экзогенные* и *лаговые эндогенные* переменные.

Замечание 1. При составлении спецификации модели количество уравнений в ней определяется количеством эндогенных переменных.

Замечание 2. Случайные составляющие включаются только в уравнения спецификации модели, в тождества они обычно не входят.

Задача 1. Составить спецификацию упрощенной макромодели для экономического объекта, состояние которого описывается следующими экономическими положениями:

- потребление C возрастает с ростом дохода населения[1] (но оно растет медленнее, чем доход);
- объем инвестиций I повышается с ростом национального дохода Y и падает при росте ставки процента R;
- национальный доход – сумма потребления, инвестиций и государственных закупок товаров и услуг G .

Решение. Для формулировки условия задачи на математическом языке выбираюi наиболее простые из возможных форм соотношений для переменных. Выберем линейную зависимость для анализируемых переменных. Тогда первое утверждение может быть представлено так:

$$C_t = a_0 + a_1(Y_t - T_t) + l_t, \text{ где } 0<a_1<1.$$

Анализируя второе утверждение, понимаем, что текущие инвестиции лучше объясняются лаговым значением национального дохода, поэтому:

$$I_t = b_0 + b_1 Y_{t-1} + b_2 R_t + v_t, \text{ где } b_1>0,\ b_2<0.$$

Для третьего положения будет справедливо тождество:

$$Y_t = C_t + I_t + G_t.$$

Таким образом спецификация модели будет иметь вид:

$$\begin{aligned} &C_t = a_0 + a_1(Y_t - T_t) + l_t, \\ &I_t = b_0 + b_1 Y_{t-1} + b_2 R_t + v_t, \\ &Y_t = C_t + I_t + G_t, \end{aligned} \qquad (1.1)$$

где $0<a_1<1$, $b_1>0$, $b_2<0$.

Отметим, что в нашей спецификации содержится два уравнения, объясняющие поведение потребителей и инвесторов, и одно тождество. Это соответствует количеству рассматриваемых эндогенных переменных модели. Их тоже три: C_t, Y_t, I_t. Объясняются они через набор из четырех предопределенных переменных: T_t, R_t, G_t – текущие экзогенные, Y_{t-1} –лаговая эндогенная переменная. Случайные возмущения l_t, v_t включены только в уравнения модели.

Получение спецификации модели - важнейший этап эконометрического моделирования, поскольку именно от него в решающей степени зависит успех всего

[1] Под *доходом населения* понимается разность между национальным доходом Y и подоходным налогом T

исследования. Правильное построение спецификации зависит от того, насколько реалистичны наши предположения о составе эндогенных и предопределенных переменных, о структуре самой системы уравнений, стохастической природе случайных остатков. При ее составлении исследователь опирается на утверждения экономической теории, использует специальные знания из области математики, статистики, теории вероятности, собственную интуицию.

2. Схема построения эконометрических моделей.

В эконометрическом моделировании можно выделить четыре основных этапа:

I. Построение спецификации эконометрической модели финансово-экономического объекта.
II. Сбор необходимой статистической информации по всем переменным, входящим в спецификацию, в виде конкретных значений показателей факторов, участвующих в модели.
III. Статистическое оценивание неизвестных параметров, входящих в модель.
IV. Верификация модели (проверка адекватности оцененной модели).

Приведенное разделение условно, поскольку эти этапы могут пересекаться и дополнять друг друга.

Пусть на ***первом этапе*** спецификация модели получена нами в виде изолированного уравнения:

$$y_t = a_0 + a_1 x_t + u_t \qquad (2.1)$$

Уравнение (2.1) задает экономическую тенденцию (закон), по которой меняется уровень y_t вслед за изменением величины x_t. В свою очередь, величина u_t в (2.1) описывает воздействие на y_t факторов, которые объективно существуют, но не отражены в модели (2.1). Эти факторы отклоняют реальное значение y_t от ожидаемого. Возникает вопрос: как измерить диапазон отклонения? Или, что равносильно, как оценить разброс возможных значений u_t вокруг нуля? Мерилом такого разброса служит в эконометрике среднее квадратическое отклонение (с.к.о.) $\sigma(u_t|x_t)$. Оно имеет следующий смысл:

$$\sigma(u_t \mid x_t) = \sqrt{\frac{u_1^2 + u_2^2 + \dots + u_n^2}{n-2}} \qquad (2.2)$$

Если $\sigma(u_t|x_t)$ не зависит от x_t, то обозначают с.к.о. просто как σ_u. Следует подчеркнуть, что σ_u является параметром модели (2.1) наряду с коэффициентами (a_0,a_1).

Таким образом, спецификация (2.1) содержит три неизвестных параметра:

$$(a_0, a_1; \sigma_u) \qquad (2.3)$$

Эти параметры надо определить какими-то методами и при помощи какой-то информации об объекте оригинале. Для этого ***на втором этапе*** построения модели **исследователь многократно определяет значения экономических переменных объекта, включенных в модель.** Для модели (2.1) измеряются значения x_t и потребления y_t в ряде периодов времени t=1, 2, …, n. В итоге этих измерений в распоряжении эконометриста оказываются известные значения

$$(x_1,y_1), (x_2,y_2), \ldots, (x_n,y_n). \qquad (2.4)$$

Именно по этой информации будут, с одной стороны, оцениваться неизвестные параметры модели (2.3), а с другой стороны - проверяться адекватность оцененной модели.

