И.В. Лутошкин

Учебно-методический комплекс по курсу

ЭКОНОМЕТРИЧЕСКОЕ МОДЕЛИРОВАНИЕ
Часть 2

Ульяновск
Издательство Ульяновского государственного университета
2008

УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Институт экономики и бизнеса

Факультет финансов и учета

Для студентов специальности "Математические методы в экономике"

Оглавление

Введение

"*Существует две вещи, процесс изготовления которых лучше не видеть: сосиски и эконометрические оценки*" [8]. Эти слова Э. Лимера, с одной стороны, можно принимать с юмором, с другой, они достаточно четко отражают сущность эконометрики: качество получаемых оценок в определяющей мере зависит от квалификации исследователя, от целей, поставленных им, таким образом, эконометрические оценки в большой степени субъективны.

При этом стоит отметить, что некоторые специалисты из смежных областей (математики, экономики) не до конца четко понимают суть эконометрики. Наиболее точно объяснил ее сущность один из основателей Р.Фриш, который и ввел этот название в 1926 г.: "*Эконометрика - это не то же самое, что экономическая статистика. Она не идентична и тому, что мы называем экономической теорией, хотя значительная часть этой теории носит количественный характер. Эконометрика не является синонимом приложений математики к экономике. Как показывает опыт, каждая из трех отправных точек – статистика, экономическая теория и математика – необходимое, но не достаточное условие для понимания количественных соотношений в современной экономической жизни. Это единство всех трех составляющих. И это единство образует эконометрику*"[9].

Эконометрика – это самостоятельная научная дисциплина, объединяющая совокупность теоретических результатов, приемов, методов и моделей, предназначенных для того, чтобы на базе экономической теории, экономической статистики и экономических измерений, математико-статистического инструментария придавать конкретное количественное выражение общим (качественным) закономерностям, обусловленным экономической теорией.

Для описания сущности эконометрической модели удобно разбить весь процесс моделирования на шесть основных этапов:

1-й этап (постановочный) – определение конечных целей моделирования, набора участвующих в модели факторов и показателей, их роли;

2-й этап (априорный) – предмодельный анализ экономической сущности изучаемого явления, формирование и формализация априорной информации, в частности, относящейся к природе и генезису исходных статистических данных и случайных остаточных составляющих;

3-й этап (параметризация) – собственно моделирование, т.е. выбор общего вида модели, в том числе состава и формы входящих в нее связей;

4-й этап (информационный) – сбор необходимой статистической информации, т.е. регистрация значений участвующих в модели факторов и показателей на различных временных или пространственных тактах функционирования изучаемого явления;

5-й этап (идентификация модели) – статистический анализ модели и в первую очередь статистическое оценивание неизвестных параметров модели;

6-й этап (верификация модели) – сопоставление реальных и модельных данных, проверка адекватности модели, оценка точности модельных данных.

Эконометрическое моделирование реальных социально-экономических процессов и систем обычно преследует два типа конечных прикладных целей (или одну из них):

1) прогноз экономических и социально-экономических показателей, характе-

ризующих состояние и развитие анализируемой системы;

2) имитацию различных возможных сценариев социально-экономического развития анализируемой системы (многовариантные сценарные расчеты, имитационное моделирование).

При постановке задач эконометрического моделирования следует определить их иерархический уровень и профиль. Анализируемые задачи могут относиться к макро- (страна, межстрановой анализ), мезо- (регионы внутри страны) и микро- (предприятия, фирмы, семьи) уровням и быть направленными на решение вопросов различного профиля инвестиционной, финансовой или социальной политики, ценообразования, распределительных отношений и т.п.

Эконометрическое моделирование – это построение аналитических зависимостей для экономических показателей. Цель изучения – ознакомление с методами исследования, т.е. методами проверки, обоснования, оценивания количественных закономерностей и качественных утверждений (гипотез) в микро- и макроэкономике на основе анализа статистических данных. Задачи: научиться строить экономические модели и оценивать их параметры; научиться проверять гипотезы о свойствах экономических показателей и формах их связи; научиться использовать результаты экономического анализа для прогноза и принятия обоснования экономических решений.

Дисциплина "Эконометрическое моделирование" является одной из завершающих по учебному плану и не является базовой для других курсов. "Эконометрическое моделирование" некоторыми своими разделами смежна с такими дисциплинами, как "Экономико-математические модели и методы", "Эконометрика". Базовыми для курса "Эконометрическое моделирование" являются дисциплины экономического цикла, такие, как "Микроэкономика", "Макроэкономика". Математической основой курса являются дисциплины "Теория вероятностей", "Математическая статистика". Рекомендуется при изучении дисциплины "Эконометрическое моделирование" использовать примеры из предшествующих курсов, проводить заимствования и аналогии с ранее изученным, использовать приобретенные теоретические и практические знания для анализа реальных экономических ситуаций. Знания, приобретенные при изучении дисциплины "Эконометрическое моделирование", могут найти применение при выполнении творческих индивидуальных заданий, курсовом и дипломном проектировании.

1. Линейная регрессия среднего и МНК

В эконометрическом исследовании главной задачей является оценка зависимой переменной (объясняемого фактора) y от независмых (объясняющих факторов) x. Обычно исследователь, обладая совокупностью наблюдений $\{y_i, x_i\}$, $i = 1, 2, \ldots, N$, соответствующих факторов, хочет оценить $E[y|x]$. Существует несколько подходов к решению данной задачи: непараметрическое, параметрическое и полупараметрическое оценивания, подробнее [3].

В данной главе рассмотрим параметрический подход на основе линейной модели и выясним свойства оценок рассматриваемой модели.

1.1. Метод наименьших квадратов

Пусть $y \in \Re$, $x \in \Re^n$, предположим, что $E[y|x] = \langle x, \beta \rangle$, тогда регрессия среднего записывается как

$$y = \langle x, \beta \rangle + e, \qquad E[x|e] = 0,$$

здесь $e \in \Re$ играет роль случайной ошибки.

Если матрица $E[xx^T]$ невырожденная, то параметр β, минимизирующий среднеквадратичную ошибку, будет единственным решением задачи

$$\beta = \arg\min_b E[(y - \langle x, b \rangle)^2].$$

Предположим, что имеющиеся наблюдений $\{y_i, x_i\}$, $i = 1, 2, \ldots, N$, независимы и одинаково распределены, тогда, пользуясь принципом аналогий, можно построить оценку для β :

$$\hat{\beta} = \arg\min_b \frac{1}{N} \sum_{i=1}^{N} (y_i - \langle x_i, b \rangle)^2 = \left(\frac{1}{N} \sum_{i=1}^{N} x_i x_i^T \right)^{-1} \frac{1}{N} \sum_{i=1}^{N} x_i y_i.$$

Это и есть оценка метода наименьших квадратов, или МНК-оценка.

1.2. Асимптотические свойства МНК-оценки

Для выяснения асимптотических свойств МНК-оценки перепишем ее в следующем виде:

$$\hat{\beta} = \beta + \left(\frac{1}{N} \sum_{i=1}^{N} x_i x_i^T \right)^{-1} \frac{1}{N} \sum_{i=1}^{N} x_i e_i.$$

Покажем, что МНК-оценка состоятельна, т.е. $\hat{\beta} \xrightarrow{p} \beta$ (сходится по вероятности), и асимптотически распределена по нормальному закону (сходится по распределению):

$$\sqrt{N}(\hat{\beta} - \beta) = \left(\frac{1}{N} \sum_{i=1}^{N} x_i x_i^T \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i e_i \xrightarrow{d} \mathcal{N}\left(0, Q_{xx}^{-1} Q_{e^2xx} Q_{xx}^{-1}\right),$$

где $Q_{xx} = E[xx^T]$, $Q_{e^2xx} = E[e^2 xx^T] = Var[xe]$.

Приведенные асимптотические свойства следуют из закона больших чисел (ЗБЧ) и центральной предельной теоремы (ЦПТ). Из ЗБЧ следует, что

$$\frac{1}{N}\sum_{i=1}^{N} x_i x_i^T \xrightarrow{p} E[xx^T] = Q_{xx},$$

$$\frac{1}{N}\sum_{i=1}^{N} x_i e_i \xrightarrow{p} E[xe] = E[xE[e|x]] = 0,$$

что влечет состоятельность МНК-оценки. Из ЦПТ для независимых одинаково распределенных величин следует, что

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} x_i e_i \xrightarrow{d} \mathcal{N}(0, Var[xe]) = \mathcal{N}(0, Q_{e^2xx}),$$

что влечет асимптотическую нормальность МНК-оценки.

Рассмотрим специальный случай, когда регрессонная ошибка условно гомоскедастична, т.е. $E[e^2|x] = \sigma^2 = const$. В этом случае $Q_{e^2xx} = \sigma^2 Q_{xx}$ и асимптотическое распределение МНК-оценки имеет дисперсионную матрицу в упрощенном виде:

$$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \sigma^2 Q_{xx}^{-1}).$$

Кроме того, легко построить состоятельную оценку этой дисперсионной матрицы:

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \langle x_i, \hat{\beta}\rangle\right)^2 \xrightarrow{p} \sigma^2,$$

$$\hat{Q}_{xx} = \frac{1}{N}\sum_{i=1}^{N} x_i x_i^T \xrightarrow{p} Q_{xx}.$$

Состоятельность первой оценки следует из ЗБЧ. Состоятельность последней довольно легко показать:

$$\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \langle x_i, \hat{\beta}\rangle\right)^2 = |\text{ используем прием прибавить и отнять } \langle x_i, \beta\rangle| =$$
$$= \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \langle x_i, \beta\rangle\right)^2 + \frac{1}{N}\sum_{i=1}^{N}\left\langle x_i, \beta - \hat{\beta}\right\rangle^2 + \frac{2}{N}\sum_{i=1}^{N}\left(y_i - \langle x_i, \beta\rangle\right)\langle x_i, \beta - \hat{\beta}\rangle =$$
$$= \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \langle x_i, \beta\rangle\right)^2 + (\beta - \hat{\beta})^T\left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i^T\right)(\beta - \hat{\beta}) + \frac{2}{N}\sum_{i=1}^{N}\left(y_i - \langle x_i, \beta\rangle\right)\langle x_i, \beta - \hat{\beta}\rangle.$$

Далее, применяя ЗБЧ и теорему Слуцкого, получим:

$$\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \langle x_i, \beta\rangle\right)^2 \xrightarrow{p} \sigma^2,$$

$$(\beta - \hat{\beta})^T \left(\frac{1}{N} \sum_{i=1}^{N} x_i x_i^T \right) (\beta - \hat{\beta}) \xrightarrow{p} 0,$$

$$\frac{2}{N} \sum_{i=1}^{N} (y_i - \langle x_i, \beta \rangle) \langle x_i, \beta - \hat{\beta} \rangle \xrightarrow{p} 0.$$

Все вместе влечет состоятельность оценки условной дисперсии регрессионной ошибки:

$$\frac{1}{N} \sum_{i=1}^{N} \left(y_i - \langle x_i, \hat{\beta} \rangle \right)^2 \xrightarrow{p} \sigma^2.$$

Теперь рассмотрим общий случай условной гетероскедастичности. В этом случае нам нужно состоятельно оценить матрицу Q_{e^2xx}. Можно показать, что состоятельной оценкой этой матрицы будет следующая:

$$\hat{Q}_{e^2xx} = \frac{1}{N} \sum_{i=1}^{N} x_i x_i^T (y_i - \langle x_i, \hat{\beta} \rangle)^2 \xrightarrow{p} Q_{e^2xx}.$$

Итак, состоятельная оценка дисперсионной матрицы МНК-оценки в случае условной гетероскедастичности запишется как

$$\hat{V}_\beta = \hat{Q}_{xx}^{-1} \hat{Q}_{e^2xx} \hat{Q}_{xx}^{-1}.$$

Определение 1.1. Величину

$$se(\hat{\beta}_j) = \sqrt{\frac{1}{N} [\hat{V}_\beta]_{jj}}$$

будем называть *стандартной ошибкой* оценки $\hat{\beta}_j$.

Тогда соответствующая статистика t_j будет иметь асимптотически стандартное нормальное распределение:

$$t_j = \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \xrightarrow{d} \mathcal{N}(0, 1).$$

1.3. Свойства МНК-оценки в конечных выборках

Введем следующие обозначения:

$$X = (x_1, x_2, \ldots, x_N)^T, \quad Y = (y_1, y_2, \ldots, y_N)^T, \quad \varepsilon = (e_1, e_2, \ldots, e_N)^T.$$

Тогда уже знакомую нам регрессионную модель линейного условного среднего можно переписать в матричном виде

$$Y = X\beta + \varepsilon, \quad E[\varepsilon | X] = 0.$$

МНК-оценка в таком случае запишется как

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \beta + (X^T X)^{-1} X^T \varepsilon.$$

Эта оценка обладает следующими свойствами в конечных выборках:

– Условная несмещенность:

$$E[\hat{\beta}|X] = \beta + (X^TX)^{-1}X^TE[\varepsilon|X] = \beta.$$

– Безусловная несмещенность следует из условной несмещенности.

– Условная дисперсия оценки есть

$$Var[\hat{\beta}|X] = (X^TX)^{-1}X^T\Omega X(X^TX)^{-1},$$

где $\Omega = Var[Y|X] = E[\varepsilon\varepsilon^T|X]$.

1.4. Обобщенный метод наименьших квадратов

Определение 1.2. Пусть $E[Y|X] = X\beta$. Классом *линейных оценок* β называется класс, содержащий оценки вида $A(X)Y$, где $A(X)$ - матрица $n \times N$, которая зависит только от X.

К примеру, для МНК-оценки $A(X) = (X^TX)^{-1}X^T$.

Определение 1.3. Пусть $E[Y|X] = X\beta$. Классом *линейных несмещенных оценок* β называется класс, содержащий оценки вида $A(X)Y$, где $A(X)$ – матрица $n \times N$, зависящая только от X и удовлетворяющая условию $A(X)X = I_n$.

Опять же для МНК-оценки $A(X)X = (X^TX)^{-1}X^TX = I_n$.

Заметим, что $Var[A(X)Y|X] = A(X)\Omega A(X)^T$. Предположим, что мы хотим найти линейную несмещенную оценку, которая минимизирует $Var[A(X)Y|X]$.

