МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ УКРАИНЫ
ДОНЕЦКИЙ НАЦИОНАЛЬНЫЙ
УНИВЕРСИТЕТЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ

КРАТКИЙ ЭКСКУРС В РЕГРЕССИОННЫЙ АНАЛИЗ

(В помощь студентам, изучающим курс «Эконометрия»)

*Утверждено на заседании кафедры
математики и математических
методов в экономике
Протокол № 7 от 16.03.2000*

Донецк, ДонНУ - 2000

ББК 65.053
Л 84

Краткий экскурс в регрессионный анализ. В помощь студентам, изучающим курс «Эконометрия». / Сост.: Кривенчук О.Г. – Донецк: ДонГУ, 2000. - 24 с./

Содержит лекционный материал по двум темам курса «Эконометрия»: «Краткий экскурс в регрессионный анализ», «Показатели адекватности регрессионной модели». Материал изложен в соответствии с программой курса, утвержденной Министерством образования и науки Украины.

Рекомендуется студентам всех форм обучения и всех специальностей, изучающих курс «Эконометрия».

Составитель: ***О.Г. Кривенчук***, к.э.н., доцент
Электронный вариант ***В.Д. Породникова,*** доцент

Отв. за выпуск ***В.В. Христиановский***, проф.

Введение

Эконометрия является сравнительно молодой отраслью науки. Буквально термин «Эконометрия» обозначает «измерения в экономике».

Эконометрию в узком понимании можно определить как дисциплину, которая с помощью статистических методов пытается установить количественные взаимосвязи между экономическими переменами.

Применяемые в эконометрии методы базируются на разделах регрессионного анализа. Взаимосвязи, которые исследуются с помощью этих методов, например, функции спроса или производственные функции, являются важнейшими моментами экономической теории.

Классическая линейная регрессионная модель является основой для построения обобщенных эконометрических моделей. Представления о ее построении студенты получают при изучении курса «Теория вероятностей и математическая статистика». Однако знания по этой проблеме недостаточны. В предлагаемом учебном пособии тема построения регрессионной модели расширенна за счет изложения методов проверки адекватности модели, а также экономической интерпретации параметров и расчета некоторых экономических показателей.

Краткий экскурс в регрессионный анализ

Исходным пунктом любого регрессионного анализа является следующая ситуация: объект исследования представлен наблюдаемыми величинами Y_1 X_1, $X_к$, между которыми имеется объективная связь. Эту связь можно представить линейной функцией.

Однако в действительности наблюдаемые величины отклоняются от этой функциональной связи. Эти отклонения включаются в модель дополнительной случайной переменной U.

Регрессионное уравнение имеет вид:

$$Y = \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n + U \qquad (1)$$

Математическое ожидание зависимой переменной (регрессанда Y при заданных значениях независимых переменных (регрессоров) $x_1, x_2, \ldots x_к$.

$$E\,(Y/x_1, x_2,\ldots,x_k)=\beta_1 x_1+\ldots+\beta_k x_k \qquad (2)$$

Числовой пример: Предприятие имеет большое количество филиалов и руководство этого предприятия хотело бы знать, как Y (годовой товарооборот одного филиала) функционально зависит от:

x_2 - торговой площади (тыс м^2),

x_3 - среднедневной интенсивности потока покупателей (тыс. чел/день)

Для каждого t филиала (T=12) имеем данные

Таблица 1

№ филиала t	Y_t	X_{t2}	X_{t3}
1	2,93	0,31	10,24
2	5,27	0,98	7,51
3	6,85	1,21	10,81
4	7,01	1,29	9,89
5	7,02	1,12	13,72
6	8,35	1,49	13,92
7	4,33	0,78	8,54
8	5,77	0,94	12,36
9	7,68	1,29	12,27
10	3,16	0,48	11,01
11	1,52	0,24	8,25
12	3,15	0,55	9,31

Следуя традиции включаем в модель параметр β_1 (свободный член), геометрически интерпретируемый отрезком, отсекаемым прямой регрессии на координатной оси. Переменной X_1 при β_1 присвоим для каждого филиала значение $x_{t1}=1(t=1, 12)$.

Чтобы получить конкретную оценку трех коэффициентов β_1, β_2, β_3 (или рассчитать множественную регрессию), необходимо применить статистический метод.

Здесь используем одношаговый метод наименьших квадратов (1-МНК) - оценщик.