Третий этап построения модели заключается в оценивании (приближенном определении) ее неизвестных параметров методами математической статистики, поэтому его именуют этапом оценивания (или настройки) модели. Так, в итоге третьего этапа построения для модели (2.1) вычисляются по статистической информации (2.4) оценки (приближенные значения) ее неизвестных параметров

$$(\widetilde{a}_0, \widetilde{a}_1; \widetilde{\sigma}_u) \qquad (2.5)$$

Оценки коэффициентов $(\widetilde{a}_0, \widetilde{a}_1)$ согласно **методу наименьших квадратов** (МНК) *выбираются таким образом, чтобы сумма квадратов отклонений эмпирических значений* y_i *из (2.4) от значений* $\widetilde{y}_i$, *найденных по уравнению регрессии*

$$\widetilde{y}_i = \widetilde{a}_0 + \widetilde{a}_1 x_i \qquad (2.6)$$

была минимальной:

$$\sum_{i=1}^{n} (\widetilde{y}_i - y_i)^2 = \sum_{i=1}^{n} (\widetilde{a}_0 + \widetilde{a}_1 x_i - y_i)^2 \to \min.$$

Для этого $\widetilde{a}_1$ вычисляется в процессе решения системы нормальных уравнений или линейного уравнения R$\widetilde{a}_1$=*S*:

$$\widetilde{a}_1 = S/R, \qquad (2.7)$$

где

$$R = \sum_{t=1}^{n} x_t^2 = x_1^2 + x_2^2 + \ldots + x_n^2,$$
$$S = \sum_{t=1}^{n} y_t \cdot x_t = y_1 x_1 + y_2 x_2 + \ldots + y_n x_n. \qquad (2.8)$$

Оценку свободного члена $\widetilde{a}_0$ получаем как:

$$\widetilde{a}_0 = \bar{y} - \widetilde{a}_1 \bar{x}, \qquad (2.9)$$

где $\bar{y}$ – среднее значение эндогенной переменной по выборке (2.4),

$\bar{x}$ - среднее значение экзогенной переменной по выборке (2.4).

Одновременно с вычислением $(\widetilde{a}_0, \widetilde{a}_1)$ эконометристы определяют и характеристики точности этих оценок – средние квадратические ошибки. S_{a_0} и S_{a_1} соответственно оценок $\widetilde{a}_0$ и $\widetilde{a}_1$ (они используются при вычислении критических уровней ошибок прогнозных значений эндогенных переменных). Например,

$$S_{a_1} = \widetilde{\sigma}_u \cdot \sqrt{R^{-1}}. \qquad (2.10)$$

Здесь

$$\widetilde{\sigma}_u = \sqrt{\frac{\widetilde{u}_1^2 + \widetilde{u}_2^2 + \ldots + \widetilde{u}_{n^*}^2}{n^* - 1}} \qquad (2.11),$$

при чем здесь n*=n-1 – количество пар значений переменных модели из набора (2.4) по которым находятся оценки $(\widetilde{a}_0, \widetilde{a}_1)$.

$$\widetilde{u}_t = y_t - (\widetilde{a}_0 + \widetilde{a}_1 x_t) \qquad (2.12)$$

для всех t=1, …, n*.

Полученную на третьем этапе оцененную (настроенную) модель записывают так:

$$\left.\begin{array}{l} y_t = \widetilde{a}_0 + \widetilde{a}_1 \cdot x_t + u_t \\ \quad (S_{a_0})\,(S_{a_1})\ \ (\widetilde{\sigma}_u) \end{array}\right\}. \qquad (2.13)$$

Под соответствующими величинами, в скобках, приводятся количественные характеристики разброса этих величин относительно их ожидаемых значений.

Суть проверки модели на адекватность на ***четвертом этапе*** в общих чертах[2] проста. Для ее осуществления выборку значений (2.4) делим на две части:

1) $(y_1,x_1), (y_2,x_2), \ldots, (y_{n-1},x_{n-1})$.
2) (y_n,x_n).

Получили две выборки: первую назовем *обучающей*, вторую - *контрольной*. Пусть модель (2.13) оценена именно по обучающей выборке. При помощи оценочной модели (2.13) осуществим прогноз величины y_n:

[2] Подробно проверка адекватности модели будет рассмотрена отдельно.

$$\widetilde{y}_n = \widetilde{a}_0 + \widetilde{a}_1 \cdot x_n \qquad (2.14)$$

Он, как правило, будет отличаться от реального значения y_n, которое известно. Ошибка прогноза (2.14), то есть величина

$$(\widetilde{y}_n) = \widetilde{y}_n - y_n, \qquad (2.15)$$

в ситуации адекватной модели (2.13) не должна выходить за определенный критический уровень $_{\text{крит}}$:

$$(\widetilde{y}_n) \qquad _{\text{крит}}. \qquad (2.16)$$

При выполнении неравенства (2.16) модель (2.13) признается адекватной. В противном случае – модель неадекватна и эконометрист возвращается на первый этап схемы: меняет спецификацию и вновь проходит третий и четвертый этапы, поскольку лишь адекватная модель может использоваться как средство изучения объекта-оригинала (в частности, для прогноза текущих эндогенных переменных).

Задача 2. Согласно рассмотренной схемы построить спецификацию инвестиционной модели и проверить ее адекватность, если известно, что уровень текущих инвестиций в закрытой экономике прямо пропорционален приросту ВВП в предыдущем периоде.

Решение. Спецификацию модели будем строить для объяснения уровня инвестиций I в текущем периоде t. Тогда I_t – эндогенная переменная, а $\Delta Y_{t-1} = Y_{t-1} - Y_{t-2}$ - предопределенная переменная модели. Учитывая, что инвестиции можно рассматривать как зависимую величину не только от прироста ВВП, но и от ставки процента, уровня инфляции и т.п., включим в спецификацию модели случайный остаток v_t. Тогда спецификацию модели можно представить так:

$$I_t = b \cdot \Delta Y_{t-1} + v_t, \qquad (2.17)$$
$$b > 0$$

Статистические данные по переменным позаимствуем из таблицы 1.