Теорема 1 (Гаусса-Маркова). *Наилучшей линейной несмещенной оценкой линейной регрессии среднего является оценка* $\tilde{\beta} = A^*(X)Y$, *где*

$$A^*(X) = (X^T\Omega^{-1}X)^{-1}X^T\Omega^{-1}.$$

В этом случае дисперсионная матрица оценки имеет вид

$$Var[\tilde{\beta}|X] = (X^T\Omega^{-1}X)^{-1}.$$

Доказательство. Оценка $\tilde{\beta}$ принадлежит классу линейных несмещенных оценок, так как $A^*(X)X = I_n$. Возьмем произвольную матрицу $A(X)$, такую, что $A(X)X = I_n$. В этом случае имеют место следующие равенства:

$$\Big(A(X) - A^*(X)\Big)X = 0,$$

$$\Big(A(X) - A^*(X)\Big)\Omega A^*(X)^T = \Big(A(X) - A^*(X)\Big)\Omega\Omega^{-1}X(X^T\Omega^{-1}X)^{-1} = 0.$$

Тогда

$$Var[A(X)Y|X] = A(X)\Omega A(X)^T =$$

$$= \Big(A(X) - A^*(X) + A^*(X)\Big)\Omega\Big(A(X) - A^*(X) + A^*(X)\Big)^T =$$

$$= \Big(A(X) - A^*(X)\Big)\Omega\Big(A(X) - A^*(X)\Big)^T + Var[A^*(X)Y|X] \geq Var[\hat{\beta}|X].$$

Следовательно, оценка $\tilde{\beta}$ является наилучшей в классе линейных несмещенных оценок. ■

Определение 1.4. Пусть $E[Y|X] = X\beta$. Оценка $\tilde{\beta} = (X^T\Omega^{-1}X)^{-1}X^T\Omega^{-1}Y$ называется оценкой *обобщенного метода наименьших квадратов* (ОМНК).

Следствие 1. ОМНК-оценка $\tilde{\beta}$ является эффективной в классе линейных несмещенных оценок.

Следствие 2. Если ошибка линейной регрессии среднего обладает свойством условной гомоскедастичности, то $\tilde{\beta} = \hat{\beta}$, т.е. МНК- и ОМНК-оценки совпадают.

Ниже приведена таблица, содержащая условные дисперсионные матрицы МНК- и ОМНК-оценок в конечных выборках для случаев условной гетеро- и гомоскедастичности:

	МНК	ОМНК
Гомоскедастичность	$\sigma^2(X^TX)^{-1}$	$\sigma^2(X^TX)^{-1}$
Гетероскедастичность	$(X^TX)^{-1}X^T\Omega X(X^TX)^{-1}$	$(X^T\Omega^{-1}X)^{-1}$

Замечание. ОМНК-оценка $\tilde{\beta}$ является *недоступной*, так как матрица Ω неизвестна.

1.5. Асимптотические свойства ОМНК-оценок

Рассмотрим асимптотические свойства ОМНК-оценки. Для этого представим ее в следующем виде:

$$\tilde{\beta} = \beta + \left(\frac{1}{N}\sum_{i=1}^{N}\frac{x_ix_i^T}{\sigma^2(x_i)}\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\frac{x_ie_i}{\sigma^2(x_i)}.$$

Пользуясь законом больших чисел и центральной предельной теоремой, получим:

$$\frac{1}{N}\sum_{i=1}^{N}\frac{x_ix_i^T}{\sigma^2(x_i)} \xrightarrow{p} Q_{xx/\sigma^2} = E\left[\frac{xx^T}{\sigma^2(x)}\right],$$

$$\frac{1}{N}\sum_{i=1}^{N}\frac{x_ie_i}{\sigma^2(x_i)} \xrightarrow{p} E\left[\frac{xe}{\sigma^2(x)}\right] = 0,$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{x_ie_i}{\sigma^2(x_i)} \xrightarrow{d} N(0, Q_{xx/\sigma^2}).$$

Последнее выражение следует ввиду

$$E\left[\left(\frac{xe}{\sigma^2(x)}\right)^2\right] = E\left[\frac{xx^T}{\sigma^2(x)}E[e^2|x]\right] = Q_{xx/\sigma^2}.$$

Таким образом, ОМНК-оценка является состоятельной и асимптотически нормальной. Ниже приведена таблица, содержащая асимптотические дисперсионные матрицы МНК- и ОМНК-оценок для случаев условной гетеро- и гомоскедастичности:

	МНК	ОМНК
Гомоскедастичность	$\sigma^2 Q_{xx}^{-1}$	$Q_{xx/\sigma^2}^{-1} = \sigma^2 Q_{xx}^{-1}$
Гетероскедастичность	$Q_{xx}^{-1} Q_{e^2 xx} Q_{xx}^{-1}$	Q_{xx/σ^2}^{-1}

Теорема 2. *ОМНК-оценка* $\tilde{\beta}$ *асимптотически эффективна в классе оценок вида*

$$\hat{\beta}_z = \left(\frac{1}{N} \sum_{i=1}^{N} z_i x_i^T \right)^{-1} \frac{1}{N} \sum_{i=1}^{N} z_i y_i,$$

где $z_i = f(x_i), \quad f : \Re^n \longrightarrow \Re^n$.

Доказательство. Заметим, что МНК- и ОМНК-оценки принадлежат указанному классу, т.к. для МНК $z_i = x_i$, а для ОМНК $z_i = x_i/\sigma^2(x_i)$. Рассмотрим оценку

$$\hat{\beta}_z = \left(\frac{1}{N} \sum_{i=1}^{N} z_i x_i^T \right)^{-1} \frac{1}{N} \sum_{i=1}^{N} z_i y_i.$$

Легко показать, что она состоятельна и асимптотически нормальна с асимтотической дисперсионной матрицей

$$V_{zz} = Q_{zx}^{-1} Q_{e^2 zz} Q_{xz}^{-1},$$

где $Q_{zx} = E[zx^T]$, а $Q_{e^2 zz} = E[zz^T e^2] = E[zz^T \sigma^2(x)]$. Зная, что асимптотическая дисперсия ОМНК-оценки равна Q_{xx/σ^2}^{-1}, рассмотрим разность

$$V_{zz} - Q_{xx/\sigma^2}^{-1} = (E[zx^T])^{-1} E[zz^T \sigma^2(x)] (E[xz^T])^{-1} - \left(E\left[\frac{xx^T}{\sigma^2(x)} \right] \right)^{-1} =$$

$$= (E[vu^T])^{-1} E[vv^T] (E[uv^T])^{-1} - (E[uu^T])^{-1} =$$

$$= (E[vu^T])^{-1} \left[E[vv^T] - E[vu^T](E[uu^T])^{-1} E[uv^T] \right] (E[uv^T])^{-1} =$$

$$= (E[vu^T])^{-1} E[ww^T] (E[uv^T])^{-1} \geqslant 0.$$

Здесь $v = z\sigma(x), \quad u = x/\sigma(x)$ и $w = v - E[vu^T](E[uu^T])^{-1} u$. Таким образом, мы показали, что ОМНК-оценка асимптотически эффективна в указанном классе. ■

1.6. Доступная ОМНК-оценка

Чтобы получить ОМНК-оценку, необходимо знать дисперсионную матрицу ошибок Ω, на главной диагонали которой находятся величины $\sigma^2(x_i)$, а на остальных местах стоят нули. Естественно полагать, что эти параметры являются неизвестными априори, поэтому они должны быть оценены. Обычно предполагают, что дисперсия ошибок есть линейная функция от некоторого преобразования x :

$$\sigma^2(x) = E[e^2|x] = z^T \gamma,$$

где $z = f(x), \quad f : \Re^n \to \Re^n$. Если предположение правильное, то можно оценить *скедастичную регрессию*

$$e = z^T\gamma + \varepsilon, \quad E[\varepsilon|z] = 0.$$

Таким образом, мы получаем состоятельные оценки дисперсий ошибок:

$$\hat{\sigma}^2(x_i) = z_i^T\hat{\gamma},$$

после чего можно построить доступную оценку обобщенного метода наименьших квадратов (ДОМНК):

$$\tilde{\beta}_F = \left(\sum_{i=1}^N \frac{x_i x_i^T}{\hat{\sigma}^2(x_i)}\right)^{-1} \sum_{i=1}^N \frac{x_i Y_i}{\hat{\sigma}^2(x_i)} = (X^T\hat{\Omega}^{-1}X)^{-1}X^T\hat{\Omega}^{-1}Y.$$

Алгоритм построения ДОМНК-оценки:

1. Используя МНК, оценить исходную регрессию и получить остатки $\hat{e}_i$ для $i = 1, \ldots, N$. Прогнать скедастичную регрессию, получить оценки $\hat{\gamma}$ и построить оценки дисперсий ошибок $\hat{\sigma^2}(x_i)$ (или $\hat{\Omega}$).

2. Построить ДОМНК-оценку

$$\tilde{\beta}_F = \left(\sum_{i=1}^N \frac{x_i x_i^T}{\hat{\sigma}^2(x_i)}\right)^{-1} \sum_{i=1}^N \frac{x_i Y_i}{\hat{\sigma}^2(x_i)} = (X^T\hat{\Omega}^{-1}X)^{-1}X^T\hat{\Omega}^{-1}Y.$$

Такой алгоритм построения оценок дисперсии ошибок не гарантирует их положительность, приведем некоторые способы избежать условия $\hat{\sigma^2}(x_i) < 0$.

1. Выбрать некоторое малое $\delta > 0$. Положить $\hat{\sigma}^2(x_i) = \max(z_i^T\hat{\gamma}, \delta)$.

2. Выбросить те наблюдения, для которых $\hat{\sigma}^2(x_i) < 0$.

3. Положить

$$\hat{\sigma}^2(x_i) = \frac{1}{N}\sum_{j=1}^N z_j^T\hat{\gamma}$$

для тех наблюдений, для которых $\sigma^2(x_i) < 0$.

Стоит отметить, что если скедастичная функция правильно специфицирована, то ДОМНК- оценка $\tilde{\beta}_F$ асимптотически эквивалентна ОМНК-оценке $\tilde{\beta}$, т.е.

$$\sqrt{N}(\tilde{\beta}_F - \beta) \xrightarrow{d} \mathcal{N}(0, Q^{-1}_{xx/\sigma^2}).$$

Состоятельная оценка асимптотической дисперсии в этом случае есть

$$\hat{V}_\beta = N\left(\sum_{i=1}^N \frac{x_i x_i^T}{\hat{\sigma}^2(x_i)}\right)^{-1}.$$

Если скедастичная функция специфицирована неправильно, то оценка $\tilde{\beta}_F$, тем не менее, остается состоятельной и асимптотически нормальной:

$$\sqrt{N}(\tilde{\beta}_F - \beta) \xrightarrow{d} \mathcal{N}(0, Q^{-1}_{xx/\sigma^2}Q_{xx/\sigma^4e^2}Q^{-1}_{xx/\sigma^2}),$$

где использованы следующие обозначения:

$$Q_{xx/\sigma^2} = E\left[\frac{xx^T}{z^T\gamma}\right], \qquad Q_{xx/\sigma^4 e^2} = E\left[\frac{xx^T}{(z^T\gamma)^2}e^2\right].$$

Состоятельная оценка асимптотической дисперсии в этом случае равна

$$\hat{V}_\beta = N\left(\sum_i \frac{x_i x_i^T}{z_i^T\hat{\gamma}}\right)^{-1} \sum_i \frac{x_i x_i^T}{(z_i^T\hat{\gamma})^2}\hat{e}_i^2 \left(\sum_i \frac{x_i x_i^T}{z_i^T\hat{\gamma}}\right)^{-1}.$$

2. Метод максимального правдоподобия в эконометрии

2.1. Базовые понятия

Пусть Y – реализация N -мерной случайной величины, имеющей плотность $P_\theta(x)$ (в случае непрерывного распределения) или вероятность $p_\theta(x)$ (в случае дискретного распределения).

Здесь $P_\theta(x)$ $(p_\theta(x))$ характеризует семейство распределений задаваемое параметром $\theta \in \Theta$, $\Theta \subset \Re^m$ – пространство параметров. Будем считать, что рассматриваемый вектор наблюдений (выборка) порожден распределением из этого семейства с параметром $\theta_0 \in \Theta$, которое будем называть истинным распределением, а θ_0 – истинным параметром.

Определение 2.5. Функция $\mathcal{L}(Y,\theta) = P_\theta(Y)$ (для дискретного случая $\mathcal{L}(Y,\theta) = \rho_\theta(Y)$) называется *функцией правдоподобия*. *Оценкой максимального правдоподобия* (МП) $\hat\theta$ называется решение задачи $\hat\theta = \hat\theta(Y) = \underset{\theta\in\Theta}{\operatorname{argmax}}\,\mathcal{L}(Y,\theta)$.

В большинстве прикладных задач решение поставленной оптимизационной задачи единственно. Вообще, данный метод оценивания распределения генеральной совокупности называют *методом максимального правдоподобия*. На практике чаще используется *логарифмическая функция правдоподобия*

$$\ell(Y,\theta) = \ln\left(\mathcal{L}(Y,\theta)\right).$$

В силу того, что логарифм – строго монотонно возрастающая функция, то $\underset{\theta\in\Theta}{\operatorname{argmax}}\,\mathcal{L}(Y,\theta) = \underset{\theta\in\Theta}{\operatorname{argmax}}\,\ell(Y,\theta)$, что позволяет искать $\widehat{\theta}$ максимизацией логарифмической функции правдоподобия.

В частном случае вектор наблюдений представляет собой выборку независимых одинаково распределенных случайных величин: $Y_i \sim IID$, $i = 1,\ldots,N$. При этом

$$\mathcal{L}(Y,\theta) = \prod_{i=1}^{N}\mathcal{L}_i(Y_i,\theta), \qquad \ell(Y,\theta) = \sum_{i=1}^{N}\ell_i(Y_i,\theta).$$

Вообще говоря вектор наблюдений Y состоит из зависимых между собой и/или неодинаково распределенных случайных величин, поэтому не является выборкой в обычном смысле слова. В общем случае это равенство тоже будет верным если обозначить

$$\mathcal{L}_i(Y_i,\theta) = P_\theta(Y_i|Y_{i-1},\ldots,Y_1) \quad \text{и} \quad \ell_i(Y_i,\theta) = \ln\left(\mathcal{L}_i(Y_i,\theta)\right).$$

Тем самым задается разбиение функции правдоподобия на вклады отдельных наблюдений.

Поскольку Y – случайная величина, то функция правдоподобия – случайная величина при данном значении параметров. Оценка максимального правдоподобия является функцией вектора наблюдений: $\hat\theta = \hat\theta(Y)$, поэтому это тоже случайная величина. Соответственно, точно так же случайными величинами является значение функции правдоподобия в максимуме $\hat{\mathcal{L}}(Y) = \hat{\mathcal{L}}(Y,\hat\theta)$ и многие другие рассматриваемые далее величины (градиент, гессиан и т. п.).