Составим систему регрессионных уравнений.

$$
\begin{aligned}
&y_1=\beta_1 x_{11}+\ldots+\beta_k x_{1k}+\ldots+\beta_K x_{1K}+u_1;\\
&y_2=\beta_1 x_{21}+\ldots+\beta_k x_{2k}+\ldots+\beta_K x_{2K}+u_2;\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&y_t=\beta_1 x_{t1}+\ldots+\beta_k x_{tk}+\ldots+\beta_K x_{tK}+u_t;\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&y_T=\beta_1 x_{T1}+\ldots+\beta_k x_{Tk}+\ldots+\beta_K x_{TK}+u_T.
\end{aligned}
\tag{3}
$$

В матричном виде

$$Y=X\beta+U. \tag{4}$$

Вектор наблюдений Y и матрица X вместе образуют матрицу данных

$D :=[Y:X]$

Систематическая часть регрессанда равна

$$\tilde{y}_t = \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + \ldots + \beta_K x_{tK}, \tag{5}$$

$$y_t = \tilde{y}_t + u_t.$$

Систематическая часть совпадает с математическим ожиданием регрессанда

$$\tilde{y}_t = E(Y_t). \tag{6}$$

Линейное уравнение множественной регрессии, состоящее из систематической части с конкретными значениями $\check{\beta}_k$ (к=1, 2…К) является эмпирической функцией регрессии:

$$\check{y}_t = \check{\beta}_t\, x_t + \ldots + \check{\beta}_{tk} + \ldots + \check{\beta}_K\, x_{tK}\,. \qquad (7)$$

$\check{y}_t$ - прогнозируемая величина математического ожидания регрессанда (вычисляется по конкретной эмпирической функции регрессии).

Погрешности в регрессионном уравнении (или остатки, или ошибки, или возмущения) рассчитываются по формуле:

$$\check{u}_t := y_t - \check{y}_t\,.$$

При оценке коэффициентов $\beta_к$ по 1МНК – оценщику применяется следующая целевая функция: сумма квадратов ошибок должна быть минимальной. Это и есть мера качества адаптации эмпирической регрессионной функции к данным.

Формирование целевой функции:

$$\check{u}_t^{\,2} = \left(y_t - \check{\beta}_1\, x_{t1} - \ldots - \check{\beta}_k\, x_{tk} - \ldots - \beta_K x_{tK} \right)^2$$

сумма квадратов для всех T наблюдений

$$S\left(\check{\beta}\right) = \sum_{t=1}^{T} \check{u}_t^{\,2} = \sum_{t=1}^{T} \left(y_t - \check{\beta}_1 x_{t1} - \ldots - \check{\beta}_K x_{tK} \right)^2 = (Y - X\check{\beta})'(Y - X\check{\beta}) = \check{u}'\check{u} \qquad (8)$$

где $\check{Y} = X\check{\beta}$ и $\check{u} = Y - \check{Y}$.

Необходимым условием существования минимума S (β) является равенство нулю градиентного вектора $\partial S(\check{\beta})/\partial\check{\beta}$ (частные производные)

$$\frac{\partial S\left(\check{\beta}\right)}{\partial \check{\beta}} = \frac{\partial\left(y'y\right)}{\partial \check{\beta}} - 2\frac{\partial\left(\check{\beta}' x'y\right)}{\partial \check{\beta}} + \frac{\partial\left(\check{\beta}' x'x\check{\beta}\right)}{\partial \check{\beta}} = \underbrace{2x'y}_{к\,-вектор} + \underbrace{2x'x\check{\beta}}_{к-вектор} = 0.$$

Разделив на 2 и упорядочив систему К уравнений относительно неизвестных β_k(k=1,…K), получим <u>систему нормальных уравнений.</u>

Обозначим $\hat{\beta}_k$ то значения $\check{\beta}_k$, при котором система равна 0

$$x'x\hat{\beta} = x'y. \tag{9}$$

Решение системы в матричном виде

$$\hat{\beta} = (X'X)^{-1} X'Y. \tag{10}$$

Вектор $\hat{\beta}$ минимизирует сумму квадратов ошибок.

Для нашего примера

$$x = \begin{pmatrix} 1 & 0{,}31 & 10{,}24 \\ 1 & 0{,}98 & 7{,}51 \\ & & \\ 1 & 0{,}55 & 9{,}31 \end{pmatrix} \quad x' = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 0{,}31 & 0{,}98 & \dots & 0{,}55 \\ 10{,}24 & 7{,}51 & \dots & 9{,}31 \end{pmatrix}$$

$$y = \begin{pmatrix} 2{,}93 \\ 5{,}27 \\ 3{,}15 \end{pmatrix}$$

Порядок действий: 1) умножим $X'X$;

2) умножим $X'Y$;

3) найдем обратную матрицу $(X'X)^{-1}$;

4) умножим обратную матрицу на $(X'Y)$.

Найдем вектор $\hat{\beta} = \begin{pmatrix} -0{,}832 \\ 4{,}743 \\ 0{,}175 \end{pmatrix}$

$$\hat{Y} = -0{,}832 + 4{,}743X_2 + 0{,}175X_3$$

$\hat{Y}$ – оцененная множественная функция регрессии.