Таблица 1.

t	0	1	2	3	4	5
I_t	955	936	876	942	1052	1186
Y_t	5330	5591	2742	6054	6424	7028
$\Delta Y_{t-1} = Y_{t-1} - Y_{t-2}$	-	261	151	312	370	604

Имеющийся набор соответствующих в рамках модели (2.17) значений переменных разделим на две выборки:

(876, 261), (942, 151), (1052,312) – обучающая выборка;

(1186, 370) – контрольная выборка.

Используя обучающую выборку находим оценки параметров (b; σ_v). Для вычисления $\widetilde{b}$ по правилу (2.7) рассчитываем вначале R=188266 и S=699102. Тогда $\widetilde{b}$=699102/188266=3,71.

Теперь для t=2, 3, 4 по формуле (2.12) определяем оценки значений случайных остатков:

$$\widetilde{v}_2 = I_2 - \widetilde{b}\Delta Y_1 = 876 - 3{,}71 \cdot 261 = -92{,}3$$
$$\widetilde{v}_3 = I_3 - \widetilde{b}\Delta Y_2 = 942 - 3{,}71 \cdot 151 = 381{,}8$$
$$\widetilde{v}_4 = I_4 - \widetilde{b}\Delta Y_3 = 1052 - 3{,}71 \cdot 312 = -105{,}5$$

Подставив их в формулу (2.11) при n*=3 находим значение $\widetilde{\sigma}_v \approx 288$. Средняя квадратическая ошибка прогноза согласно (2.10) $S_b = \widetilde{\sigma}_v \cdot \sqrt{R^{-1}} = 288 \cdot \sqrt{188266^{-1}} \approx 0{,}7$. Оцененную модель (2.17) можно теперь представить так:

$$I_t = 3{,}71 \cdot \Delta Y_{t-1} + v_t \qquad (2.18)$$

$$(S_b=0{,}7) \quad (\widetilde{\sigma}_v = 288.)$$

Полученные результаты с экономической точки зрения можно проинтерпретировать следующим образом: коэффициент $\widetilde{b}$ =3,71 показывает, что каждая единица прироста ВВП в предыдущем периоде (в денежном выражении) индуцирует 3,71 денежную единицу дополнительных текущих инвестиций. При этом оцененное значение коэффициента отличается от истинного неизвестного значения b в среднем квадратическом на 0,7.

Проведем верификацию модели (2.18). Для этого используем значение экзогенной переменной из контрольной выборки $\Delta Y_4 = 370$ и по формуле

$$\widetilde{I}_t = 3{,}71 \cdot \Delta Y_{t-1}$$

вычислим прогнозное значение $\widetilde{I}_5$:

$$\widetilde{I}_5 = 3{,}71*370=1373.$$

Истинная ошибка такого прогноза составляет $(\widetilde{I}_5)=\widetilde{I}_5-I_5$=1373-1186=187 и вполне укладывается в рамки полученного значения $\widetilde{\sigma}_v = 288$, поэтому и без более изощренных расчетов[3] сейчас можно сделать вывод о том, что оцененная модель (2.18) адекватна объекту оригиналу.

3. Использование Excel при оценивании линейных эконометрических моделей

Оценки параметров (b; σ_v) в решении задачи 2 можно получить в Excel, воспользовавшись функцией ЛИНЕЙН.

Общий алгоритм применения этой функции для оценивания моделей линейной множественной регрессии:

1) ввести по столбикам известные значения: по первому столбцу - эндогенной переменной, начиная со второго – предопределенных переменных (Рис. 1);

	A	B	C	D
1	y_1	x_{11}	x_{21}	
2	y_2	x_{12}	x_{22}	
…	…			
n	y_n	x_{1n}	x_{2n}	

Рис. 1.

2) поместив курсор в ячейку с адресом A,n+1 вызвать *Мастер функций* (*Вставка-Функция* или через знак f_x на панели инструментов), из категории *Статистические* выбрать функцию *Линейн*;

3) заполнить открывшееся диалоговое окно функции(Рис. 2):

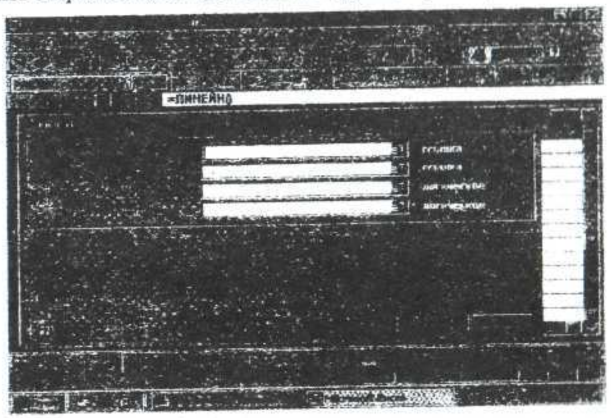

Рис. 2.

[3] Более подробные расчеты приведены в разделе «Проверка адекватности модели»

- в первой строчке указать диапазон ячеек, в которых находятся значения эндогенной переменной – A1:An;
- во второй – диапазон ячеек со значениями объясняющих переменных – B1:Cn;
- в третью строку поставить: 1 (*Истина*), если в линейном уравнении есть свободный член; 0 (*Ложь*), если свободного члена нет;
- в четвертой строке также можно поставить 0 или 1 – в первом случае в качестве выходной информации будут представлены только оценки коэффициентов; поскольку интересна и другая статистическая информация для модели, то лучше всегда заполнять эту строчку 1.

Закрыть диалоговое окно нажатием *Ok*.