Пусть функция правдоподобия дифференцируема по θ и достигает максимума во внутренней точке $\hat{\theta} \in int(\Theta)$, тогда оценка МП $\hat{\theta}$, согласно теореме Ферма, должна удовлетворять условию первого порядка:

$$\frac{\partial \mathcal{L}}{\partial \theta}\left(Y, \hat{\theta}\right) = 0 \quad \text{или} \quad \frac{\partial \ell}{\partial \theta}\left(Y, \hat{\theta}\right) = 0.$$

Полученная система уравнений называется *уравнениями правдоподобия.*

Таким образом, градиент логарифмической функции правдоподобия $g(Y,\theta) = \dfrac{\partial \ell}{\partial \theta}(Y,\theta)$ при $\theta = \hat{\theta}$ должен быть равен нулю. (Будем предполагать, что $g(Y,\theta)$ – столбец.)

Для того, чтобы оценки, удовлетворяющие уравнениям правдоподобия, действительно давали максимум правдоподобия, достаточно, чтобы были выполнены условия второго порядка (предполагаем, что функция правдоподобия дважды дифференцируема). А именно, матрица Гессе (гессиан) логарифмической функции правдоподобия должна быть всюду отрицательно определена. Матрица Гессе $\mathcal{H}$ по определению есть матрица вторых производных:

$$\mathcal{H}_{ij}(Y,\theta) = \frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j}(Y,\theta), \quad i,j = 1,\ldots,m.$$

В некоторых моделях функция правдоподобия неограничена сверху и не существует оценок максимального правдоподобия в смысле приведенного выше определения. Согласно альтернативному определению оценками максимального правдоподобия называют корни уравнения правдоподобия, являющиеся локальными максимумами функции правдоподобия, корнями уравнения правдоподобия. Существуют модели, для которых такие оценки состоятельны.

Определение 2.6.Информационной матрицей для вектора наблюдений размерностью N будем называть матрицу

$$\mathcal{I}(\theta) = \mathcal{I}^N(\theta) = E_\theta\left(g(Y,\theta)g^T(Y,\theta)\right).$$

Заметим, что по этому определению информационная матрица – функция некоторого вектора параметров $\theta \in \Theta$. Индекс θ у символа математического ожидания E означает, что ожидание вычисляется в предположении, что θ – точка истинных значений параметров.

В дальнейшем будет использоваться следующее очевидное свойство функции правдоподобия. Пусть $\varphi(Y)$ есть некоторая функция вектора наблюдений Y, тогда ее математическое ожидание равно

$$E(\varphi(Y)) = \int\limits_{\mathcal{Y}} \varphi(Y) L(\theta_0, Y) dY,$$

где $\mathcal{Y}$ обозначает пространство элементарных событий (пространство переменной Y).

Таким образом, можно переписать определение информационной матрицы в виде

$$\mathcal{I}(\theta) = \int_{\mathcal{Y}} g(Y,\theta) g^T(Y,\theta)\, \mathcal{L}(Y,\theta)\, dY.$$

Определение 2.7. Асимптотической информационной матрицей называется предел

$$\mathcal{I}^{\infty}(\theta) = \lim_{N\to\infty} \frac{1}{N} \mathcal{I}^N(\theta).$$

Заметим, что информационная матрица $\mathcal{I}^N(\theta)$ является величиной порядка $O(N)$, поэтому множитель $1/N$ добавлен в определение для существования конечного предела.

Рассмотрим выборку, состоящую из независимых и одинаково распределеных наблюдений, тогда, применяя определение информационной матрицы к отдельным наблюдениям $(\mathcal{I}_i)$, имеем

$$\mathcal{I}^N = N\mathcal{I}_i.$$

В этом случае можно заключить, что информация растет пропорционально количеству наблюдений.

Линейная регрессия с нормально распределенными ошибками

Пусть ошибки $\varepsilon_i \sim NID(0,\sigma^2)$. Эта аббревиатура означает, что случайные величины ε_i независимы и имеют нормальное распределение с параметрами $(0,\sigma^2)$ (normally and independently distributed). Ковариационная матрица вектора ошибок – это единичная матрица с точностью до множителя: $E(\varepsilon\varepsilon^T) = \sigma^2 I_N$.

Зависимая переменная связана с ошибками соотношением:

$$Y = X\beta + \varepsilon,$$

где X – матрица регрессоров $(N \times m)$, β – вектор-столбец неизвестных коэффициентов длины m. Таким образом, Y_i имеет нормальное распределение с параметрами $(\langle x_i, \beta\rangle, \sigma^2)$, где x_i – i -я строка матрицы X :

$$Y_i \sim \mathcal{N}(\langle x_i, \beta\rangle, \sigma^2).$$

Напомним, что плотность нормального распределения $\mathcal{N}(a,\sigma^2)$ определяется формулой

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right).$$

Функция правдоподобия для набора наблюдений (Y_i, x_i), $i = 1,2,\ldots,N$ имеет вид

$$\mathcal{L} = (2\pi\sigma^2)^{-N/2} \prod_{i=1}^{N} \exp\left(-\frac{(Y_i - \langle x_i,\beta\rangle)^2}{2\sigma^2}\right).$$

Логарифмическая функция правдоподобия:

$$\ell = -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(Y_i - \langle x_i, \beta\rangle)^2 =$$
$$= -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(Y - X\beta)^T(Y - X\beta).$$

Если обозначить через e вектор остатков $e = Y - X\beta$, то можно получить другое представление логарифмической функции

$$\ell = -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}e^T e.$$

В данном случае вектор неизвестных параметров состоит из двух компонент:

$$\theta = \begin{pmatrix} \beta \\ \sigma^2 \end{pmatrix}.$$

Соответственно, градиент логарифмической функции правдоподобия тоже состоит из двух частей:

$$g_\beta = \frac{\partial \ell}{\partial \beta} = \frac{1}{\sigma^2}X^T(Y - X\beta) = \frac{1}{\sigma^2}X^T e.$$

$$g_{\sigma^2} = \frac{\partial \ell}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{e^T e}{2\sigma^4} = \frac{1}{2\sigma^4}\left(\langle e, e\rangle - N\sigma^2\right).$$

Оценка максимального правдоподобия $\hat\theta$ должна удовлетворять равенству $g(\hat\theta) = 0$, откуда получим

$$\hat\beta = \left(X^T X\right)^{-1} X^T Y \quad \text{и} \quad \hat\sigma^2 = \frac{\langle \hat e, \hat e\rangle}{N} = \frac{\langle Y - X\hat\beta, Y - X\hat\beta\rangle}{N}.$$

Из полученного видно, что в случае линейной регрессии ММП дает ту же оценку вектора коэффициентов регрессии β, что и МНК. Как известно, оценка дисперсии $\hat\sigma^2$ является смещенной:

$$E(\hat\sigma^2) = \frac{N - m}{N}\sigma^2.$$

Покажем, каким образом связаны ММП и МНК. Для этого выразим, используя равенство $g_{\sigma^2} = 0$, дисперсию через β:

$$\sigma^2(\beta) = \frac{\langle Y - X\beta, Y - X\beta\rangle}{N}.$$

Если подставить ее в функцию правдоподобия, то получится *концентрированная функция правдоподобия*:

$$\ell^c = -\frac{N}{2}\ln(2\pi\sigma^2(\beta)) - \frac{\langle e, e\rangle}{2\sigma^2(\beta)} = -\frac{N}{2}\ln\left(2\pi\frac{\langle Y - X\beta, Y - X\beta\rangle}{N}\right) - \frac{N}{2}.$$

Максимизация концентрированной функции правдоподобия эквивалентна минимизации суммы квадратов остатков $\mathrm{RSS}(\beta) = \langle Y - X\beta, Y - X\beta\rangle$ по β.

Гессиан логарифмической функции правдоподобия состоит из следующих компонент:

$$H_{\beta\beta} = \frac{\partial^2 \ell}{\partial\beta\partial\beta^T} = -\frac{1}{\sigma^2}X^T X, \quad H_{\beta\sigma^2} = \frac{\partial^2 \ell}{\partial\beta\partial\sigma^2} = -\frac{1}{\sigma^4}e^T X,$$

$$H_{\sigma^2\beta} = \frac{\partial^2 \ell}{\partial\sigma^2\partial\beta^{\mathrm{T}}} = -\frac{1}{\sigma^4}X^T e, \quad H_{\sigma^2\sigma^2} = \frac{\partial^2 \ell}{\left(\partial\sigma^2\right)^2} = \frac{N}{2\sigma^4} - \frac{e^T e}{\sigma^6}.$$

В точке истинных значений параметров θ_0 остатки совпадают с погрешностями $e = \varepsilon$. Используя это, получим, что компоненты информационной матрицы, вычисленной в точке θ_0 равны:

$$\begin{aligned}
\mathcal{I}_{\beta\beta}(\theta_0) &= E\left(g_\beta(\theta_0)\, g_\beta(\theta_0)^T\right) = E\left(\frac{1}{\sigma^4}X^T\varepsilon\varepsilon^T X\right) = \frac{1}{\sigma^4}X^T E\left(\varepsilon\varepsilon^T\right)X = \\
&= \frac{1}{\sigma^4}\sigma^2 X^T I X = \frac{1}{\sigma^2}X^T X, \\
\mathcal{I}_{\beta\sigma^2}(\theta_0) &= E(g_{\sigma^2}(\theta_0)g_\beta(\theta_0)^T) = E\left(\frac{1}{\sigma^6}(\varepsilon^T\varepsilon - N\sigma^2)\varepsilon^T X\right) = 0^T, \\
\mathcal{I}_{\sigma^2\beta}(\theta_0) &= 0 \quad \text{(аналогично)}, \\
\mathcal{I}_{\sigma^2\sigma^2}(\theta_0) &= E\left(g_{\sigma^2}(\theta_0)^2\right) = E\left(\frac{1}{4\sigma^8}\left(\varepsilon^T\varepsilon - N\sigma^2\right)^2\right) = \\
&= \frac{1}{4\sigma^8}E\left(\left(\varepsilon^T\varepsilon\right)^2\right) - 2N\frac{1}{4\sigma^6}E\left(\varepsilon^T\varepsilon\right) + N^2\frac{1}{4\sigma^4} = \\
&= \frac{1}{4\sigma^8}\left(\sum_{i=1}^{N}E(\varepsilon_i^4) + 2\sum_{i=1}^{N}\sum_{j=i+1}^{N}E(\varepsilon_i\varepsilon_j)\right) - \frac{N}{2\sigma^6}\sum_{i=1}^{N}E(\varepsilon_i^2) + \frac{N^2}{4\sigma^4} = \\
&= \frac{3N\sigma^4 + (N^2 - N)\,\sigma^4}{4\sigma^8} - \frac{N}{2\sigma^6}N\sigma^2 + \frac{N^2}{4\sigma^4} = \frac{N}{2\sigma^4}.
\end{aligned}$$

2.2. Характеристика ММП.

В статистике применяются три основных метода оценивания:

- Метод наименьших квадратов.
- (Обобщенный) метод моментов.
- Метод максимального правдоподобия.

Интересно сравнить ММП с двумя другими методами. Условия, при которых можно использовать ММП более ограничительны, так как метод требует явного задания вида распределения. С другой стороны, ММП более универсален. Его можно использовать для любых моделей, задающих вид распределения наблюдаемых переменных. Два другие метода можно использовать лишь тогда, когда распределение переменных можно представить в определенном виде. Если есть гипотеза о точном виде распределения, то всегда понятно, как получать оценки параметров, распределений параметров и различных статистик, как проверять гипотезы, хотя сами расчеты могут быть сложными.

Еще одно свойство – *инвариантность* по отношению к переобозначению параметров. Пусть $\varphi(\cdot) : r^k \to r^k$ – однозначная обратимая функция, тогда можно подставить в функцию правдоподобия вместо θ величину $\varphi(\tau)$, где τ – новый вектор параметров, $\tau \in \varphi^{-1}(\theta)$. При этом, если $\hat{\tau}$ – оценка МП в новой задаче, то $\hat{\theta}$ – оценка МП в старой задаче. Из инвариантности следует, что оценка МП, как правило, несмещенна.

Если правильно выбрать параметризацию, то распределение оценок в малых выборках может быть близко к асимптотическому, если неправильно, то асимптотическое распределение будет очень плохой аппроксимацией. ММП получил широкое распространение благодаря своим хорошим асимптотическим свойствам:

- состоятельность,
- асимптотическая нормальность,
- асимптотическая эффективность.

С точки зрения эффективности сильные предположения о виде распределения, которые приходится делать, применяя ММП, окупаются (в большей или меньшей степени), так как мы делаем очень ограничительные предположения, то можем доказать более сильные утверждения.

2.3. Связь гессиана и матрицы вкладов в градиент с информационной матрицей

2.3.1. Гессиан и информационная матрица

Покажем, какая связь существует между информационной матрицей и гессианом. Сделаем это только в случае непрерывного распределения. Тот же метод доказательства очевидным образом распространяется на дискретные распределения. Применяя правило дифференцирования логарифма к логарифмической функции правдоподобия, получим следующее тождество:

$$\frac{\partial \ell}{\partial \theta} = \frac{1}{\mathcal{L}} \frac{\partial \mathcal{L}}{\partial \theta}.$$

Продифференцируем по θ^T :

$$\frac{\partial^2 \ell}{\partial \theta \partial \theta^T} = \frac{1}{\mathcal{L}} \frac{\partial^2 \mathcal{L}}{\partial \theta \partial \theta^T} - \frac{1}{\mathcal{L}^2} \frac{\partial \mathcal{L}}{\partial \theta^T} \frac{\partial \mathcal{L}}{\partial \theta}.$$

Отсюда, воспользовавшись правилом дифференцирования логарифма, получим

$$\mathcal{H} = \frac{\partial^2 \ell}{\partial \theta \partial \theta^T} = \frac{1}{\mathcal{L}} \frac{\partial^2 \mathcal{L}}{\partial \theta \partial \theta^T} - \frac{\partial \ell}{\partial \theta^T} \frac{\partial \ell}{\partial \theta}.$$

Найдем теперь математическое ожидание обеих частей в точке θ_0 (при истинных параметрах распределения):

$$E(\mathcal{H}(Y, \theta_0)) = E\left(\frac{\partial^2 \ell}{\partial \theta \partial \theta^T}(\theta_0)\right) = \int\limits_{\mathcal{Y}} \mathcal{L}(\theta_0, Y) \frac{1}{\mathcal{L}(\theta_0, Y)} \frac{\partial^2 \mathcal{L}(\theta_0, Y)}{\partial \theta \partial \theta^T} dY - E\left(\frac{\partial \ell(\theta_0)}{\partial \theta^T} \frac{\partial \ell(\theta_0)}{\partial \theta}\right).$$

Второй член разности есть по определению информационная матрица $\mathcal{I}(\theta_0)$, что касается первого слагаемого, то интеграл равен нулю при условии, что операции интегрирования и дифференцирования перестановочны (для этого достаточно, в частности, чтобы пространство зависимой переменной $\mathcal{Y}$ не зависело от θ, или плотность распределения по краям $\mathcal{Y}$ стремилась к нулю):

$$\frac{\partial^2}{\partial \theta \partial \theta^T} \int\limits_{\mathcal{Y}} \mathcal{L}(\theta_0, Y) dY = \frac{\partial^2}{\partial \theta \partial \theta^T} 1 = 0.$$

Таким образом,

$$-E\left(\mathcal{H}(Y,\theta_0)\right)=\mathcal{I}(\theta_0),$$

т.е. информационная матрица равна математическому ожиданию гессиана функции правдоподобия со знаком минус. То же самое свойство верно асимптотически:

$$-\lim_{N\to\infty}\frac{1}{N}E\left(\mathcal{H}(Y,\theta_0)\right)=\mathcal{I}^{\infty}(\theta_0).$$

2.3.2. Матрица вкладов в градиент и информационная матрица

Прежде всего докажем, что математическое ожидание градиента в точке (в предположении, что операции интегрирования и дифференцирования перестановочны) θ_0 равно нулю $(E(g(Y,\theta_0))=0)$:

$$E\left(g(Y,\theta_0)\right)=\int\limits_{\mathcal{Y}}g(Y,\theta_0)\mathcal{L}(Y,\theta_0)dY=\int\limits_{\mathcal{Y}}\frac{\partial\ell}{\partial\theta}(Y,\theta_0)\mathcal{L}(Y,\theta_0)dY=$$

$$=\int\limits_{\mathcal{Y}}\frac{1}{\mathcal{L}(Y,\theta_0)}\frac{\partial\mathcal{L}}{\partial\theta}(Y,\theta_0)\mathcal{L}(Y,\theta_0)dY=\int\limits_{\mathcal{Y}}\frac{\partial\mathcal{L}}{\partial\theta}(Y,\theta_0)dY=$$

$$=\frac{\partial}{\partial\theta}\int\limits_{\mathcal{Y}}\mathcal{L}(Y,\theta_0)dY=\frac{\partial}{\partial\theta}1=0.$$

Как уже говорилось, функцию правдоподобия можно разбить по вкладам отдельных наблюдений: $\ell(Y,\theta)=\sum\limits_{i=1}^{N}\ell_i(Y_i,\theta)$. Очевидно, что свойство аддитивности сохраняется для градиента.