Интерпретация оценок:

$\hat{\beta}_2 = 4{,}743$: если при прочих равных условиях переменная X_2 увеличится на единицу своего измерения, то Y увеличится на 4,743 ед. своего измерения.

Увеличение торговой площади на 1тыс.$м^2$ приведет к увеличению годового товарооборота филиала на 4,743 тыс.грн. (при прочих равных условиях)

Уравнение множественной регрессии можно использовать как эконометрическую прогнозную модель.

<u>Пример.</u> Спрогнозировать годовой товарооборот филиала, который должен быть построен в относительно заселенном районе со средней интенсивностью потока покупателей 15000 (чел/в день) $x_{t3}=15$ и торговой площадью 1200$м^2$ ($x_{t2}=1{,}2$).

$$\hat{Y}_t = -0{,}832 + 4{,}743 * 1{,}2 + 0{,}175 * 15 = 7{,}485 \text{ (тыс.грн).}$$

Эконометрическую модель также можно использовать для принятия решений с фиксированной целью.

<u>Пример.</u> Оценим, какое значение, при прочих равных условиях должна иметь переменная x, чтобы при заданных величинах других переменных была получена необходимая величина y (переменной цели).

Построить филиал на улице со средним потоком покупателей 100000 (чел/в день). При этом желательная величина товарооборота – 7 тыс.грн.

Инструмент достижения цели – торговая площадь

$$\mathbf{Z=-0{,}832+4{,}743x_{t2}+0{,}175*10;}$$

$$\mathbf{X_{t2}=1{,}282.}$$

На основе оценщика $\hat{\beta}$ определим оценки для других величин модели: для $E(y_t)$; u_t и σ_u^2.

$\hat{y}_t$ - прогноз величины y (ее математическое ожидание) для наблюдаемого объекта t (или периода времени t)

$$\hat{y} = x\hat{\beta} \qquad (11)$$

$$\hat{y} = \begin{pmatrix} 1 & 0,31 & 10,24 \\ 1 & 0,98 & 7,51 \\ 1 & 0,55 & 9,31 \end{pmatrix} \begin{pmatrix} -0,832 \\ 4,743 \\ 0,175 \end{pmatrix}$$

Оценка возмущений:

$$\hat{u} = y - \hat{y} = y - x\hat{\beta} . \qquad (12)$$

Свойство вектора ошибок при оценщике 1МНК

$$\sum_{t=1}^{T} \hat{u}_t = 0 \quad \text{(но не их квадратов).} \qquad (13)$$

Относительная ошибка прогноза рассчитывается по формуле

$$\frac{\hat{u}_t}{y_t} \cdot 100 \text{ (в \%).} \qquad (14)$$

Дисперсия возмущений (ошибок) u_t – важная характеристика регрессионной модели. Ее величина должна быть как можно меньше.

$$\hat{\sigma}_u^2 = \frac{\text{сумма квадратов ошибок}}{\text{число степеней свободы}} = \frac{\sum_{t=1}^{T} u_t^2}{T - K} = \frac{y'y - \hat{y}y}{T - K} = \frac{y'y - \hat{\beta}x'y'}{T - K}, \qquad (15)$$

где T - длина ряда;

К - количество регрессоров.

Если ошибки уже оценены, то

$$\hat{\sigma}_u^2 = \frac{\hat{u}'\hat{u}}{T - K} .$$

Стандартизованные регрессионные коэффициенты не учитывают единицы измерения регрессоров, поэтому применяются для сравнения влияния объясняющих переменных на Y. Они рассчитываются по формуле:

$$\hat{\beta}_k^S = \hat{\beta}_k \frac{S_k}{S_y}; \qquad (k=2,\ldots K; \quad S_y > 0), \qquad (16)$$

где S_k - эмпирическое стандартное (среднеквадратическое) отклонение k-го регрессора $x_к$;

S_y - эмпирическое стандартное (среднеквадратическое) отклонение регрессанда Y.

Так как для x_1 среднеквадратическое отклонение =0, то $\hat{\beta}_1^s$ не имеет смысла.

Для филиалов предприятий

$$\hat{\beta}_2^S = \hat{\beta}_2 \cdot \frac{S_2}{S_y} = 4{,}742 \cdot \frac{0{,}416}{2{,}205} = 0{,}894$$

$$S_k = \sqrt{\hat{\sigma}_k^2}$$

$\hat{\beta}_k S_k$ - типичный эффект k-го регрессора,

S_y - типичный эффект регрессанда.

Чем больше $\hat{\beta}_k^s$, тем более значимым, при прочих равных условиях является k-й регрессор.

Коэффициенты эластичности также не учитывают единицы измерения регрессоров. Они рассчитываются по формуле:

$$\varepsilon_k = \frac{\partial y_t}{\partial x_{tk}} \cdot \frac{x'_k}{y'} \quad (y' \neq 0;\ k = 2,\ldots K), \qquad (17)$$

где y' и x'_k - значения y и x_k, определяющие точку регрессионной функции, для которой вычисляется коэффициент эластичности.