4) выделить курсором диапазон ячеек A,n+1:C,n+5;

5) активизировать строку формул, поставив в нее курсор;

6) одновременно нажать комбинацию клавиш Ctrl+Shift+Enter.

В результате *Мастером функций* в выделенном диапазоне будут размещены результаты оценивания модели (Рис. 3). Сейчас обратим внимание только на выделенные значения, поскольку они нам необходимы для записи оцененного вида модели. К информации по другим показателям обратимся позже.

	A	B	C	D
1	y_1	x_{11}	x_{21}	
2	y_2	x_{12}	x_{22}	
…	…			
n	y_n	x_{1n}	x_{2n}	
n+1	$\tilde{a}_2$	$\tilde{a}_1$	$\tilde{a}_0$	
n+2	S_{a_2}	S_{a_1}	S_{a_0}	
n+3	R^2	$\tilde{\sigma}$	#Н/Д	
n+4	F	f_2	#Н/Д	
n+5	RSS	ESS	#Н/Д	

Рис. 3.

Описанная процедура применяется для оценивания в Excel линейных эконометрических моделей с любым количеством объясняющих переменных.

Задача 3. Оценить в Excel модель (2.17), полученную в ходе решения задачи 2.

Решение. По условию задачи в модель входят две переменные:

- эндогенная I_t, ее значения из обучающей выборки внесем по столбцу А,
- предопределенная ΔY_{t-1}, значения которой из обучающей выборки внесем по столбцу В.

Результаты выполнения последующих 2)-6) шагов описанного выше алгоритма показаны на рисунке 4.

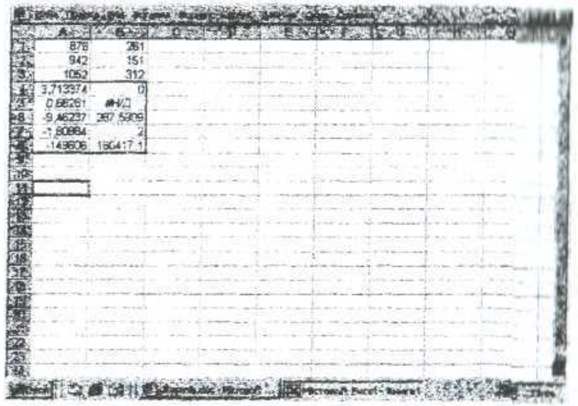

Рис. 4.

Используя полученные оценки параметров, запишем модель (2.17) в оцененном виде:

$$I_t = 3{,}71 \cdot \Delta Y_{t-1} + v_t \qquad (3.1)$$

$$(S_b{=}0{,}66) \quad (\widetilde{\sigma}_v = 287{,}6)$$

Можно убедиться, что полученная таким образом оцененная модель (3.1) совпадает с моделью (2.18) из задачи 2 (с учетом допущенных там округлений чисел). Это свидетельствует о том, что и параметры в (3.1) также рассчитаны на основе метода наименьших квадратов.

4. Основные положения регрессионного анализа. Теорема Гаусса-Маркова.

Методы и модели регрессионного анализа занимают центральное место в математическом аппарате эконометрики.

Рассматриваемая в регрессионном анализе зависимость переменной y от x может задаваться уравнением регрессии:

$$y_t = a_0 + a_1 x_t \qquad (4.1).$$

В силу воздействия неучтенных случайных факторов и причин отдельные наблюдаемые значения переменной y будут в большей или меньшей степени отклоняться от функции регрессии. В этом случае уравнение (4.1) дополняют случайным членом u_t и запись

$$y_t = a_0 + a_1 x_t + u_t \qquad (4.2)$$

называется *парной регрессионной моделью.*

Предположим, что для оценки линейной регрессии (4.2) взята выборка, содержащая n пар значений переменных (x_i, y_i), где i=1, 2, … ,n. Тогда (4.2) можно представить в виде:

$$y_i = a_0 + a_1 x_i + u_i \qquad (4.3)$$

Отметим основные предпосылки регрессионного анализа.

1. В модели (4.3) возмущение u_i есть величина случайная, а объясняющая переменная x_i - величина неслучайная.
2. Математическое ожидание u_i равно 0:
   $$M(u_i)=0. \qquad (4.4)$$
3. Дисперсия возмущения u_i постоянна для любого i[4]:
   $$D(u_i)=\sigma^2 \qquad (4.5)$$
4. Случайные возмущения u_i и u_j не коррелированы:
   $$Cov(u_i, u_j)=0. \qquad (4.6)$$
5. Возмущение u_i есть нормально распределенная случайная величина.

В этом случае модель (4.3) называется *классической нормальной линейной регрессионной моделью.*

Требование выполнение предпосылки 5 необходимо для оценки точности параметров уравнения регрессии. Сами параметры определяются при этом методом наименьших квадратов (см. раздел 2).

Возникает вопрос, являются ли оценки параметров $(\overline{a_0}, \overline{a_1}; \overline{\sigma_u})$ наилучшими? Ответ на него дает следующая теорема.

Теорема Гаусса-Маркова. Если регрессионная модель (4.3) удовлетворяет предпосылкам 1)-4), то оценки коэффициентов модели имеют наименьшую дисперсию в классе всех несмещенных линейных оценок.