Определим *матрицу вкладов в градиент* отдельных наблюдений G как

$$G_{ij}(Y,\theta)=\frac{\partial\ell_i}{\partial\theta_j}\left(Y,\theta\right).$$

При этом

$$\sum_{i=1}^{N}G_{ij}\left(Y,\theta\right)=\sum_{i=1}^{N}\frac{\partial\ell_i}{\partial\theta_j}\left(Y,\theta\right)=\frac{\partial}{\partial\theta_j}\sum_{i=1}^{N}\ell_i\left(Y,\theta\right)=\frac{\partial\ell}{\partial\theta_j}\left(Y,\theta\right)=g_j\left(Y,\theta\right).$$

Используя рассуждения, аналогичные приведенным выше, можно показать, что $E\left(G_{ij}(Y,\theta_0)\right)=0$. Пусть $G_i(Y,\theta_0)$ – строка матрицы $G^0=(Y,\theta_0)$, относящаяся к наблюдению i, тогда, если $i\neq j$, то $E\left(G_i(Y,\theta_0)G_j(Y,\theta_0)^T\right)=0$. Так как элементы матрицы G^0 имеют нулевое математическое ожидание, то это означает что строки матрицы G^0, относящиеся к разным наблюдениям, некоррелированы.

Действительно, функция правдоподобия i -го наблюдения по определению есть плотность распределения Y_i (в случае непрерывного распределения) условная по информации, содержащейся в наблюдениях $1,\ldots,i-1$ (условная по $Y_1,\ldots,Y_{i-1}$). Обозначим это информационное множество Ω_i. Будем вычислять математическое ожидание по частям – сначала условное, а потом от него безусловное (правило полного математического ожидания). Предположим, что $i<j$, тогда

$$E\left(G_i(Y,\theta_0)G_j(Y,\theta_0)^T\right) = E\left(E\left(G_i(Y,\theta_0)G_j(Y,\theta_0)^T|\Omega_i\right)\right) =$$
$$= E\left(G_i(Y,\theta_0)E\left(G_j(Y,\theta_0)^T|\Omega_i\right)\right) = 0.$$

Равенство $E\left(G_j(Y,\theta_0)^T|\Omega_i\right) = 0$ доказывается по той же схеме, что и доказанное выше $E(g(Y,\theta_0)) = 0$. Используя это свойство, получим

$$E\left(G^{0T}G^0\right) = E\left(\sum_{i=1}^N G_i^{0T}G_i^0\right) = E\left(\left(\sum_{i=1}^N G_i^0\right)^T\left(\sum_{i=1}^N G_i^0\right)\right) = E\left(g(Y,\theta_0)g^T(Y,\theta_0)\right).$$

Последнее выражение есть по определению информационная матрица. Таким образом,

$$E\left(G^{0T}G^0\right) = \mathcal{I}^N(\theta_0).$$

2.3.3. Вычисление информационной матрицы

Рассмотрим теперь, как вычислить для конкретной модели информационную матрицу $\mathcal{I}^N(\theta)$. Здесь существуют три способа. Очевидно, что все три способа должны для "хороших" моделей давать один и тот же результат.

Во-первых, можно воспользоваться определением информационной матрицы: $\mathcal{I}(\theta_0) = E(g(Y,\theta_0)g^T(Y,\theta_0))$, во-вторых, можно воспользоваться равенством $\mathcal{I}(\theta_0) = -E(H(Y,\theta_0))$. Самым простым часто (когда функцию правдоподобия можно простым образом разбить на вклады наблюдений) оказывается третий способ, который использует только что рассмотренное свойство

$$\mathcal{I}(\theta_0) = \mathcal{I}^N(\theta_0) = E\left(G^{0T}G^0\right) = E\left(\sum_{i=1}^N G_i^{0T}G_i^0\right).$$

Выше было получено выражение для информационной матрицы в случае линейной регрессии с нормально распределенными ошибками прямо по определению. Вычислим теперь ее двумя другими способами.

Гессиан уже был вычислен выше. Математическое ожидание от него со знаком минус равно.

$$\mathcal{I}(\theta_0) = -E(H(Y,\theta_0)) = -E\left(\begin{bmatrix} -\frac{1}{\sigma^2}X^TX & -\frac{1}{\sigma^4}X^T\varepsilon \\ -\frac{1}{\sigma^4}\varepsilon^TX & \frac{N}{2\sigma^4} - \varepsilon^T\varepsilon\frac{1}{\sigma^6} \end{bmatrix}\right) = \begin{bmatrix} X^TX\frac{1}{\sigma^2} & 0 \\ 0^T & \frac{N}{2\sigma^4} \end{bmatrix}.$$

Вклад в логарифмическую функцию правдоподобия i-го наблюдения равен

$$\ell_i = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(Y_i - x_i\beta)^2.$$

Продифференцировав его, получим вклад в градиент i-го наблюдения в точке истинных параметров:

$$G_i^0 = \left(\frac{1}{\sigma^2} x_i \varepsilon_i, \frac{\varepsilon_i^2}{2\sigma^4} - \frac{1}{2\sigma^2} \right).$$

Вклад в информационную матрицу i-го наблюдения в точке истинных параметров равен

$$\mathcal{I}_i(\theta_0) = E\left(G_i^{0T} G_i^0 \right) = \begin{bmatrix} \frac{1}{\sigma^2} X_i^T X_i & 0 \\ 0^T & \frac{1}{2\sigma^4} \end{bmatrix}.$$

Таким образом,

$$\mathcal{I}(\theta_0) = \sum_{i=1}^{N} \mathcal{I}_i(\theta_0) = \begin{bmatrix} \frac{X^T X}{\sigma^2} & 0 \\ 0^T & \frac{N}{2\sigma^4} \end{bmatrix}.$$

Все три способа, как и следовало ожидать, привели к одному и тому же результату.

Заметим, что $\mathcal{I}_i(\theta_0)$ – положительно определенная матрица, $\mathcal{I}(\theta_0)$ при любом количестве наблюдений – положительно определенная матрица (в предположении, что матрица регрессоров имеет полный ранг). Из этого можно сделать вывод, что информация в некотором смысле увеличивается с ростом количества наблюдений. Это одно из объяснений названия "информационная матрица". В частности, определитель информационной матрицы увеличивается с ростом количества наблюдений:

$$|\mathcal{I}^{N+1}(\theta_0)| > |\mathcal{I}^N(\theta_0)|.$$

2.4. Распределение градиента и оценок максимального правдоподобия

Оценки максимального правдоподобия имеют нормальное асимптотическое распределение. Для доказательства этого воспользуемся предположением, что градиент функции правдоподобия в точке истинных значений параметров θ_0 имеет асимптотически нормальное распределение.

Градиент $g(Y, \theta_0)$ будет иметь нормальное распределение (асимптотически), если к нему применима *центральная предельная теорема*. Надо представить $g(Y, \theta_0)$ как сумму некоторой последовательности случайных величин. Для этого подходит разложение градиента на вклады отдельных наблюдений

$$g_j(Y, \theta_0) = \sum_{i=1}^{N} G_{ij}(Y, \theta_0).$$

Как сказано выше, каждое из слагаемых здесь имеет нулевое математическое ожидание. Если выполнены условия регулярности, то $\sum\limits_{i=1}^{N} G_{ij}(Y, \theta_0)$ стремится к нормальному распределению с ростом количества наблюдений. Ковариационная

матрица градиента в точке θ_0 есть информационная матрица, поскольку его математическое ожидание равно нулю: $V(g(Y,\theta_0)) = E(g(Y,\theta_0)g^T(Y,\theta_0)) = \mathcal{I}^N(\theta_0)$. Последнее равенство выполнено по определению.

Окончательно получаем

$$\frac{1}{\sqrt{N}} g(Y,\theta_0) \overset{a}{\sim} \mathcal{N}\left(0, \mathcal{I}^\infty(\theta_0)\right)$$

Используя это свойство градиента можно доказать асимптотическую нормальность оценок ММП. Для этого используем разложение в ряд Тейлора в точке θ_0 до членов первого порядка:

$0 = g(\hat{\theta}) = g(\theta_0) + \mathcal{H}(\bar{\theta})(\hat{\theta} - \theta_0)$,

где $\mathcal{H}$ – гессиан (матрица вторых производных от логарифмической функции правдоподобия), $\bar{\theta}$ – выпуклая комбинация $\hat{\theta}$ и θ_0 . Так как $\hat{\theta}$ – состоятельная оценка параметра θ_0 , то $\bar{\theta}$ тоже должна быть состоятельной оценкой θ_0. В силу того, что $-\frac{1}{N}\mathcal{H}(\theta_0) \overset{a}{=} \mathcal{I}^\infty(\theta_0)$, то имеем асимптотическое равенство: $-\frac{1}{N}\mathcal{H}(\bar{\theta}) \overset{a}{=} \mathcal{I}^\infty(\theta_0)$. Таким образом,

$$\sqrt{N}\left(\hat{\theta} - \theta_0\right) \overset{a}{=} \left(\mathcal{I}^\infty(\theta_0)\right)^{-1} \frac{1}{\sqrt{N}} g(\theta_0) \overset{a}{\sim} \mathcal{N}\left(0, \left(\mathcal{I}^\infty(\theta_0)\right)^{-1} \mathcal{I}^\infty(\theta_0) \left(\mathcal{I}^\infty(\theta_0)\right)^{-1}\right)$$

Окончательно получим

$$\sqrt{N}\left(\hat{\theta} - \theta_0\right) \overset{a}{\sim} \mathcal{N}\left(0, \left(\mathcal{I}^\infty(\theta_0)\right)^{-1}\right).$$

Это соотношение позволяет оценить ковариационную матрицу оценок $\hat{\theta}$. С этой точки зрения оценка обратной информационной матрицы является оценкой ковариационной матрицы МП-оценок $\hat{V}_\theta$ (с точностью до множителя $\frac{1}{N}$), и эти термины можно использовать как синонимы. Понятно, что для этого должны быть выполнены соответствующие условия, гарантирующие, что операции интегрирования и дифференцирования коммутируют и что справедлива центральная предельная теорема.

2.5. Численные методы нахождения оценок максимального правдоподобия

Рассмотрим семейство алгоритмов вычисления оценок максимального правдоподобия. Эти алгоритмы являются итеративными градиентными методами

$$\theta^{k+1} = \theta^k + (\hat{\mathcal{I}}^k)^{-1} g(\theta^k), \quad k = 0, 1, \ldots.$$

Стационарная точка этого процесса $\theta^{k+1} = \theta^k$ будет удовлетворять уравнениям правдоподобия $g = 0$ и (с соответствующими оговорками) будет оценкой максимального правдоподобия.

Если в качестве $\hat{\mathcal{I}}^k$ взять информационную матрицу в точке оценок $\mathcal{I}^N(\theta^k)$, то мы получаем метод, называемый по-английски *method of scoring*:

$$\theta^{k+1} = \theta^k + (\mathcal{I}^N(\theta^k))^{-1} g(\theta^k).$$

Если в качестве $\hat{\mathcal{I}}^k$ взять минус гессиан $-H(\theta^k)$, то мы получаем классический *метод Ньютона*:

$$\theta^{k+1} = \theta^k - (H(\theta^k))^{-1} g(\theta^k).$$

Метод Ньютона, как правило, быстрее сходится в ближайшей окрестности оценок МП, зато метод, использующий информационную матрицу обычно менее чувствителен к выбору начальных приближений.

Особой простотой, и потому притягательностью (требуется найти только первые производные), отличается третий способ оценивания информационной матрицы, использующий матрицу вкладов в градиент G. Этот способ предложен в статье Berndt, Hall, Hall, and Hausman (1974) и поэтому называется BHHH. Другое название - метод внешнего произведения градиента (outer product of the gradient, сокращенно OPG). Этот способ основан на том, что $\mathcal{I}(\theta_0) = E\left(G^{0T} G^0\right)$, таким образом, в качестве $\hat{\mathcal{I}}^k$ можно взять $G(Y,\hat{\theta})^T G(Y,\hat{\theta})$.

Шаг метода BHHH (OPG) можно получить с помощью вспомогательной (искусственной) регрессии, зависимой переменной в которой будет вектор, составленный из единиц (обозначим его $\mathbf{1}$), а матрицей регрессоров – матрица $G(\theta^k)$. Если $\Delta\theta^k$ – оценки коэффициентов в этой вспомогательной регрессии на k -м шаге, то итерация имеет вид

$$\theta^{k+1} = \theta^k + \Delta\theta^k, \quad \text{где} \quad \Delta\theta^k = \left(G\left(\theta^k\right)^T G\left(\theta^k\right)\right)^{-1} G\left(\theta^k\right)^T 1.$$

Хотя этот последний алгоритм является самым простым, но, как правило, сходится очень медленно. Если учесть, что обычно при использовании этого метода $\left(G\left(\theta^k\right)^T G\left(\theta^k\right)\right)^{-1}$ берут в качестве оценки ковариационной матрицы оценок, то использовать его нежелательно.