Для линейного регрессионного уравнения

$$\frac{\partial y_t}{\partial x_{tk}} = \hat{\beta}_k$$

Значит:

$$\hat{\varepsilon}_k = \hat{\beta}_k \frac{x'_k}{y'}; \quad (y' \neq 0). \qquad (18)$$

Средняя эластичность определяется при $x'_k = \overline{x}$; $y' = \overline{y}$:

$$\hat{\overline{\varepsilon}}_k = \hat{\beta}_k \frac{\overline{x}_k}{\overline{y}}. \qquad (19)$$

Для нашего примера

$$\hat{\varepsilon}_2 = \hat{\beta}_2 \frac{\overline{x}_2}{\overline{y}} = 4{,}743 \ \frac{0{,}9}{5{,}253} = 0{,}813 \ . \qquad (20)$$

<u>Интерпретация</u>: Если X_2 изменится на 1%, то Y изменится на 0,813% (при прочих равных условиях: неизменной X_3 и среднем воздействии остальных неучтенных в модели случайных величин).

В классической регрессионной модели вектор U - случайный, значит и зависящий от него вектор Y - тоже случайный.

Так как вектор Y входит в функцию оценивания коэффициентов методом 1МНК, то и результат - вектор $\hat{\beta}$ интерпретируется как k-вектор случайных переменных.

Для характеристики случайных переменных $\hat{\beta}_k$ наряду с величинами математического ожидания используются также дисперсии $\sigma^2_{\hat{\beta}_k}$ и ковариации $\sigma_{\hat{\beta}_k \hat{\beta}_{k'}}$ (k; k'=1,…,K; $k \neq k'$)

Истинные значения этих параметров классической регрессионной модели образуют дисперсно-ковариационную (кратко: ковариационную) матрицу размерности К х К. Обозначим ее $\sum_\beta$.

Эта ковариационная матрица неизвестна, она должна быть оценена.

Оцененная методом 1МНК ковариационная матрица для параметров $\hat{\beta}$

$$\hat{\Sigma}_{\hat{\beta}} = \hat{\sigma}_u^2 (x'x)^{-1} = \begin{bmatrix} \sigma_{\hat{\beta}_1}^2 & \sigma_{\hat{\beta}_1\hat{\beta}_2} & \cdots & \sigma_{\hat{\beta}_1\hat{\beta}_k} \\ \sigma_{\hat{\beta}_2\hat{\beta}_1} & \sigma_{\hat{\beta}_2}^2 & \cdots & \sigma_{\hat{\beta}_2\beta_k} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{\hat{\beta}_k\hat{\beta}_1} & \sigma_{\hat{\beta}_k\beta_2} & \cdots & \sigma_{\hat{\beta}_k}^2 \end{bmatrix} \tag{21}$$

Элементы этой симметричной матрицы необходимы для t-тестирования гипотез, а также для расчетов доверительных интервалов. Чем больше дисперсии и ковариации, тем больше будут интервалы.

А так как желательны узкие доверительные и прогнозные интервалы, то для модели желательны малые значения элементов ковариационной матрицы.

Эту матрицу несложно вычислить

$$\hat{\sum}_{\hat{\beta}} = \hat{\sigma}_u^2 (x'x)^{-1}, \text{ где } \quad \hat{\sigma}_u^2 = \frac{1}{T-K} \hat{u}' \hat{u}. \tag{22}$$

Сначала находим обратную матрицу $(x'x)^{-1}$. Тогда обозначим (k, k') - й элемент этой матрицы $C_{kk'}$.

Элементы на главной диагонали

$$\hat{\sigma}_{\hat{\beta}_k}^2 = \hat{\sigma}_u^2 C_{kk'}; \quad (k=1,2\ldots K). \tag{23}$$

Для элементов вне диагонали

$$\hat{\sigma}_{\hat{\beta}_k\hat{\beta}_{k'}} = \sigma_u^2 c_{kk'}; \qquad (k \neq k').$$

Дисперсию возмущений σ_u^2 находим по формуле (15).

Отрицательный знак при $\hat{\sigma}_{\hat{\beta}_k\hat{\beta}_{k'}}$ говорит об обратной связи.

Однако о силе этой связи лучше свидетельствуют соответствующие <u>коэффициенты корреляции</u>.

$$R_{\hat\beta_k \hat\beta_{k'}} = \frac{\hat\sigma_{\hat\beta_k \hat\beta_{k'}}}{\hat\sigma_{\hat\beta_k} \hat\sigma_{\hat\beta_{k'}}} ; \qquad (24)$$

Например: $\hat{\Sigma}_{\hat\beta} = \begin{bmatrix} 1,8 & 0 & -0,4 \\ 0 & 0,4 & -0,2 \\ -0,4 & -0,2 & 0,2 \end{bmatrix}$

$\hat\sigma_{\hat\beta_1 \hat\beta_3} = -0,4$ (обратная связь ниже средней) .