---

[4] Случайные возмущения, удовлетворяющие этому условию, называются *гомоскедастичными.*

Таким образом, оценки коэффициентов $(\tilde{a}_0, \tilde{a}_1)$ в определенном смысле являются наиболее *эффективными*. Оценка $\tilde{\sigma}_u$ при этом рассчитывается по известной формуле

$$\tilde{\sigma}_u = \sqrt{\frac{\tilde{u}_1^2 + \tilde{u}_2^2 + \ldots + \tilde{u}_n^2}{n-2}}.$$

5. Проверка адекватности предпосылок теоремы Гаусса-Маркова

Компактную запись модели (4.3) представим в развернутом виде:

$$\left.\begin{aligned} y_1 &= a_0 + a_1 x_1 + u_1 \\ y_2 &= a_0 + a_1 x_2 + u_2 \\ &\ldots\ldots \\ y_n &= a_0 + a_1 x_n + u_n \end{aligned}\right\} \qquad (5.1)$$

Проверим для полученной системы уравнений наблюдений (5.1) справедливость второй предпосылки теоремы Гаусса-Маркова, согласно которой

$$D(u_1)= D(u_2)= \ldots = D(u_n)=\sigma^2. \qquad (5.2)$$

Для этого проведем

ТЕСТ ГОЛДФЕЛДА-КВАНДТА.

Чтобы проверить, не нарушается ли условие (5.2) из-за положительной (или отрицательной) зависимости $D(u_t)$ от $|x_t|$ упорядочим уравнения в (5.1) в порядке возрастания значений $|x_t|$ объясняющей переменной. Разобьем упорядоченную систему уравнений на примерно одинаковые две части. По первым m упорядоченным уравнениям, где

$$k<m \quad n/2, \qquad (5.3)$$

k=2 – количество оцениваемых коэффициентов для системы (5.1), вычисляем методом наименьших квадратов оценки параметров этой модели. Используя их, вычислим сумму квадратов остатков (ESS_1) для m первых уравнений (5.1).

Замечание 3. При нахождении мнк-оценок модели в Excel величина ESS находится в ячейке B,n+5 (Рис. 3).

Аналогичным образом оценив модель (5.1) по второй половине уравнений (5.1), найдем значение ESS_2.

Вычисляем величину *статистики Голдфелда-Квандта*

$$GQ = \frac{ESS_1}{ESS_2} = \frac{\sum_{t=1}^{m} \tilde{u}_t^2}{\sum_{t=m+1}^{n} \tilde{u}_t^2}. \qquad (5.4)$$

Статистика GQ является случайной величиной, подчиняющейся закону распределения, который называется *распределением Фишера(F)* со степенями свободы

$$f_1=f_2=m-k. \qquad (5.5)$$

Таблица значений $F(f_1,f_2)$ приведена в *Приложении 1*. Находим в ней значение $F_{крит}$ по степеням свободы из (5.5). Если одновременно будут выполняться условия

$$GQ \quad F_{крит}, \qquad (5.6)$$

$$GQ^{-1} \quad F_{крит},$$

то предпосылка (5.2) признается адекватной[5].

Задача 4. Проверить условие гомоскедастичности случайных возмущений для модели

$$Y_t=b*X_t+\varepsilon_t.$$

Результаты наблюдений указаны в таблице.

Y	70	65	55	60	50	35	40	30	25	32
X	5	11	15	17	20	22	25	27	30	35

Решение. Записав уравнения наблюдения типа (5.1) по заданной в задачи модели отмечаем, что они уже упорядочены по возрастанию |X|. Поскольку коэффициентов в модели, которые нужно оценить, всего один, то k=1, n=10, отсюда m=n/2=5.

Оценим методом наименьших квадратов нашу модель по первым пяти уравнениям:

$$Y_t=\underset{(0,95)}{3{,}689}*X_t+\underset{(30,93)}{\varepsilon_t}$$

Найдем оценки случайных остатков для каждого из пяти наблюдений:

$$\tilde{\varepsilon}_1=70-3{,}689*5=51{,}56$$

$$\tilde{\varepsilon}_2=65-3{,}689*11=24{,}42$$

$$\tilde{\varepsilon}_3=55-3{,}689*15=-0{,}33$$

$$\tilde{\varepsilon}_4=60-3{,}689*17=-2{,}7$$

$$\tilde{\varepsilon}_5=50-3{,}689*20=-23{,}77$$

Рассчитаем, используя полученные результаты, сумму квадратов остатков:

[5] Если вторая предпосылка теоремы Гаусса-Маркова не выполняется, то случайные возмущения в (5.1) называются *гетероскедастичными*.

$$ESS_1 = 51{,}56^2 + 24{,}42^2 + (-0{,}33)^2 + (-2{,}7)^2 + (-23{,}77)^2 = 3827{,}264$$

Оценим методом наименьших квадратов модель по оставшимся пяти уравнениям

$$Y_t = \underset{(0,15)}{1{,}123} * X_t + \underset{(9,71)}{\varepsilon_t}$$

и аналогичным образом найдем соответствующую им сумму квадратов остатков:

$$ESS_2 = 377{,}1542.$$

Вычисляем статистику Голдфелда-Квандта:

$$GQ = 3827{,}264/377{,}1542 = 10{,}14774;$$

$$GQ^{-1} = 1/10{,}14774 = 0{,}098544.$$

Определим степени свободы $f_1=f_2=m-k=5-1=4$ и из Таблицы значений дроби Фишера (*Приложение 1*) находим $F_{крит}= 6{,}39$. Проверим выполнение условий (5.6): GQ $F_{крит}$, хотя GQ^{-1} $F_{крит}$. Поскольку первое из них не выполняется, признаем предпосылку (5.2) теоремы Гаусса-Маркова неадекватной. Поэтому случайные возмущения в нашей модели можно признать гетероскедастичными.

ТЕСТ ДАРБИНА-УОТСОНА

Тест Дарбина – Уотсона применяется для проверки четвертой предпосылки теоремы Гаусса-Маркова о некоррелированности случайных остатков в (5.1).