Возможны различные модификации этой основной идеи. Шаг алгоритма можно вычислять, варьируя длину исходного шага параметром λ :

$$\theta^{k+1} = \theta^k + \lambda(\hat{\mathcal{I}}^k)^{-1} g(\theta^k).$$

Разумно выбирать параметр λ, максимизируя по нему функцию правдоподобия в точке θ :

$$\lambda = arg \max_{\lambda} \ell(\theta^k + \lambda(\hat{\mathcal{I}}^k)^{-1} g(\theta^k)).$$

В частном случае матрица $\mathcal{I}^N(\theta_0)$ является блочно-диагональной. Тогда шаг алгоритма можно разбить на несколько “подшагов”, один для каждого блока. Изменяются при этом только параметры, соответствующие данному блоку.

Если из условий первого порядка выразить одни оцениваемые параметры через другие и подставить их в функцию правдоподобия, то получится *концентрированная функция правдоподобия*. Действуя таким образом, задачу поиска оценок МП можно упростить, сведя к задаче максимизации концентрированной функции правдоподобия по меньшему числу параметров. Задача может упроститься до одномерного поиска.

Существует много других алгоритмов. Есть алгоритмы специально сконструированные для конкретной модели. Есть универсальные методы, которые можно применять к широкому классу моделей, такие как метод удвоенной регрессии и итеративный обобщенный МНК. Можно, конечно, использовать универсальные опти-

мизационные алгоритмы, которые подходят не только для максимизации функции правдоподобия.

2.6. Асимптотическое распределение и асимптотическая эквивалентность трех классических статистик

Предположим, что мы хотим проверить гипотезу о том, что вектор истинных параметров θ_0 удовлетворяет набору ограничений, который в векторном виде можно записать как $r(\theta_0) = 0$.

Тогда с учетом этой информации задача получения оценки максимального правдоподобия эквивалентна задаче нахождения точки локального максимума функции Лагранжа (при условии, что решение регулярно):

$$L(\theta, \lambda) = \ell(\theta) - r^T(\theta)\lambda.$$

Ограниченная оценка $\tilde{\theta}$ должна вместе с вектором множителей Лагранжа $\tilde{\lambda}$ удовлетворять следующей системе условий первого порядка:

$$\begin{cases} g(\tilde{\theta}) = R^T(\tilde{\theta})\tilde{\lambda}, \\ r(\tilde{\theta}) = 0, \end{cases}$$

где $R(\theta)$ – матрица первых производных ограничений: $R = \dfrac{\partial r(\theta)}{\partial \theta}$.

Для вывода распределений интересующих нас статистик используем тот же прием, с помощью которого выше получено распределение оценок. Поскольку мы предполагаем, что оценки МП состоятельны и нас интересуют асимптотические распределение, то для разложений в ряд Тейлора будем писать приближенные равенства. Разложим градиент и ограничения в ряд Тейлора до членов первого порядка в точке истинных параметров θ_0 :

$g(\tilde{\theta}) \approx g(\theta_0) + H(\theta_0)(\tilde{\theta} - \theta_0)$,
$r(\tilde{\theta}) \approx R(\theta_0)(\tilde{\theta} - \theta_0)$.

При получении второго соотношения мы использовали, что в точке истинных параметров ограничения выполняются: $r(\theta_0) = 0$.

Подставив эти приближения в условия первого порядка, получим следующие асимптотические равенства:

$g(\theta_0) + H(\theta_0)(\tilde{\theta} - \theta_0) \overset{a}{=} R^T(\tilde{\theta})\tilde{\lambda}$,
$R(\theta_0)(\tilde{\theta} - \theta_0) \overset{a}{=} 0$.

Перепишем систему в блочной форме:

$$\begin{pmatrix} -H(\theta_0) & R^T(\theta_0) \\ R(\theta_0) & 0 \end{pmatrix} \begin{pmatrix} \tilde{\theta} - \theta_0 \\ \tilde{\lambda} \end{pmatrix} \overset{a}{=} \begin{pmatrix} g(\theta_0) \\ 0 \end{pmatrix}.$$

Домножая соответствующие строки на $1/\sqrt{N}$, чтобы получились величины порядка $O(1)$, получим асимптотическое равенство:

$$\begin{pmatrix} \sqrt{N}\left(\tilde{\theta} - \theta_0\right) \\ \dfrac{1}{\sqrt{N}}\tilde{\lambda} \end{pmatrix} \overset{a}{=} \begin{pmatrix} \mathcal{I}^{\infty}(\theta_0) & R^T(\theta_0) \\ R(\theta_0) & 0 \end{pmatrix}^{-1} \begin{pmatrix} \dfrac{1}{\sqrt{N}} g(\theta_0) \\ 0 \end{pmatrix}.$$

Можно показать, что решение полученной системы записывается через выражения, асимптотически эквивалентные оценкам $\tilde{\theta}$ и множителям Лагранжа $\tilde{\lambda}$:

$$\begin{aligned}\sqrt{N}\left(\tilde{\theta}-\theta_0\right) &\stackrel{a}{=} \left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1}\times\\ &\times\left(I-R^T(\theta_0)\left(R(\theta_0)\left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1}R^T(\theta_0)\right)^{-1}R(\theta_0)\left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1}\right)\left(\frac{1}{\sqrt{N}}g(\theta_0)\right),\\ \frac{1}{\sqrt{N}}\tilde{\lambda} &\stackrel{a}{=} \left(R(\theta_0)\left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1}R(\theta_0)^T\right)^{-1}R(\theta_0)\left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1}\left(\frac{1}{\sqrt{N}}g(\theta_0)\right).\end{aligned}$$

Вспомним, что $\frac{1}{\sqrt{N}}g(\theta_0)\stackrel{a}{\sim}\mathcal{N}\left(0,\mathcal{I}^{\infty}(\theta_0)\right)$, отсюда получим асимптотическое распределение вектора множителей Лагранжа:

$$\begin{aligned}\frac{1}{\sqrt{N}}\tilde{\lambda} &\stackrel{a}{\sim}\mathcal{N}\left(0,\left(R(\theta_0)\left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1}R^T(\theta_0)\right)^{-1}\right)\times\\ &\times\left(R(\theta_0)\left(I^{\infty}(\theta_0)\right)^{-1}\mathcal{I}^{\infty}(\theta_0)\left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1}R(\theta_0)^T\left(R(\theta_0)\left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1}R(\theta_0)^T\right)^{-1}\right)\\ \frac{1}{\sqrt{N}}\tilde{\lambda} &\stackrel{a}{\sim} N\left(0,\left(R(\theta_0)\left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1}R^T(\theta_0)\right)^{-1}\right).\end{aligned}$$

Определение 2.8. *Статистикой множителя Лагранжа* называют следующую величину:

$$LM\equiv\tilde{\lambda}^TR(\tilde{\theta})\tilde{\mathcal{I}}^{-1}R^T(\tilde{\theta})\tilde{\lambda}$$

Здесь $\tilde{\mathcal{I}}$ – матрица, полученная на основании выборочной информации в точке $\tilde{\theta}$, такая что $\frac{1}{N}\tilde{\mathcal{I}}$ – состоятельная оценка $\mathcal{I}^{\infty}(\theta_0)$. Величина LM имеет распределение χ^2 с p степенями свободы, где p – размерность вектора ограничений $r(\theta)$:

$$LM\stackrel{a}{\sim}\chi^2(p).$$

Это следует из формулы для распределения $\tilde{\lambda}$, состоятельности оценки $\tilde{\theta}$ и невырожденности матрицы $R(\theta_0)\left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1}R^T(\theta_0)$.

Вспомним, что одно из условий первого порядка максимума функции правдоподобия имеет вид $g(\tilde{\theta})=R^T(\tilde{\theta})\tilde{\lambda}$. Это позволяет выразить статистику множителя Лагранжа через градиент логарифмической функции правдоподобия:

$$LM=g^T(\tilde{\theta})\tilde{\mathcal{I}}^{-1}g(\tilde{\theta})\stackrel{a}{\sim}\chi^2(p).$$

Хотя статистика множителя Лагранжа получила свое название благодаря тому, что ее можно выразить через множители Лагранжа, на практике гораздо чаще используют градиентную форму (score form of LM test).

Если вспомнить асимптотическое выражение для $\tilde{\lambda}$, то можно выразить (асимптотически) LM-тест через $g(\theta_0)$:

$$LM\stackrel{a}{=}\frac{1}{N}g^T(\theta_0)\left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1}R^T(\theta_0)\left(R(\theta_0)\left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1}R^T(\theta_0)\right)^{-1}R(\theta_0)\left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1}g(\theta_0).$$

Определение 2.9. *Статистика отношения правдоподобия* по определению есть

$$LR\equiv 2\left(\ell(\tilde{\theta})-\ell(\hat{\theta})\right).$$

Найдем ее асимптотическое распределение. Используем для этого разложение в ряд Тейлора:

$$\ell(\tilde{\theta}) = \ell(\hat{\theta}) + \frac{1}{2}\left(\tilde{\theta} - \hat{\theta}\right)^T \mathcal{H}\left(\bar{\theta}\right)\left(\tilde{\theta} - \hat{\theta}\right),$$

где $\hat{\theta}$ – выпуклая линейная комбинация $\tilde{\theta}$ и $\hat{\theta}$. Поскольку $\tilde{\theta}$ и $\hat{\theta}$ – состоятельные оценки θ_0, то $-\mathcal{H}\left(\bar{\theta}\right) \overset{a}{=} N\mathcal{I}^{\infty}(\theta_0)$. В этом случае

$$LR = 2\left(\ell(\tilde{\theta}) - \ell(\hat{\theta})\right) \overset{a}{=} N\left(\tilde{\theta} - \hat{\theta}\right)^T \mathcal{I}^{\infty}(\theta_0)\left(\tilde{\theta} - \hat{\theta}\right).$$

Асимптотический эквивалент этой статистики также можно записать в терминах θ_0, $\mathcal{I}^{\infty}(\theta_0)$, $R(\theta_0)$ и $g(\theta_0)$.

Отняв от $\sqrt{N}\left(\tilde{\theta} - \theta_0\right) \overset{a}{=} \left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1}\left(\frac{1}{\sqrt{N}} g(\theta_0)\right)$ доказанное ранее равенство об асимптотическом разложении $\sqrt{N}\left(\tilde{\theta} - \theta_0\right)$ получаем, что

$$\sqrt{N}\left(\tilde{\theta} - \theta_0\right) \overset{a}{=} \left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1} R^T(\theta_0)\left(R(\theta_0)\left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1} R^T(\theta_0)\right)^{-1} R(\theta_0)\left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1} g(\theta_0).$$

Отсюда следует, что статистика отношения правдоподобия асимптотически равна той же самой случайной величине, что и статистика множителя Лагранжа:

$$LR \overset{a}{=} \frac{1}{N} g^T(\theta_0)(\mathrm{I}(\theta_0)^{\infty})^{-1} R^T(\theta_0)\left(R(\theta_0)\left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1} R^T(\theta_0)\right)^{-1} R(\theta_0)\left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1} g(\theta_0).$$

Это означает, что статистика отношения правдоподобия также имеет асимптотическое распределение χ^2 с p степенями свободы.

Третья классическая статистика основана на распределении $r(\hat{\theta})$. Так как $\hat{\theta}$ – оценка, полученная без учета ограничений, то в общем случае $r(\hat{\theta}) \neq 0$, однако, если верна нулевая гипотеза, то $r(\hat{\theta}) \overset{a}{=} 0$. Разложим $r(\hat{\theta})$ в ряд Тейлора в окрестности точки θ_0, учитывая, что $r(\theta_0) = 0$:

$$r(\hat{\theta}) \approx r(\theta_0) + R(\theta_0)(\hat{\theta} - \theta_0) = R(\theta_0)(\hat{\theta} - \theta_0).$$

Ранее было выведено, что $\sqrt{N}(\hat{\theta} - \theta_0) \overset{a}{\sim} \mathcal{N}\left(0, \left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1}\right)$.

Определение 2.10. Определим *статистику Вальда*:

$$W \equiv r^T(\hat{\theta})\left(R(\hat{\theta})\mathcal{I}^{-1}(\hat{\theta})R^T(\hat{\theta})\right)^{-1} r(\hat{\theta}).$$

Как и в случае двух других тестов $W \overset{a}{\sim} \chi^2(p)$. В пределе

$$r(\hat{\theta}) \overset{a}{=} R(\theta_0)\left(\tilde{\theta} - \theta_0\right) \overset{a}{=} \frac{1}{N} R(\theta_0)\left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1} g(\theta_0).$$

Значит,

$$LR \overset{a}{=} \frac{1}{N} g^T(\theta_0)\left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1} R^T(\theta_0)\left(R(\theta_0)\left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1} R^T(\theta_0)\right)^{-1} R(\theta_0)\left(\mathcal{I}^{\infty}(\theta_0)\right)^{-1} g(\theta_0).$$

Тем самым мы показали, что с ростом количества наблюдений все три статистики стремятся к одной и той же случайной переменной, которая имеет распределение $\chi^2(p)$. Другими словами, три классических теста **асимптотически эквивалентны**.

Все три статистики совпадают, если логарифмическая функция правдоподобия является квадратичной. Это верно, в частности, для линейной регрессии с известной дисперсией, например,

$$Y = X\beta + \varepsilon, \quad \text{где} \quad \varepsilon_i \sim NID(0,1).$$

Рассмотрим, к примеру, логарифмическую функцию правдоподобия с единственным (скалярным) параметром θ:

$$\ell(\theta) = -(a - b\theta)^2 + const.$$

Гессиан $\mathcal{H}(\theta) = \dfrac{\partial^2 \ell}{\partial \theta^2} = -2b^2$ является постоянной величиной, информационная матрица, таким образом, равна $\mathcal{I} = 2b^2$ при всех θ.