Рассчитаем коэффициент корреляции

$R_{\hat\beta_1\hat\beta_3} = \frac{\sigma_{\hat\beta_1\hat\beta_3}}{\sigma_{\hat\beta_1}\sigma_{\hat\beta_3}} = \frac{-0,4}{\sqrt{1,8} * \sqrt{0,2}} = -\frac{2}{3}$ (достаточно тесная обратная связь).

<u>Следующие свойства МНК оценщика</u> $\hat\beta$ <u>в классической модели</u> говорят о том, что он является наиболее эффективным.

а) линейность по Y

$\hat\beta = AY + a$, где $A = (X'X)^{-1} X'$; $a = 0$;

б) несмещенность

$E\left(\hat\beta\right) = \beta$ (без доказательства);

в) МНК оценщик минимизирует покомпонентную дисперсию.

Теорема Гаусса-Маркова доказывает, что

$\sigma^2_{\hat\beta_k} \le \sigma^2_{\check\beta_k}$ (для к=1, 2…К), где

$\hat\beta$ - 1МНК – оценщик;

$\hat\beta$ - любой оценщик, например, по максимальному правдоподобию;

г) нормальное распределение $\hat\beta$.

В классической линейной модели с нормальным распределением возмущений u_t вектор $\hat{\beta}$ - нормально распределен;

д) состоятельность $\hat{\beta}$.

При этом $T \to \infty$ матрица ковариаций $\sum_{\hat{\beta}}$ сводится к нулю.

Выводы. Одношаговый метод наименьших квадратов обеспечивает в классической линейной регрессионной модели многие желанные статистические свойства оценщиков. 1МНК в этом плане "идеальный".

Но для эмпирических исследований часто нереальной является сама классическая регрессионная модель.

Она больше подходит к лабораторным условиям с контролируемыми регрессорами. При эмпирических исследованиях часть статистических свойств классической модели теряется. И все-таки, классическая регрессионная модель и относящиеся к ней методы оценки и проверки статистических гипотез образуют основу для развития обобщенных моделей, а также методов их оценки и тестирования.

Этим и занимается эконометрия.

Показатели адекватности классической регрессионной модели

В эмпирических экономических и социальных исследованиях из множества вариантов уравнений, которые отличаются регрессорами - "входящими х-ми", необходимо выбрать наиболее адекватную регрессионную функцию.

Для оценки адекватности функции имеющимся выборочным данным очень часто применяется коэффициент детерминации R^2.

а) R^2 - квадрат эмпирического коэффициента корреляции между двумя рядами наблюдений - y_t (t=1,…T) (теоретическими значениями регрессанда) и $\hat{y}_t (t = 1,...T)$ (его расчетными значениями)

$$R^2 = \frac{\left[\sum_{t=1}^{T}(y_t - \overline{y})(\hat{y}_t - \overline{y})\right]^2}{\sum_{t=1}^{T}(y_t - \overline{y})^2 \sum_{t=1}^{T}(\hat{y}_t - \overline{y})^2} \tag{25}$$

$0 \le R^2 \le 1$,

б) $$R^2 = \frac{\sum_{t=1}^{T}(\hat{y}_t - \overline{y})^2}{\sum_{t=1}^{T}(y_t - \overline{y})^2}. \tag{26}$$

Частное от деления суммы квадратов регрессии около средней на сумму квадратов наблюдаемой совокупности около средней.

в) $$R^2 = 1 - \frac{\sum_{t=1}^{T}\hat{u}_t^2}{\sum_{t=1}^{T}(y_t - \overline{y})^2} \tag{27}$$

в числителе: сумма квадратов погрешностей

в знаменателе: сумма квадратов совокупности около средней.

Если R^2=1 - случай полной адекватности, когда все наблюдаемые значения лежат в регрессионной гиперплоскости R^2=0. Функция регрессии в этом случае ничего не объясняет .

Это возможно при

$$\hat{\beta}_1 = \overline{y} \text{ и } \hat{\beta}_2 = \ldots = \hat{\beta}_k = 0$$

$$\hat{y}_t = \overline{y} \text{ для всех } t.$$

Регрессионное уравнение оценено тем лучше, чем больше R^2.

Из двух вариантов регрессионных уравнений, отличающихся регрессорами, лучшим считается тот, у которого R^2 больше (при прочих равных условиях).

<u>Связь между R^2 и количеством регрессоров R .</u>

Если включить в модель дополнительный регрессор (К+1-й), то всегда коэффициент детерминации будет больше

$R^2(K+1) \geq R^2(K)$.

Их разница ΔR^2 может быть определена после двойного расчета коэффициентов детерминации, однако это вычисление можно упростить.