Найдем мнк-оценки модели (5.1) по всем уравнениям наблюдения. Вычислим величину

$$DW = \frac{\sum_{t=2}^{n}(\tilde{u}_t - \tilde{u}_{t-1})^2}{\sum_{t=1}^{n}\tilde{u}_t^2}, \qquad (5.7)$$

которая называется *статистикой Дарбина-Уотсона*. Значение DW находится между 0 и 4. Критическое значение DW зависит от числа n – количества наблюдений в (5.1) и k – количества объясняющих переменных. По этим значениям из таблицы критических значений DW (*Приложение 2*) определим нижнюю и верхнюю границы интервала ($d_н$,$d_в$). Нанесем полученные результаты на прямую (Рис. 5). Определим, в какое из образовавшихся пяти подмножеств попадает рассчитанное значение DW.

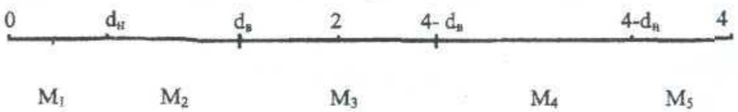

Рис. 5.

- Если DW принадлежит M_3, то четвертая предпосылка теоремы Гаусса-Маркова адекватна, т.е. случайные возмущения в (5.1) не коррелированны;

- Если DW принадлежит M_1, то между ними существует положительная корреляция, т.е. $Cov(u_i, u_j)>0$;
- Если DW принадлежит M_5, то между ними существует отрицательная корреляция, т.е. $Cov(u_i, u_j)=0$;
- Если DW принадлежит M_2 или M_4, то принять решение об адекватности или неадекватности предпосылки (4.6) нельзя.

Задача 5. Заданы реальный доход на душу населения y (тыс. долл.), процент рабочей силы, занятой в сельском хозяйстве x_1 и средний уровень образования населения в возрасте после 25 лет x_2 (число лет, проведенных в учебных заведениях) для 15 развитых стран в 1983 году. Составить спецификацию и проверить выполнение предпосылки (4.6) о некоррелированности случайных возмущений, включенных в модель.

Страна	y	x_1	x_2	Страна	y	x_1	x_2
1	7	8	9	9	10	6	12
2	9	9	13	10	11	7	14
3	9	7	11	11	11	6	11
4	8	6	11	12	12	4	15
5	8	10	12	13	9	8	15
6	4	4	16	14	10	5	10
7	9	5	11	15	12	8	13
8	8	5	11				

Решение. Представим в качестве спецификации модели по заданному условию множественную регрессию

$$y_t = a_0 + a_1 x_1 + a_2 x_2 + v_t.$$

Используя статистические данные, оценим неизвестные параметры модели:

$$y = \underset{(2,55)}{5{,}084} - \underset{(0,19)}{0{,}409}x_1 + \underset{(0,17)}{0{,}602}x_2 + \underset{(1,26)}{v_t}$$

Рассчитывая, как и в задаче 4 оценки случайных возмущений для каждого уравнения наблюдения, по (5.7) находим значение статистики Дарбина-Уотсона:

$$DW = \frac{29{,}8413}{19{,}073} = 1{,}565.$$

Поскольку в нашей модели количество объясняющих переменных k=2, количество уравнений наблюдений n=15, то ($d_н$,$d_в$)=(095;1,54) (см. *Приложение 2*). Согласно рисунку

5 это означает, что DW принадлежит M_3, что говорит об отсутствии корреляции между случайными остатками.

Положительное прохождение теста Дарбина-Уотсона дает основание полагать, что регрессионная зависимость между переменными указана правильно.

Часто причиной неадекватности предпосылки (4.6) является ошибка в построенной спецификации регрессионной модели (например, пропущена значимая объясняющая переменная). Поэтому тест Дарбина-Уотсона в эконометрике считается одним из самых важных. Однако прежде, чем проводить его, нужно убедиться в качественности полученной спецификации модели. Для этого используют F-тест.

6. Проверка качества спецификации модели

F-ТЕСТ

В качестве меры объясняющей способности регрессора x_t в рамках модели (4.2) может служить *коэффициент детерминации*

$$R^2 = \frac{\widetilde{D}(\widetilde{y}_t)}{\widetilde{D}(y_t)} = 1 - \frac{\widetilde{D}(\widetilde{u}_t)}{\widetilde{D}(y_t)}. \qquad (6.1)$$

Замечание 4. Статистическое значение этой величины, рассчитанное функцией Линейн, находится в ячейке А,n+3 (Рис.3).

Значение R^2 находятся в пределах от 0 до 1 включительно. Однако в силу стохастичного характера этой величины нельзя однозначно утверждать о хорошем (даже при $R^2=1$) или, наоборот, плохом (если $R^2=0$) качестве спецификации модели. Рассчитаем величину

$$F = \frac{\frac{R^2}{k}}{(1-R^2)/(n-(1+k))}, \qquad (6.2)$$

которая также случайна. Она имеет F-распределение со степенями свободы

$$f_1 = k, \qquad f_2 = n-(k+1),$$

где k- количество объясняющих переменных модели. Отыскав по степеням свободы критическое значение $F_{крит}$ *дроби Фишера (Приложение 1)*, проверяем выполнение неравенства

$$F \quad F_{крит} \qquad (6.3)$$

Если оно выполняется, то регрессоры не обладают объясняющей способностью в рамках модели. Ее спецификация признается некачественной и подлежит уточнению.

Если же неравенство (6.3) не выполняется, то объясняющая способность регрессоров характеризуется значением R^2, а качество спецификации модели признается удовлетворительным.

Замечание 5. Значение F при использовании функции Линейн находится в ячейке А,n+4; степень свободы f_2 - в B,n+4 (Рис.3).