Возьмем ограничение вида $r(\theta) = 3\theta - 1$, тогда получим $\hat{\theta} = \dfrac{a}{b}, \quad \tilde{\theta} = \dfrac{1}{3}$, откуда $\ell(\hat{\theta}) = 0 + const, \quad \ell(\tilde{\theta}) = -\left(a - \dfrac{b}{3}\right)^2 + const.$

$$LR = 2\left(\ell(\hat{\theta}) - \ell(\tilde{\theta})\right) = 2\left(a - \frac{b}{3}\right)^2.$$

Градиент равен $g(\theta) = 2b(a - b\theta)$, таким образом, $g(\tilde{\theta}) = 2b\left(a - \dfrac{1}{3}b\right)$.

$$LM = g^T(\tilde{\theta})\mathcal{I}^{-1}(\tilde{\theta})g(\tilde{\theta}) = 2b\left(a - \frac{1}{3}b\right)\frac{1}{2b^2}2b\left(a - \frac{1}{3}b\right) = 2\left(a - \frac{b}{3}\right)^2.$$

Найдем ту же статистику через множитель Лагранжа

$$L(\theta) = -(a - b\theta)^2 - \lambda(3\theta - 1) \to \max_{\theta}.$$

По теореме Лагранжа

$$\frac{\partial L}{\partial \theta} = 2(a - b\theta) - 3\lambda = 0, \Rightarrow \tilde{\lambda} = \frac{2}{3}\left(a - b\tilde{\theta}\right) = \frac{2}{3}\left(a - \frac{b}{3}\right).$$

Заметим, что

$$R(\theta) = \frac{\partial r}{\partial \theta} = 3, \quad \Rightarrow \quad R(\tilde{\theta}) = 3.$$

$$LM = \tilde{\lambda}^T R(\tilde{\theta})\mathcal{I}^{-1}(\tilde{\theta})R^T(\tilde{\theta})\tilde{\lambda} = \frac{2}{3}\left(a - \frac{b}{3}\right)3\frac{1}{2b^2}3\frac{2}{3}\left(a - \frac{b}{3}\right) = 2\left(a - \frac{b}{3}\right)^2.$$

Теперь найдем статистику Вальда, вначале определим следующее

$$r(\hat{\theta}) = 3\hat{\theta} - 1 = 3\frac{a}{b} - 1, \quad R(\hat{\theta}) = 3.$$

$$W = r^T(\hat{\theta})\left(R(\hat{\theta})\mathcal{I}^{-1}(\hat{\theta})R^T(\hat{\theta})\right)^{-1}r(\hat{\theta}) = \left(3\frac{a}{b} - 1\right)\left(3\frac{1}{2b^2}3\right)^{-1}\left(3\frac{a}{b} - 1\right) = 2\left(a - \frac{b}{3}\right)^2.$$

2.6.1. Соотношения между статистиками.

Все три классические статистики совпадают в случае "бесконечно большой выборки". В выборках конечных размеров их поведение может существенно отличаться от асимптотического. Поэтому не всегда на эти классические статистики можно полагаться. В этом их отличие от F -статистик, которые имеют *точное* распределение в конечной выборке в случае классической линейной регрессии с нормальными ошибками и линейной проверяемой гипотезой. Рассмотрим этот случай более подробно.

Предположим, что проверяется гипотеза $Q\beta = q$, где Q – известная матрица $(p \times m)$, q – известный вектор $(p \times 1)$. В использованных выше обозначениях $r(\theta) = r(\beta, \sigma^2) = Q\beta - q$, матрица $R(\theta)$ равна $(Q\ 0)$ при всех значениях θ (нулевой вектор относится к параметру σ^2). Проверяемой гипотезе соответствует следующая статистика Вальда (вспомним, что $\mathcal{I}_{\beta\beta}^{-1} = \sigma^2 \left(X^T X\right)^{-1}$) :

$$W = r^T(\hat{\theta}) \left(R(\hat{\theta})\mathcal{I}^{-1}(\hat{\theta})R^T(\hat{\theta})\right)^{-1} r(\hat{\theta}) = \left(Q\hat{\beta} - q\right)^T \left(Q\mathcal{I}_{\beta\beta}^{-1}Q^T\right)^{-1} \left(Q\hat{\beta} - q\right) =$$
$$= \frac{1}{\hat{\sigma}^2} \left(Q\hat{\beta} - q\right)^T \left(Q\left(X^T X\right)^{-1} Q^T\right)^{-1} \left(Q\hat{\beta} - q\right).$$

Нам нужно максимизировать функцию правдоподобия при ограничении $Q\beta = q$. Функция Лагранжа рассматриваемой задачи условной максимизации имеет вид

$$L(\theta, \lambda) = -\frac{N}{2} \ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2} (Y - X\beta)^T (Y - X\beta) - (Q\beta - q)^T \lambda.$$

В максимуме должно выполняться

$$\frac{\partial L}{\partial \beta^T}\left(\tilde{\theta}, \tilde{\lambda}\right) = \frac{1}{\tilde{\sigma}^2} X^T \left(Y - X\tilde{\beta}\right) - Q^T\tilde{\lambda} = 0.$$

Отсюда

$$\tilde{\beta} = \left(X^T X\right)^{-1} X^T Y - \tilde{\sigma}^2 \left(X^T X\right)^{-1} Q^T\tilde{\lambda} = \hat{\beta} - \tilde{\sigma}^2 \left(X^T X\right)^{-1} Q^T\tilde{\lambda},$$

где $\tilde{\beta} = \left(X^T X\right)^{-1} X^T Y$ – оценка ОМНК (без ограничений). Домножая это равенство слева на Q и учитывая, что $Q\tilde{\beta} = q$, получим

$$\tilde{\sigma}^2\tilde{\lambda} = \left(Q\left(X^T X\right)^{-1} Q^T\right)^{-1} \left(Q\hat{\beta} - q\right).$$

Таким образом, оценки с учетом ограничений равны

$$\tilde{\beta} = \hat{\beta} - \left(X^T X\right)^{-1} Q^T \left(Q\left(X^T X\right)^{-1} Q^T\right)^{-1} \left(Q\hat{\beta} - q\right).$$

Из условия $\frac{\partial L}{\partial \beta^T}\left(\tilde{\theta}, \tilde{\lambda}\right) = 0$ несложно получить, что $\tilde{\sigma}^2 = \frac{\widetilde{RSS}}{N}$, где $\widetilde{RSS}$ – сумма квадратов остатков в регрессии с ограничениями (так же как $\hat{\sigma}^2 = \frac{\widehat{RSS}}{N}$ в регрессии без ограничений).

Статистика множителя Лагранжа равна:

$$LM = \tilde{\lambda}^T R(\tilde{\theta})\mathcal{I}^{-1}(\tilde{\theta})R^T(\tilde{\theta})\tilde{\lambda} = \tilde{\sigma}^2\tilde{\lambda}^T Q\left(X^T X\right)^{-1}Q^T\tilde{\lambda} = \\ = \frac{1}{\tilde{\sigma}^2}\left(Q\hat{\beta}-q\right)^T\left(Q\left(X^T X\right)^{-1}Q^T\right)^{-1}\left(Q\hat{\beta}-q\right).$$

Можно также показать, что

$$\left(Q\beta-\hat{\beta}\right)^T\left(Q\left(X^T X\right)^{-1}Q^T\right)^{-1}\left(Q\hat{\beta}-q\right) = \widehat{\widetilde{RSS}} - \widehat{RSS}.$$

Это позволяет выразить LM и W через суммы квадратов остатков:

$$LM = N\frac{\widehat{\widetilde{RSS}} - \widehat{RSS}}{\widehat{\widetilde{RSS}}}, \quad W = N\frac{\widehat{\widetilde{RSS}} - \widehat{RSS}}{\widehat{RSS}}.$$

Логарифмическая функция правдоподобия в максимуме равна

$\ell(\tilde{\theta}) = -\frac{N}{2}\ln\left(2\pi\frac{\widehat{\widetilde{RSS}}}{N}\right) - \frac{N}{2}$, в регрессии без ограничений

$$\ell(\hat{\theta}) = -\frac{N}{2}\ln\left(2\pi\frac{\widehat{RSS}}{N}\right) - \frac{N}{2}.$$

Отсюда найдем статистику отношения правдоподобия:

$$LR = 2\left(\ell(\tilde{\theta}) - \ell(\hat{\theta})\right) = N\left(ln\left(\widehat{\widetilde{RSS}}\right) - ln\left(\widehat{RSS}\right)\right).$$

Так как логарифм – строго вогнутая функция, то выполнено следующее точное неравенство:

$$W > LR > LM.$$

Таким образом, тест Вальда будет чаще отвергать гипотезу, тест множителя Лагранжа – реже. Это же неравенство верно и для нелинейных регрессий.

F -статистика для проверки той же гипотезы равна

$$F = \frac{\widehat{\widetilde{RSS}} - \widehat{RSS}}{\widehat{RSS}}\frac{N-m}{p}.$$

Она распределена как $F(p, N-m)$.

В линейной регрессии лучше, конечно использовать $t-$ и $F-$ статистики. Кроме того, распределение этих статистик лучше аппроксимируется их номинальным распределением и в других моделях: нелинейных регрессиях, некоторых моделях, являющихся развитием регрессионных, некоторых искусственных регрессиях и т. п. Хотя здесь $t-$ и $F-$ статистики не будут иметь точного распределения, но, как показывают эксперименты, они, как правило, лучше в конечных выборках, чем их асимптотические аналоги ($\mathcal{N}$ и χ^2 соответственно). Такие $t-$ и $F-$ статистики называют **асимптотическими** $t-$ **и** $F-$ **статистистиками**. Три классические статистики можно преобразовать в асимптотические $F-$ статистики по следующим формулам:

$$F_W = \frac{W}{N}\frac{N-m}{p}, \quad F_{LM} = \frac{LM}{N-LM}\frac{N-m}{p}, \quad F_{LR} = \left(\exp\left(\frac{LR}{N}\right)-1\right)\frac{N-m}{p}.$$

Все эти статистики распределены приближенно как $F(p, N-m)$.

Понятно, что тесты на основе W, LR, LM и асимптотические $F-$ тесты дают противоречивые результаты в конечных выборках. Одни из них могут отвергать гипотезу при выбранном уровне значимости, другие же говорить в пользу принятия гипотезы.

Для того, чтобы исследовать поведение асимптотических статистик в конечных выборках, используют метод Монте-Карло. С помощью этого метода можно, в частности, выяснить, какой из тестов более подходит для данного типа моделей, какую оценку ковариационной матрицы оценок лучше всего использовать.

3. Форма регрессионной модели

В большинстве случаев на первом этапе моделирования используется линейная функциональная форма регрессионной модели

$$Y = \langle X, \beta \rangle + e, \quad Y, e \in \Re,\ X, \beta \in \Re^n.$$

при этом предполагается для наблюдаемой статистики $\{Y_i, X_i\}$, $Y_i \in \Re,\ X_i \in \Re^n$, $i = 1, 2, \ldots, N$ выполнение равенств

$$Y_i = \langle X_i, \beta \rangle + \varepsilon_i, \quad \varepsilon_i \in \Re.$$

В силу простоты интерпретации линейной модели дальнейшее изменение функциональной формы не производится. Однако, необходимость изменения функциональной формы модели возникает, если неверна одна из следующих гипотез, выполнение которых требуется для того, чтобы обычный метод наименьших квадратов (ОМНК) в применении к регрессионной модели давал хорошие результаты:

1. Ошибки имеют нулевое математическое ожидание, или, что то же самое, математическое ожидание зависимой переменной является линейной комбинацией регрессоров:
$$E(\varepsilon_i) = 0,\ E(Y_i) = X_i\beta,\ (i = 1, 2, \ldots, N).$$

2. Ошибки гомоскедастичны, т. е. имеют одинаковую дисперсию для всех наблюдений:
$$E(\varepsilon_i^2) = \sigma^2,\ (i = 1, 2, \ldots, N).$$

3.1. Тестирование правильности спецификации регрессионной модели

Если ошибка имеет ненулевое математическое ожидание, то оценки ОМНК окажутся смещенными. Другими словами, в ошибке осталась детерминированная (неслучайная) составляющая, которая может быть функцией входящих в модель регрессоров, что и означает, что функциональная форма выбрана неверно. Заметить эту ошибку спецификации можно с помощью графиков остатков по „подозрительным“ переменным: регрессорам и их функциям (в т. ч. произведениям разных регрессоров), расчетным значениям и их функциям. Остатки дают представления об ошибках, поэтому они должны в правильно заданной регрессии иметь везде нулевое среднее. Если остатки (e), например, для каких-то значений некоторой переменной x в среднем больше нуля, а для каких-то – меньше, то это служит признаком неправильно специфицированной модели (см. Рис. 1).

Похожим образом обнаруживается и гетероскедастичность (отсутствие гомоскедастичности). Она проявляется в том, что разброс остатков меняется в зависимости от некоторой переменной x (см. Рис. 2).

Дисперсия ошибок может меняться в зависимости от регрессоров и их функций, расчетных значений и их функций. Формальный тест можно провести с помощью вспомогательной регрессии — регрессии квадратов остатков по „подозрительным“ переменным и константе. Соответствующая статистика — обычная

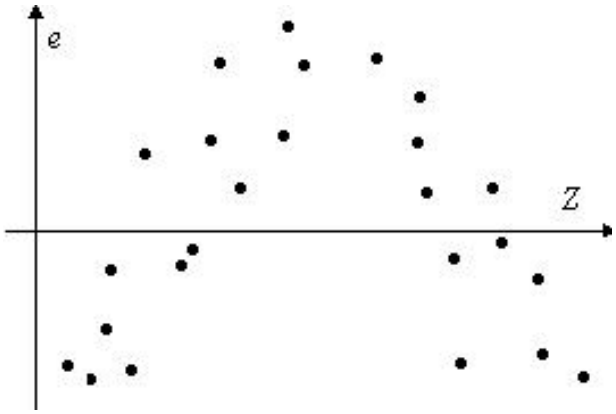

Рис. 1.

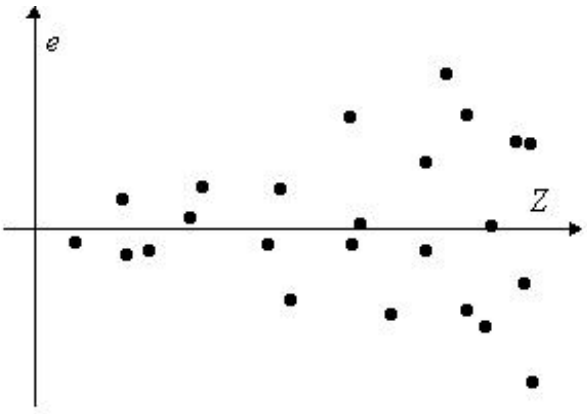

Рис. 2.

$F-$ статистика для гипотезы о равенстве нулю коэффициентов при всех переменных кроме константы, выдаваемая любым статистическим пакетом. Ошибки в спецификации функциональной формы обнаруживаются также тестами на автокорреляцию остатков, такими как статистика Дарбина-Уотсона, если наблюдения упорядочены по каком-либо признаку, например, по порядку возрастания одного из регрессоров. Понятно, что это тест неформальный.