Частный коэффициент детерминации ΔR_k^2 показывает предельный (граничный) вклад k-го регрессора в R^2, или показывает, на какую величину уменьшится коэффициент детерминации, если k-й регрессор (и только он!) будет исключен из группы регрессоров.

$$\Delta R_k^2 = \frac{(1 - R^2) t_k^2}{T - K}, \quad (28)$$

где $t_k = \dfrac{\hat{\beta}_k}{\hat{\sigma}_{\beta_k}}$ - это t - статистика для k -го регрессионного коэффициента

R^2 – коэффициент детерминации;

Т – число наблюдений;

К – число регрессоров;

Т-К – число степеней свободы.

Увеличение регрессоров приводит к росту R^2, однако снижает число степеней свободы (Т-К), что отрицательно сказывается при применении t – тестов и F – тестов, а также при построении доверительных и прогнозных интервалов.

Поэтому преимущество имеет скорректированный коэффициент детерминации, учитывающий число степеней свободы.

Скорректированный коэффициент детерминации по Тейлору:

$$\overline{R}_T^2 = 1 - (1 - R^2)\frac{T-1}{T-K}. \quad (29)$$

Скорректированный коэффициент детерминации по Анемии:

$$\overline{R}_A^2 = 1 - (1 - R^2)\frac{T+K}{T-K}. \quad (30)$$

Важное свойство скорректированных коэффициентов детерминации заключается в том, что приращение $\Delta\overline{R}_T^2$ и $\Delta\overline{R}_A^2$ может оказаться как положительным, так и отрицательным, то есть при отрицательном новая модель с дополнительным регрессором будет хуже, чем модель с меньшим числом регрессоров.

При этом $\overline{R}_A^2$ изменяется на большую величину, чем $\overline{R}_T^2$ после включения дополнительного регрессора, значит лучше отражает уменьшение числа степеней свободы.

t-тесты, как показатели адекватности модели

R^2; $\overline{R}_T^2$; $\overline{R}_A^2$; ΔR_k^2 - эти критерии относятся к регрессионному уравнению, как к совокупности, то есть к уравнению в целом .

t-тесты и F – тесты используются для проверки гипотез об истинных, но неизвестных значениях отдельных коэффициентов регрессии (или нескольких коэффициентов).

Условия проведения t-теста – наличие классической регрессионной модели с выполнением предпосылки о нормальном распределении:

а) вектор возмущений $U := (U_1; U_2; U_T)'$ является t-мерным нормально распределенным с нулевым вектором математического ожидания и единичной ковариационной матрицей σ_u^2;

б) регрессионная матрица X детерминирована и имеет полный ранг K.

Такие предпосылки дают возможность посредством t-тестов статистически проверить определенные гипотезы о числовых значениях коэффициентов $\beta_к$ ($1\leq k \leq K$) и о значениях отдельных линейных комбинаций этих коэффициентов.

t- тест двусторонней пары гипотез

<u>Вопрос:</u> существенно ли влияет k-й регрессор в генеральной выборке на регрессанд, иначе отличается ли истинное значения коэффициента $\beta_к$ от нуля?

Утверждение, которое статистически должно быть подтверждено тестом, формируется как альтернативная гипотеза.

Двусторонняя альтернативная гипотеза:

$H_0 \quad : \quad \beta_k = \beta_k^*$;

$H_A \quad : \quad \beta_k \neq \beta_k^*$.

Частный случай, когда регрессор не влияет на регрессанд $\beta_k^* = 0$.

$H_A \quad : \quad \beta_k \neq 0$.

Нулевая гипотеза

$H_0 \quad : \quad \beta_k = 0$.

Ошибки первого ряда : нулевая гипотеза отвергается, хотя она верна.

Ошибки второго ряда: нулевая гипотеза не отвергается, хотя она неверна.

Расчетная (или эмпирическая) величина t-статистики:

$$t = \frac{\hat{\beta}_k - \beta_k^*}{\hat{\sigma}_{\hat{\beta}_k}}. \tag{31}$$

Если нулевая гипотеза ($H_0 : \beta_k = \beta_k^*$) действительна, то для классической линейной модели рассчитанная по формуле (31) t- статистика является реализацией центральной t- распределенной случайной величины с Т-К степенями свободы.

Гипотеза H_0 отклоняется, если

$$|t| > t\,(\alpha, T - K),$$

где |t| - абсолютное значение t-статистики,

t (α, T-K) - табличное значение t- распределения (t- критерия) для уровня значимости α и Т-К степеней свободы.

Иначе

$$P\,(T \leq t\,(2\alpha, T - K)) = 1 - \alpha\,. \tag{32}$$

Эта величина называется (1-α) – квантиль t- распределения с (Т-К) степенями свободы.