Задача 6. Проверить при помощи F-теста качество спецификации модели, построенной по условию задачи 5.

Решение. Для спецификации

$$y_t = a_0 + a_1 x_1 + a_2 x_2 + v_t$$

при оценивании получаем коэффициент детерминации $R^2=0{,}6216$. Подставим его в (6.2) и при k=2, n=15 рассчитаем F=9,855. По степеням свободы $f_1=2$ и $f_2=15-(2+1)=12$ определяем табличное значение дроби Фишера: $F_{крит}=3{,}89$. Поскольку неравенство (6.3) не выполняется, т.к. $9{,}855 \geq 3{,}89$, делаем вывод о значимости указанных в модели регрессоров по отношению к эндогенной переменной y_t. О хорошем качестве построенной спецификации свидетельствует и достаточно высокий показатель коэффициента детерминации.

7. Проверка адекватности модели.

В разделе 2 мы рассмотрели процесс оценивания модели линейной парной регрессии

$$y_t = a_0 + a_1 x_t + u_t \qquad (7.1)$$

по обучающей выборке значений

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n). \qquad (7.2)$$

В рамках модели (7.1) значения переменных из (7.2) будут связаны уравнениями

$$\left.\begin{aligned} y_1 &= a_0 + a_1 x_1 + u_1 \\ y_2 &= a_0 + a_1 x_2 + u_2 \\ &\ldots\ldots \\ y_n &= a_0 + a_1 x_n + u_n. \end{aligned}\right\} \qquad (7.3)$$

По полученной оцененной модели

$$\left.\begin{aligned} y_t &= \widetilde{a}_0 + \widetilde{a}_1 \cdot x_t + u_t \\ &(S_{a_0})\,(S_{a_1})\ \ (\widetilde{\sigma}_u) \end{aligned}\right\} \qquad (7.4)$$

прогнозируемое значение эндогенной переменной для известного значения предопределенной переменной x_0 из контролирующей выборки (y_0,x_0) рассчитывается путем его подстановки в оцененное уравнение регрессии:

$$\tilde{y}_0 = \tilde{a}_0 + \tilde{a}_1 \cdot x_0 . \qquad (7.5)$$

Рассчитаем среднеквадратическую ошибку прогноза (7.5):

$$S_{y_0} = \tilde{\sigma}_u \cdot \sqrt{1+q_0}\,, \qquad (7.6)$$

$$q_0 = \frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}, \qquad (7.7)$$

где x_i- значения регрессоров из обучающей выборки, $\bar{x}$ - их среднее значение.

Замечание 6. Такая *процедура точечного прогноза* используется и для линейной модели множественной регрессии. Однако формула расчета величины q_0 (7.7) для множественной регрессии неприменима. Для среднеквадратической ошибки прогноза эндогенной переменной, которая находится, например, по правилу:

$$\tilde{y}_0 = \tilde{a}_0 + \tilde{a}_1 \cdot x_{1,0} + \tilde{a}_2 \cdot x_{2,0}$$

величина q_0 рассчитывается так:

$$q_0 = \vec{x}_0^{\,T} \cdot (X^T \cdot X)^{-1} \cdot \vec{x}_0, \qquad (7.8)$$

$$\vec{x}_0 = \begin{pmatrix} 1 \\ x_{1,0} \\ x_{2,0} \end{pmatrix}, \qquad (7.9)$$

$$X = \begin{pmatrix} 1 & x_{1,1} & x_{2,1} \\ 1 & x_{1,2} & x_{2,2} \\ \vdots & \vdots & \vdots \\ 1 & x_{1,n} & x_{2,n} \end{pmatrix}. \qquad (7.10)$$

В вектор (7.9), кроме 1, входят значения предопределенных переменных из контролирующей выборки, матрица (7.10) состоит из значений предопределенных переменных из обучающей выборки и расширена столбцом единиц. В (7.8) $\vec{x}_0^{\,T}$ и X^T – транспонированные соответственно вектор (7.9) и матрица (7.10), $(X^T*X)^{-1}$ – обратная матрица к матрице-произведению X^T на X[6].

[6] В Excel для транспонирования, умножения и поиска обратной матрицы используются функции ТРАНСП, МУМНОЖ и МОБР.

Точечный прогноз (7.5) можно расширить до *интервального*, если построить *доверительный* интервал $[y_0^-, y_0^+]$, который будет с принятой доверительной вероятностью β накрывать прогнозируемое значение $\tilde{y}_0$:

$$\left.\begin{aligned} y_0^- &= \tilde{y}_0 - t_{крит} \cdot S_{y_0} \\ y_0^+ &= \tilde{y}_0 + t_{крит} \cdot S_{y_0} \end{aligned}\right\}. \qquad (7.11)$$

$t_{крит}$ – критическое значение *дроби Стьюдента* (см. *Приложение 3)*, которая имеет смысл нормированной ошибка прогноза и при выполнении предпосылок (4.4), (4.5) о случайном возмущении, как случайная величина, обладает t-распределением с числом степеней свободы

$$f_2 = n-(k+1), \qquad (7.12)$$

где $(k+1)$ – количество оцениваемых коэффициентов модели.

Оцененная модель (7.4) при интервальном прогнозировании признается адекватной, если значение y_0 из контролирующей выборки принадлежит $[y_0^-, y_0^+]$. В противном случае модель неадекватна и подлежит доработке.