3.2. Линейные и нелинейные модели

Линейная форма модели в целом является более предпочтительной. Линейные модели оцениваются более простым методом наименьших квадратов. При выполнении некоторого набора гипотез оценки ОМНК для линейной модели обладают рядом хороших свойств, невыполняющихся для оценок нелинейной модели, это же относится к распределениям оценок и различных статистик. В линейной регрессионной модели мат. ожидание зависимой переменной — это линейная комбинация регрессоров с неизвестными коэффициентами, которые и являются оцениваемыми параметрами модели.Такая модель является **линейной по виду**. В матричной форме ее можно записать как $Y = X\beta + \varepsilon$. Не обязательно, чтобы влияющие на Y факторы входили в модель линейно. Регрессорами могут быть любые точно заданные (не содержащие неизвестных параметров) функции исходных факторов — это не меняет свойств ОМНК. Важно, чтобы модель была **линейной по параметрам**. Бывает, что модель записана в виде, который нелинеен по параметрам, но преобразованием уравнения регрессии и переобозначением параметров можно привести ее к линейному виду. Такую модель называют **внутренне линейной**.

Поясним введенные понятия на примерах. Модель $Y = \alpha + \beta X_1 X_2 + \varepsilon$ нелинейна по X_1 и X_2, но линейна по параметрам, и можно сделать замену $X = X_1 X_2$, так что модель примет линейный вид: $Y = \alpha + \beta X + \varepsilon$. Модель $Y = \exp(\alpha + \beta X + \varepsilon)$ нелинейна по виду, но сводится к ней логарифмированием обеих частей: $\ln Y = \alpha + \beta X + \varepsilon$. В этой новой модели зависимой переменной будет уже $\ln Y$. Модель $Y = (a-1)(b+X) + \varepsilon$ нелинейна по параметрам a и b, но сводится к линейной заменой параметров $\alpha = (a-1)b$ и $\beta = a - 1$. Тогда $Y = \alpha + \beta X + \varepsilon$.

Для применения метода наименьших квадратов важно, чтобы ошибка была **аддитивной**, то есть, чтобы зависимая переменная являлась суммой своего математического ожидания и ошибки. Об этом следует помнить, производя преобразования модели. Например, модель $Y = X^\beta + \varepsilon$ нельзя преобразовать в линейную по параметрам с аддитивной ошибкой. Аналогичную **модель с мультипликативной ошибкой** $Y = \alpha X^\beta \varepsilon$ можно преобразовать к виду $\ln Y = \ln\alpha + \beta \ln X + \ln\varepsilon$ или $\widetilde{Y} = \widetilde{\alpha} + \beta\widetilde{X} + \widetilde{\varepsilon}$, где $\widetilde{Y} = \ln Y$, $\widetilde{\alpha} = \ln\alpha$, $\widetilde{X} = \ln X$, $\widetilde{\varepsilon} = \ln\varepsilon$. Однако следует отметить, что вследствие преобразования распределение ошибки изменилось. Если $\widetilde{\varepsilon}$ оказывается нормально распределенной, это значит, что ε имела логнормальное распределение.

Экономическая теория оперирует моделями разных типов. Некоторые из них дают регрессионные уравнения линейного вида, некоторые – нелинейного. Рассмотрим это на примере однородных производственных функций. Самая популярная производственная функция — **функция Кобба-Дугласа** — легко приводится к линей-

ному виду логарифмированием:

$$Y = \alpha K^{\beta} L^{1-\beta} \Rightarrow$$
$$\ln Y - \ln L = \ln \alpha + \beta(\ln K - \ln L),$$

где Y – выпуск продукции, K – капитал, L – труд.

Функция с постоянной эластичностью замены (ПЭЗ) дает внутренне нелинейное уравнение регрессии:

$$Y = \alpha(\beta K^{\rho} + (1-\beta)L^{\rho})^{1/\rho}.$$

Достаточно гладкую функцию вблизи некоторой точки можно разложить в ряд Тейлора, получив тем самым линейную форму модели. Так, при $\rho \to 0$ функция с постоянной эластичностью замены совпадает с функцией Кобба-Дугласа. Если же приблизить функцию ПЭЗ в точке $\rho = 0$ разложением в ряд Тейлора до членов первого порядка, то получается так называемая **транслоговая производственная функция**:

$$\ln Y - \ln L = \ln \alpha + \beta(\ln K - \ln L) + \gamma(\ln K - \ln L)^2,$$

где $\gamma = \frac{1}{2}\rho\beta(1-\beta)$.

Разложение в ряд Тейлора дает **полиномиальную форму модели**. В полиномиальную регрессионную модель могут входить не только первые степени исходных переменных, но и их одночлены различных степеней: степени этих переменных и **члены взаимодействия** (произведения степеней двух или более различных переменных).

Может случиться: что „истинная“модель бывает настолько нелинейной, что полиномиальное приближение становится неудовлетворительным — количество оцениваемых параметров было бы слишком большим. Тогда приходится пожертвовать удобствами ОМНК и использовать нелинейный МНК или другие методы. Есть также много других причин, по которым предпочтительнее использовать внутренне нелинейную функциональную форму. Например, функция ПЭЗ, рассмотренная выше, включает в себя как частные случаи при разных значениях параметра ρ сразу несколько популярных видов производственных функций: функцию Кобба-Дугласа, линейную функцию (с полной взаимозаменяемостью факторов) и функцию леонтьевского типа (с полной взаимодополняемостью факторов). Оценив ее, можно сделать вывод о том, к какому из этих трех видов ближе „истинная“ функция.

Кроме натуральных степеней исходных переменных можно использовать и другие функции от них. Это и уже встречавшиеся выше логарифмы и т. п.: $\ln X$, $\sqrt{X}$, $\frac{1}{X}$, $\exp(X)$, $\frac{1}{1+\exp(-X)}$ (логиста) и др. Интересной функцией является **преобразование Бокса-Кокса**: $\frac{X^{\alpha}-1}{\alpha}$. При $\alpha \to 0$ она стремится к $\ln X$. При других значениях это некоторая степень X (с точностью до линейного преобразования). В этом отношении преобразование Бокса-Кокса схоже с функцией ПЭЗ. Оно также похоже на нее в том отношении, что дает внутренне нелинейную модель. Обычно исследователь обладает достаточной свободой при выборе функциональной формы модели. Но важно, чтобы при этом не нарушались те условия, которые необходимы для хорошей работы применяемых методов оценивания. Нужно не забывать

проводить проверку правильности спецификации модели и исправлять модель, когда получена плохая диагностика, например, добавлять одночлены более высоких степеней в полиномиальную модель.

Рассмотрим, как может помочь изменение функциональной формы в борьбе с гетероскедастичностью. Многие экономические переменные таковы, что размер отклонений, с ними связанных, зависит от величины этих переменных (например, пропорционален), а величина эта в выборке колеблется в широких пределах (изменяется в несколько раз). Возникающая при этом гетероскедастичность снижает эффективность оценок параметров. Объяснить потерю эффективности можно следующим образом. В методе наименьших квадратов все наблюдения выступают в одинаковых „весовых категориях“, и поэтому в оценках непропорционально мало используется информация от наблюдений с меньшей дисперсией. Тем самым происходит потеря информации. Поэтому, например, нехорошо в регрессию включать временные ряды для номинальных показателей, если в рассматриваемой стране высокая инфляция, или использовать непреобразованную модель в случае выборки стран, в которой есть и большие, и малые страны (США наряду с Исландией). Обычно применяют два вида преобразований. Рассмотрим их на примере функции потребительского спроса кейнсианского типа: $C = \alpha I + \beta X + \varepsilon$, где C – потребление, I – доход, X – символизирует прочие факторы. Разумно предположить, что среднеквадратическое отклонение ошибки прямо пропорционально I.

1. **Нормирование**. Пронормировать рассматриваемую модель можно, разделив ее на I:

$$C/_I = \alpha + \beta\, X/_I + \varepsilon/_I$$

Можно использовать для нормировки (взвешивания) и переменную, не входящую в модель. Обозначим ее N:

$$C/_N = \alpha + \beta\, X/_N + \varepsilon/_N.$$

Нормирование равнозначно использованию взвешенного метода наименьших квадратов. Как веса для номинальных величин можно использовать уровень цен, получив тем самым реальные величины. Как веса для стран можно использовать население, получив тем самым среднедушевые показатели (потребление на душу населения и т. п.).

2. **Логарифмирование**. Прологарифмировав уравнение $C = \alpha I + \varepsilon$ при $\sqrt{V(\varepsilon)} \ll C$ можно получить следующее линейное приближение:

$$\ln C = \ln \alpha + \ln I + \varepsilon/_{\alpha I}.$$

Вряд ли можно привести теоретические возражения и против того, чтобы сразу использовать **линейную в логарифмах модель** (эта форма модели сокращенно называется **логлинейной**), например,

$$\ln C = \alpha + \beta \ln I + \varepsilon.$$

„Кандидатами“ на логарифмирование в первую очередь служат те переменные, которые заведомо могут принимать только положительные значения. Один из

их признаков, это то, что, как правило, интересуются относительными приростами таких переменных, а не абсолютными приростами. В экономике это следующие величины: физические объемы благ, цены, стоимостные показатели, различные индексы.

Как итог, перечислим основные функциональные формы регрессионной модели (без учета ошибки) с примерами:

Функциональная форма	Пример
Линейная	$Y = \alpha_0 + \alpha_1 X_2$
Полиномиальная	$Y = \alpha_0 + \alpha_1 X^1 + \alpha_3 X^3$ $Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_{11} X_1^2 + \alpha_{22} X_2^2 + +\alpha_{12} X_1 X_2$
Логлинейная (линейная в логарифмах)	$\ln Y = \alpha_0 + \alpha_1 \ln X$
Мультипликативная	$Y = \alpha_0 X_1^{\alpha_1} X_2^{\alpha_2}$
Нормированная	$\frac{Y}{N} = \alpha_0 + \alpha_1 \frac{X}{N}$

Возможны различные комбинации этих форм. Например, часто встречается **полулогарифмическая форма**:

$$\ln Y = \alpha + \beta X, \text{ или } \quad \ln Y = \alpha + \beta \ln X,$$
$$\text{или } \quad \ln Y = \alpha + \beta \ln X + \gamma Z.$$

3.3. Выбор между альтернативными функциональными формами

Самый распространенный способ выбора между альтернативными моделями — выбор на основе точности подбора. В качестве показателя точности подбора обычно используется коэффициент детерминации $\boldsymbol{R}^2$. Не следует забывать, что этот показатель можно использовать для сравнения только моделей с одной и той же зависимой переменной. Чтобы учитывать при выборе простоту модели, делают поправку на количество регрессоров. Это дает коэффициент детерминации скорректированный на количество степеней свободы $\widetilde{\boldsymbol{R}}^{\boldsymbol{2}}$.

Оценки метода наименьших квадратов являются одновременно и оценками метода максимального правдоподобия. Поэтому предлагается сравнивать модели на основе максимума логарифмической функции правдоподобия $\hat{\ell}$. Если учесть при этом количество наблюдений N и ввести „штраф“ за большое количество регрессоров k, то получится **информационный критерий Акаике** (Akaike information criterion):

$$AIC = -\frac{2}{N}(\hat{\ell} - k).$$

Чем меньше AIC, тем лучшей считается модель.

Существует и другой подход к выбору между моделями. Одна из моделей предполагается истинной, т.е. принимается за нулевую гипотезу, и тестируется против некоторой альтернативной гипотезы, спецификация которой зависит от альтернативной модели. По сути дела, осуществляется тестирование функциональной формы „нулевой“ модели.

Если одна из моделей является частным случаем другой модели (англ. nested), то в качестве „нулевой“ берется более узкая модель, а альтернативой служит более широкая. В случае линейной регрессии применяется соответствующая $F-$ статистика, а в случае нелинейной — одна из χ^2- статистик: статистика Вальда, множителя Лагранжа или отношения правдоподобия. Если же модели не входят одна в другую (nonnested), то любая из них принимается за нулевую и дополняется за счет информации, содержащейся в другой модели, так, чтобы „нулевая“ модель была частным случаем этой расширенной. Здесь уже можно применить один из вышеупомянутых тестов. Если нулевая гипотеза отвергается, то это означает, что альтернативная модель содержит какую-то информацию, не содержащуюся в „нулевой“ модели.

Тестов такого рода предложено очень много. Опишем только концептуально наиболее простые.

Сначала рассмотрим случай, когда обе сравниваемые модели линейны и зависимая переменная одна и та же. **J− тест** заключается в том, что в „нулевую“ модель добавляется в качестве еще одного регрессора расчетные значения из альтернативной модели. Проверяется гипотеза о равенстве коэффициента при дополнительном регрессоре нулю с помощью соответствующей $t-$ статистики.

Похожий тест состоит в том, что в „нулевую“ модель добавляют из альтернативной модели все те регрессоры, которые не содержатся в нулевой и проверяют гипотезу о равенстве коэффициентов при дополнительных регрессорах нулю с помощью соответствующей $F-$ статистики. В этом тесте обе сравниваемые модели содержатся в расширенной модели.

Один из тестов для сравнения моделей с разными зависимыми переменными — $\mathbf{P}_E-$ **тест**. Пусть две сравниваемые модели заданы следующими уравнениями:

$$Y_1 = f_1(Y) = X_1\beta_1 + \varepsilon_1,$$

$$Y_2 = f_2(Y) = X_2\beta_2 + \varepsilon_2.$$

Например, $f_i(Y) = Y$, $f_i(Y) = \ln Y$ или $f_i(Y) = {}^{Y}\!/_{W}$ („взвешенная“ зависимая переменная). В „нулевую“ модель в P_E -тесте добавляется регрессор, равный разности расчетных значений из альтернативной модели и приведенных к тому же виду расчетных значений из „нулевой“ модели. Так, в первую модель нужно добавить

$$X_2\hat{\beta}_2 - f_2(f_1^{-1}(X_1\hat{\beta}_1)) = \hat{Y}_2 - f_2(f_1^{-1}(\hat{Y}_1)).$$

Пусть, к примеру, $f_1(Y) = Y$, а $f_2(Y) = \ln(Y)$. Тогда в первую модель добавляют $\hat{Y}_2 - \ln(\hat{Y}_1)$, а во вторую — $\hat{Y}_1 - \exp(\hat{Y}_2)$.

Если отвергаются обе модели, то это должно означать, что каждая из них содержит информацию, не содержащуюся в другой, и следует попытаться как-то соединить две модели в одну. Если обе модели не отвергаются, то это означает, что с точки зрения данного теста они эквивалентны.

4. Фиктивные переменные как регрессоры

Термин „фиктивные переменные“ используется как противоположность „значащим“ переменным, показывающим уровень количественного показателя, принимающего значения из непрерывного интервала. Как правило, фиктивная переменная

— это индикаторная переменная, отражающая качественную характеристику. Чаще всего применяются бинарные фиктивные переменные, принимающие два значения, 0 и 1, в зависимости от определенного условия. Например, в результате опроса группы людей 0 может означать, что опрашиваемый — мужчина, а 1 — женщина. К фиктивным переменным иногда относят регрессор, состоящий из одних единиц (т.е. константу, свободный член), а также временной тренд.