α - является вероятностью ошибки первого рода, то есть нулевая гипотеза отвергается, хотя она верна.

t- тест односторонней пары гипотез представлен в двух вариантах:

I. H_0 : $\beta_k \le \beta_k^*$;

H_A : $\beta_k > \beta_k^*$. (33)

II. H_0 : $\beta_k \ge \beta_k^*$;

H_A : $\beta_k < \beta_k^*$. (34)

Области принятия гипотез имеют следующие границы:

для первого случая (33),

если $t \le t\ (2\alpha,\ T - K)$, принимается H_0;

если $t > t\ (2\alpha,\ T - K)$, отклоняется H_0;

для второго случая (34),

если $t \ge t\ (2\alpha,\ T - K)$, принимается H_0;

если $t < t\ (2\alpha,\ T - K)$, отклоняется H_0;

Схема проведения t- теста

Шаг 1: сформулировать пару гипотез H_0 и H_A.

Шаг 2: выбрать уровень значимости α.

Шаг 3 найти в таблицах t- критерий.

Шаг 4: рассчитать t – статистику.

Шаг 5: сравнить t- рассчетное с t – табличным.

Шаг 6: интерпретировать результат теста.

Чаще всего α=0,05 или α=0,01 (иногда его получают расчетным путем, используя компьютерные программы).

t - тест подтверждает предположение, что $\beta_к \ne \beta_к^*$

Если $\beta_к^*$=0, то t - статистика в области отклонения названной гипотезы свидетельствует о том, что к-й регрессор в генеральной совокупности оказывает влияние на регрессанд.

$\beta_к$ (для α=…) - статистически значимо отличается от нуля.

$\beta_к$ (для α=…) - статистически защищено от нуля.

Говорить βк - является значимым неточно, так как статистически незначимое может быть экономически значимым.

В области принятия гипотезы $H_0 : \beta_k = 0$ t- тест не позволяет отвергнуть нулевую гипотезу.

Доверительный интервал для регрессионного коэффициента β_k при доверительном уровне 1-α является интервалом со случайно зависимыми границами; он включает (накрывает) истинное значение k-го регрессионного коэффициента с вероятностью 1-α.

Формула доверительного интервала имеет вид:

$$P\left[\hat{\beta}_k - t(\alpha, T-K)\hat{\sigma}_{\hat{\beta}_k} \le \beta_k \le \hat{\beta}_k + t(\alpha, T-K)\hat{\sigma}_{\hat{\beta}_k}\right] = 1-\alpha \,. \qquad (35)$$

F - тест гипотез для групп регрессионных коэффициентов и линейных комбинаций

F - тест также применяется к линейной модели нормальной регрессии.

С помощью F - теста можно проверить только двусторонние гипотезы о значении нескольких коэффициентов или нескольких линейных комбинаций, а также сочетаний того и другого.

t - тест применим только для одного параметра или одной линейной комбинации.

Если применить F - тест для одного параметра, то он дает тот же результат, что и t - тест.

F - тестом статистически проверяется как единое целое (разовой проверкой):

а) двусторонняя гипотеза о значении одного, двух или нескольких регрессионных коэффициентов (t - тест только для одного коэффициента);

б) двусторонняя гипотеза о значении одной, двух или нескольких линейных комбинаций регрессионных коэффициентов;

в) совокупность гипотез о значении регрессионных коэффициентов и их линейных комбинаций.

Общая линейная гипотеза:

$H_0 : C\beta = C*;$ (36)

$H_A : C\beta \neq C*.$

C* - вектор столбец, состоящий из m элементов

$C_j^* \ (j = 1,2\ldots m)$

$\beta := (\beta_1 \ldots \beta_k)'$ - вектор-столбец К регрессионных коэффициентов, подлежащих оцениванию.

С-матрица размерности m x k; m определяет количество строк в матрице С, или m - количество линейных уравнений (линейных гипотез), проверяемых F - тестом.

Пример:

а) C=(010) c*=0

проверит гипотезу $H_0 : \beta_2=0$; $H_A : \beta_2\neq 0$;

б) C=(1 1 0); c*=1

проверит гипотезу $\beta_1+\beta_2=1$;

в) $c=\begin{bmatrix} 0\,10 \\ 1\,10 \end{bmatrix}$; $c* = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

проверит гипотезы $\beta_2=0$ и $\beta_1+\beta_2=1$.

Чтобы с помощью F - теста проверить гипотезы, необходимо вычислить F-статистику по формуле:

$$F = \frac{(C\hat{\beta} - C*)'\left[C(X'X)^{-1}C'\right]^{-1}(C\hat{\beta} - C*)}{m\hat{\sigma}_m^2}. \quad (37)$$

Формула (37) в ряде случаев значительно упрощается.

Проверим гипотезу о том, что ни один регрессор не оказывает влияние на регрессанд:

$H_0 : \beta_2 = \beta_3 = \ldots = \beta_K = 0;$

H_A : $\beta_k \neq 0$, хотя бы для одного k.