Задача 7. Имеются следующие данные о годовых ставках месячных доходов по трем акциям за шестимесячный период:

Акция	Доходы по месяцам, %					
A	5,4	5,3	4,9	4,9	5,4	6,0
B	6,3	6,2	6,1	5,8	5,7	5,7
C	9,2	9,2	9,1	9,0	8,7	8,6

Есть основания полагать, что доходы Y по акции С зависят от доходов X_1 и X_2 по акциям А и В. Составить спецификацию модели, по данным за пять месяцев оценить ее. Выбрав в качестве контролирующей выборки данные за шестой месяц, для доходов по акциям А и В построить интервальный прогноз для доходаY по акции С. Проверить адекватность оцененной модели.

Решение. В качестве спецификации модели рассмотрим множественную регрессию

$$Y_t = a_0 + a_1 X_{1t} + a_2 X_{2t} + u_t.$$

По обучающей выборке оценим ее:

$$Y_t = \underset{(1.123)}{5{,}633} - \underset{(1.157)}{0{,}248} X_{1t} + \underset{(0.157)}{0{,}78} X_{2t} + \underset{(0.08)}{u_t} \qquad (7.13)$$

Прогнозное значение дохода по акции С для шестого месяца получим по формуле

$$\tilde{Y}_6 = 5{,}633 - 0{,}248 X_{16} + 0{,}78\, X_{26}$$

Для X_{16}=6 и X_{26}=5,7 прогноз дохода $\widetilde{Y}_6$=8,586797. Найдем среднеквадратическую ошибку этого прогноза по (7.6). Для этого сначала рассчитаем q_0 по (7.8).

$$\overset{\times}{x}_0=\begin{pmatrix}1\\6\\5{,}7\end{pmatrix};\quad X=\begin{pmatrix}1&5.4&6.3\\1&5.3&6.2\\1&4.9&6.1\\1&4.9&5.8\\1&5.4&5.7\end{pmatrix};\quad \overset{\times}{x}{}_0^T=(1\quad 6\quad 5{,}7);\ X^T=\begin{pmatrix}1&1&1&1&1\\5.4&5.3&4.9&4.9&5.4\\6.3&6.2&6.1&5.8&5.7\end{pmatrix};$$

$$X^T\cdot X=\begin{pmatrix}5&25.9&30.1\\25.9&134.43&155.97\\30.1&155.97&181.47\end{pmatrix};\qquad (X^T\cdot X)^{-1}=\begin{pmatrix}197.83&-15.56&-19.44\\-15.56&3.87&-0.75\\-19.44&-0.75&3.88\end{pmatrix};$$

$$\overset{\times}{x}{}_0^T\cdot(X^T\cdot X)^{-1}=(-6.33\quad 3.42\quad -1.86);$$

$$q_0=\overset{\times}{x}{}_0^T\cdot(X^T\cdot X)^{-1}\cdot\overset{\times}{x}_0=3{,}598958.$$

$$S_{y_0}=\widetilde{\sigma}_u\cdot\sqrt{1+q_0}=0{,}17196.$$

Строим доверительный интервал для $\widetilde{Y}_6$=8,586797. При f_2=5-(2+1)=2 и задавшись доверительной вероятностью β=0,95 находим $t_{крит}$=4,3. По формулам (7.11)

$$Y_6^-=8{,}586797-4{,}3*0{,}17196=7{,}85;$$

$$Y_6^+=8{,}586797+4{,}3*0{,}17196=9{,}32.$$

Размер точного дохода по акции С в шестом месяце составил Y_6=8,6. Это число попадает в прогнозный интервал (Y_6^-, Y_6^+), поэтому оцененная модель (7.13) признается адекватной и пригодной для прогнозирования величины дохода по акции С.

Приложени1.

Критические значения функции распределения Фишера F (f_1,f_2)

f_1 / f_2	1	2	3	4	5	6	7
1	161	200	216	225	230	234	237
2	18,9	19,0	19,2	19,2	19,3	19,3	19,3
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09
5	6,61	5,49	5,41	5,19	5,05	4,95	4,88
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14
11	4,84	3,98	3,36	3,20	3,09	3,01	2,90
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83
14	4,62	3,74	3,34	3,11	2,96	2,85	2,76
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51

Приложение 2.

Значение границ интервала ($d_н$,$d_в$) критических значений статистики DW критерия Дарбина-Уотсона (на уровне значимости α=0,05)

n	k=1		k=2		k=3	
	$d_н$	$d_в$	$d_н$	$d_в$	$d_н$	$d_в$
15	1,08	1,36	0,95	1,54	0,82	1,75
16	1,10	1,37	0,98	1,54	0,86	1,73
17	1,13	1,38	1,02	1,54	0,90	1,71
18	1,16	1,39	1,05	1,53	0,93	1,69
19	1,18	1,40	1,08	1,53	0,97	1,68
20	1,20	1,41	1,10	1,54	1,00	1,68

Приложение 3.

Значения $t_{крит}$ дроби Стьюдента

f_c	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
β =0,95	12,71	4,3	3,18	2,78	2,57	2,45	2,36	2,31	2,26	2,23	2,20	2,18	2,16	2,14	2,13
β =0,98	31,82	6,96	4,54	3,75	3,36	3,14	3,00	2,9	2,82	2,76	2,72	2,68	2,65	2,62	2,6
β =0,99	63,66	9,92	5,84	4,60	4,03	3,71	3,50	3,35	3,25	3,17	3,11	3,05	3,01	2,98	2,95

Литература

1. Бывшев В.А. Введение в эконометрию. Часть 2. –М.: ФА при Правительстве РФ, 2003.
2. Грицан В.Н. Эконометрика.- М.: «Дашков и К», 2002.
3. Катышев П.К. Магнус Я.Р., Пересецкий А.А. Сборник задач к начальному курсу эконометрики. – М.: «Дело», 2003.
4. Кремер Н.Ш., Путко Б.А. Эконометрика. – М.: ЮНИТИ, 2003.