Рассчитываем F-статистику по формуле:

$$F = \frac{R^2}{1-R^2} \cdot \frac{T-K}{K-1}. \quad (38)$$

Правило применения F-тестов:

Нулевая гипотеза отклоняется, если

F > F (1-α; m; T-K),

где F – рассчитано по формуле (38);

F (1-α; m; T-K) – квантиль F – распределения (табличное значение или F - критерий).

m = K –1,

K – число регрессоров, включая свободный член;

T –K – число степеней свободы;

α - уровень значимости, который в экономических расчетах часто принимается 0,05; 0,01.

Может оказаться, что F-тест подтвердит статистическую значимость всей совокупности параметров β, в тоже время каждый параметр окажется статистически не значим.

Если F-тест проверяет один коэффициент или одну линейную комбинацию, тогда m = 1, рассчетные значения $F = t^2$, а табличные значения $[t(\alpha, T-K)]^2 = F(1-\alpha; 1; T-K)$, то есть результаты F -теста и t-теста совпадают.

Изложенный в данном методическом пособии материал является основой для изучения дальнейших тем курса «Эконометрия», связанных с обобщением классической регрессионной модели, методами оценки параметров в отдельных случаях несостоятельности классической модели, а также практическим применением моделей для экономического анализа, прогнозирования и управления.

Список рекомендуемой литературы по курсу «Эконометрия»

1. Аллен Р. Математическая эконометрия. –М.: Иностранная литература, 1963. – 606 с.
2. Винн Р., Холден К. Введение в прикладной эконометрический анализ. – М.: Финансы и статистика, 1981. –283 с.
3. Гранберг А.Г. Статистическое моделирование и прогнозирование. – М.: Финансы и статистика, 1990. – 378 с.
4. Дадаян В.С. Моделирование глобальных экономических процессов. – М.: Экономика, 1984. – 276 с.
5. Джонстон Д. Эконометрические методы. – М.: Статистика, 1965. – 361 с.
6. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. – М.: Финансы и статистика, 1986. – Т1 – 365., Т. II – 379 с.
7. Доугерти К. Введение в эконометрику: Пер. с англ. – М.: ИНФРА –М, 1999.- XIV, 402 с.
8. Єлейко В. Основи економетрії: У 2 ч. – Львів: ТзОВ «МАРКА Лтд», 1995. – 192 с., Ч. 1.
9. Кейн Э. Экономическая статистика и эконометрия. Вып. 1, 2 – М.: 1977.
10. Клас А., Гергели К., Колек Ю., Шуян И. Введение в эконометрическое моделирование. – М.: Статистика, 1978.
11. Клейнер Г.Б. Производные функции. – М.: Финансы и статистика, 1986. – 221 с.
12. Колек Ю., Шуян И. Эконометрические модели в социалистических странах. – М.: Экономика, 1978. – 151 с.
13. Лизер С. Эконометрические методы и задачи. – М.: 1971. – 247 с.
14. Ляшенко И.Н. Макромодели экономического роста. – К.: Вища школа. 1979. – 151 с.
15. Маленко Э. Лекции по микроэкономическому анализу. – М.: Наука, 1985. – 422 с.
16. Маленко Э. Статистические методы эконометрии. – М.: Статистика, 1975. – 321 с.
17. Мартинос С. Методические проблемы построения и применения эконометрических моделей. – Вильнюс: Максклас. 1979. – 170 с.
18. Пирогов Г., Федоровский Ю. Проблемы структурного оценивания в эконометрии. – М.: Статистика, 1979.
19. Райцин В.Я. Математические методы и модели планирования уровня жизни. – М.: Экономика, 1970. – 389 с.
20. Теория и практика статистического моделирования экономики. (Под редакцией Б.П. Суворова). – М.: Финансы и статистика, 1985. – 271 с.
21. Гинтнер Г. Введение в эконометрию. – М.: Статистика, 1965. – 361 с.
22. Шаттелес Т. Современные эконометрические методы. – М.: Статистика, 1975. – 161 с.
23. Гинтнер Г. Введение в эконометрию. – М.: Статистика, 1964.

24. Грубер И. Эконометрия I. Введение во множественную регрессию и эконометрию. 1995. (часть 1, 2).
Грубер И. Эконометрия 2. Введение во множественную регрессию и эконометрию. 1995. (часть 3, 4). Перевод г. Киев 1995 г., доцент, к.э.н. А.Б. Воронова.
25. Христиановский В.В., Москардини А, Гузь Н.Г., Лаулер К, Кривенчук О.Г. Прикладная эконометрия: Учебное издание / Донецк: Донецкий госуниверситет, 1998, 173 с.
26. Гузь Н.Г. Выбор и регулирование в микроэкономике. – Донецк: ИЭП НАН Украины, 1997. 195 с.