Фиктивные переменные, будучи экзогенными, не создают каких-либо трудностей при применении ОМНК. Фиктивные переменные являются эффективным инструментом построения регрессионных моделей и проверки гипотез.

Пример. (Проверка гипотезы о равенстве средних в двух выборках в предположении равенства дисперсий)

Нулевая гипотеза состоит в том, что случайные величины в двух выборках имеют одинаковое математическое ожидание. Альтернативная гипотеза состоит в том, что математические ожидания равны только в пределах выборок, но не между выборками. Предполагается, что величины нормально распределены и дисперсии одинаковы для всех наблюдений. Объединим две выборки в одну. Пусть Y_i — вектор наблюдений для данной величины, D_i — фиктивная переменная принимающая значение 0 для первой выборки и 1 для второй выборки. Тогда для проверки гипотезы оценим регрессионную модель:

$$Y_i = \alpha + \beta D_i + \varepsilon_i.$$

Нулевая гипотеза $H_0 : \beta = 0$, альтернативная гипотеза $H_1 : \beta \neq 0$. Такую гипотезу можно проверить с помощью $t-$ статистики для коэффициента β. В этом случае $\hat{\alpha}$ будет оценкой математического ожидания для первой выборки, $\hat{\alpha} + \hat{\beta}$ для второй.

Предположим, что математическое ожидание зависимой переменной в регрессии увеличивается на некоторую фиксированную величину, если выполняется определенное условие. Пусть, например, для выборки предприятий одной отрасли оценивается производственная функция Кобба-Дугласа. Есть гипотеза, что для частных предприятий в этой отрасли производство при тех же труде и капитале выше, чем для государственных. Введем переменную D_i , которая принимает значение 0 для государственных предприятий и 1 для частных. Регрессионное уравнение будет иметь вид:

$$\ln Y_i - \ln L_i = \alpha_0 + \alpha_1 D_i + \beta(\ln K_i - \ln L_i).$$

Если коэффициент α_1 значимо положителен, то гипотезу нельзя отвергнуть.

Еще одна область применения фиктивных переменных — когда предполагается, что коэффициенты при „значащих“ переменных меняются в зависимости от некоторого условия.

Пусть в приведенной модели $\beta = \beta_0$ для государственных предприятий и $\beta = \beta_1$ для частных. Тогда модель запишется в виде:

$$\ln Y_i - \ln L_i = \alpha_0 + \alpha_1 D_i + \beta_0(\ln K_i - \ln L_i) + (\beta_1 - \beta_0) D_i(\ln K_i - \ln L_i).$$

Заменив параметры, получаем линейную относительно параметров модель.

В регрессионное уравнение может войти несколько фиктивных переменных. Рассмотрим два условия: A и B. Пусть D_i^A — индикатор условия A ($D_i^A = 1$, если выполнено условие A, и 0 — если нет), D_i^B — индикатор условия B. Тогда $D_i^{AB} = D_i^A D_i^B$ — индикатор одновременного выполнения условий A и B. Эти три переменные будут точно описывать, в каком состоянии находится „мир“ для данного наблюдения. Следует отметить, что четвертая фиктивная переменная (индикатор того, что одновременно не выполнены условия A и B) будет излишней в регрессии, уже включающей константу. Если ее добавить в регрессию, то матрица регрессоров будет вырожденной.

Дисперсионный анализ с фиксированными эффектами может быть сведен к регрессионному анализу с фиктивными регрессорами. Проверке гипотез с помощью ковариационного анализа будет соответствовать проверка гипотезы о равенстве нулю коэффициентов при соответствующей группе фиктивных переменных.

4.1. Использование фиктивных переменных для проверки однородности наблюдений и прогнозирования

Приведенную выше модель для двух типов предприятий

$$\ln Y_i - \ln L_i = \alpha_0 + \alpha_1 D_i + \beta_0(\ln K_i - \ln L_i) + (\beta_1 - \beta_0) D_i(\ln K_i - \ln L_i).$$

можно использовать для проверки гипотезы о том, что коэффициенты регрессии разные для гос. предприятий и для частных. Гипотеза проверяется с помощью $F-$ теста на добавление переменных D_i и $D_i(\ln K_i - \ln L_i)$.

В общем случае пусть наблюдения разбиты на две группы — I^1 и I^2 . Матрица регрессоров X распадается на две матрицы регрессоров X^1 и X^2 соответственно, а зависимая переменная Y — на Y^1 и Y^2 соответственно. Нулевая гипотеза состоит в том, что наблюдения порождены моделью $Y = X\beta + \varepsilon$. Альтернативная гипотеза состоит в том, что первая группа наблюдений порождена моделью $Y^1 = X^1\beta^1 + \varepsilon^1$, а вторая группа наблюдений — моделью $Y^2 = X^2\beta^2 + \varepsilon^2$, причем $\beta^1 \neq \beta^2$.

Введем фиктивную переменную D, такую что $D_i = 0$ при $i \in I^1$ и $D_i = 1$ при $i \in I^2$. Если все ошибки имеют одинаковую дисперсию, то гипотезу можно проверить с помощью регрессии Y по $Z = [X, X * D]$. Здесь $X * D$ обозначает прямое произведение матрицы X на D, так что $i-$ я строка матрицы Z равна $Z_i = [X_i, D_i X_i]$.

Тест на равенство коэффициентов регрессии в двух выборках, называют **тестом Чоу**. Нулевая гипотеза проверяется с помощью $F-$ статистики для гипотезы о том, что коэффициенты при всех добавленных переменных равны нулю.

Еще одно использование фиктивных переменных — проверка гипотезы о том, что некоторое наблюдение принадлежит к той же выборке, что и все остальные наблюдения. Пусть i^* — номер этого наблюдения. Введем фиктивную переменную D, такую что $D_i = 0$ при $i \neq i^*$ и $D_{i^*} = 1$. Добавим эту переменную в исходную регрессию. Нужной нам статистикой будет $F-$ или $t-$ статистика для гипотезы о том, что коэффициент при добавленной переменной равен нулю. Если нулевая гипотеза отвергается, то соответствующее наблюдение следует считать **выбросом**. Назовем этот тест тестом для выбросов.

Тот же тест можно провести для небольшой группы наблюдений I^*. Требуется добавить регрессию по одной фиктивной переменной описанного вида для каждого из наблюдений $i \in I^*$. Нужной нам статистикой будет $F-$ статистика для гипотезы о том, что коэффициенты при всех добавленных переменных одновременно равны нулю.

Фиктивные переменные, которые равны нулю для всех наблюдений кроме одного, обладают тем свойством, что при добавлении их в регрессию соответствующий остаток зануляется.

Если в тесте Чоу одна из двух выборок содержит мало наблюдений (не больше количества регрессоров), то остатки в этой выборке должны занулиться при применении ОМНК. В этом случае тест Чоу совпадает с описанным только что тестом для выбросов.

Рассмотрим теперь использование фиктивных переменных для прогнозирования. Пусть мы оценили некоторую регрессию $Y = X\beta + \varepsilon$ и у нас имеются дополнительные наблюдения, для которых известна матрица регрессоров X^*, но неизвестны значения зависимой переменной Y^*. Предсказания находятся по формуле $X^*\tilde{\beta}$, где $\tilde{\beta}$ – оценки ОМНК из регрессии Y по X. Эти предсказания можно найти с помощью следующей регрессионной модели:

$$\begin{bmatrix} Y \\ 0 \end{bmatrix} = \begin{bmatrix} X & 0 \\ X^* & I \end{bmatrix} \begin{bmatrix} \beta \\ \beta^* \end{bmatrix} + \begin{bmatrix} \varepsilon \\ \varepsilon^* \end{bmatrix}.$$

Вместо неизвестной зависимой переменной здесь стоят нули, и добавлены фиктивные переменные, каждая из которых равна нулю для соответственного добавочного наблюдения. Оценки β будут совпадать с $\hat{\beta}$, а оценки β^* будут равны $\hat{\beta}^* = -X^*\hat{\beta}$, то есть будут равны предсказаниям со знаком минус. Стандартные ошибки предсказаний будут равны стандартным ошибкам оценок $\tilde{\beta}^*$, полученным из той же регрессии.

Пусть теперь Y^* становятся известными. Интересно было бы проверить, насколько фактические значения отличаются от предсказанных. Оказывается, можно воспользоваться аналогичной регрессией, в которой слева вместо нулей стоят Y^*:

$$\begin{bmatrix} Y \\ Y^* \end{bmatrix} = \begin{bmatrix} X & 0 \\ X^* & I \end{bmatrix} \begin{bmatrix} \beta \\ \beta^* \end{bmatrix} + \begin{bmatrix} \varepsilon \\ \varepsilon^* \end{bmatrix}.$$

Оценки коэффициентов при фиктивных переменных в этом случае будут равны ошибкам предсказаний $\hat{\beta}^* = Y^* - X^*\hat{\beta}$. Тест на адекватность предсказаний проводится как тест на одновременное равенство коэффициентов при фиктивных переменных нулю: $\beta^* = 0$. Очевидно, что этот тест совпадает с тестом для выбросов.

4.2. Использование фиктивных переменных в моделях с временными рядами

В регрессионных моделях с временными рядами используется три основных вида фиктивных переменных:

1. Переменные-индикаторы принадлежности наблюдения к определенному периоду — для моделирования скачкообразных структурных сдвигов. Границы периода (моменты „скачков“) должны быть установлены из априорных соображений. Например, 1, если наблюдение принадлежит периоду 1941-45 гг. и 0 в противном случае. Это пример использования для моделирования временного структурного сдвига. Постоянный структурный сдвиг моделируется переменной равной 0 до определенного момента времени и 1 для всех наблюдений после этого момента времени.

2. Сезонные переменные — для моделирования сезонности. Сезонные переменные принимают разные значения в зависимости от того, какому месяцу или кварталу года или какому дню недели соответствует наблюдение.

3. Линейный временной тренд — для моделирования постепенных плавных структурных сдвигов. Эта фиктивная переменная показывает, какой промежуток времени прошел от некоторого „нулевого“ момента времени до того момента, к которому относится данное наблюдение (координаты данного наблюдения на временной шкале). Если промежутки времени между последовательными наблюдениями одинаковы, то временной тренд можно составить из номеров наблюдений.

Фиктивные переменные помогают отразить тот факт, что коэффициенты линейной регрессии могут меняться во времени. В простейшем случае изменяется константа, а тем самым и математическое ожидание зависимой переменной.

Пусть исходная модель имеет вид $Y_t = \alpha + \beta X_t + \varepsilon_t$ и предполагается, что α линейно зависит от фиктивной переменной F_t : $\alpha_t = \alpha_0 + \alpha_1 F_t$, тогда уравнение изменится следующим образом: $Y_t = \alpha_0 + \alpha_1 F_t + \beta X_t + \varepsilon_t$, оставаясь линейным по параметрам.

Коэффициенты при значащих переменных тоже могут быть подвержены изменениям. Проинтерпретировать это можно так, что сила их влияния на независимую переменную меняется со временем.

Например, в рассмотренном уравнении может быть $\beta_t = \beta_0 + \beta_1 F_t$, тогда $Y_t = \alpha + \beta_0 X_t + \beta_1 F_t X_t + \varepsilon$. Эта модель также остается линейной по параметрам. Коэффициент β_1 показывает, как исходный коэффициент β зависит от времени. С помощью соответствующей $t-$ статистики можно проверить гипотезу, что $\beta_1 = \beta_0$ (β не меняется со временем).

Можно предложить следующий тест на стабильность коэффициентов модели во времени. Для его проведения нужно добавить в уравнения произведения всех исходных регрессоров и фиктивной переменной. Например, в модель $Y_t = \alpha + \beta_1 X_{t1} + \beta_2 X_{t2} + \varepsilon$ следует добавить регрессоры F_t, $X_{t1} F_t$ и $X_{t2} F_t$. Если коэффициенты при добавочных переменных значимы в совокупности (применяем $F-$ статистику), то нельзя отвергнуть гипотезу о том, что коэффициенты изменяются со временем.

Тест Чоу представляет собой частный случай описанного теста. Для временных рядов тест Чоу — это тест на то, что в определенный момент времени произошло скачкообразное изменение коэффициентов регрессии.

Временной тренд отличается от бинарных фиктивных переменных тем, что имеет смысл использовать его степени: t^2, t^3 и т. д. Они помогают моделировать

гладкий, но нелинейный тренд. (Бинарную переменную нет смысла возводить в степень, потому что в результате получится та же самая переменная.)

Можно также комбинировать три указанных вида фиктивных переменных, создавая переменные „взаимодействия“ соответствующих эффектов. Пусть Y — квартальные данные по некоторому показателю. Его поведение можно смоделировать, представляя мат. ожидание как комбинацию линейного тренда и сезонности.

$$Y_t = \alpha_0 + \alpha_1 t + \beta_1 Q_{t1} + \beta_2 Q_{t2} + \beta_3 Q_{t3} + \gamma_1 Q_{t1} t + \gamma_2 Q_{t2} t + \gamma_3 Q_{t3} t + \varepsilon_t,$$

где t — тренд, Q_i — квартальные сезонные переменные

$$Q_{tj} = \begin{cases} 1, & \text{если } j\text{-й квартал;} \\ 0, & \text{иначе.} \end{cases}$$

Q_{t4} не нужно вводить в эту регрессию, так как есть константа, а $Q_{t4}t$ не нужно вводить в регрессию, так как есть временной тренд t. Если все $\gamma_j \neq 0$, то это означает, что структура сезонности линейно изменяется со временем.

Комбинация рассмотренных фиктивных переменных позволяет моделировать еще один эффект — изменение наклона тренда с определенного момента. Помимо тренда в регрессию следует тогда ввести следующую переменную: в начале выборки до некоторого момента времени она равна 0, а вторая ее часть представляет собой временной тренд (1, 2, 3 и т. д. в случае одинаковых интервалов между наблюдениями).

Регрессионные модели с фиктивными переменными являются альтернативой $ARIMA-$ моделям и регрессионным моделям с $AR-$ или $MA-$ процессом в ошибке. В первом случае изменение мат. ожидания во времени можно назвать детерминированным трендом, во втором — стохастическим (строго говоря термин „стохастический тренд“ употребляют только по отношению к нестационарным процессам). Решить, какой вид модели применять, сложно. Дело в том, что трудно отличить (в случае малых выборок), когда случайная величина имеет линейный детерминированный тренд со стационарными отклонениями от него, а когда она формируется нестационарным авторегрессионным процессом. То же самое верно для выбора способа моделирования сезонности.

Использование фиктивных переменных имеет следующие преимущества:

1. Интервалы между наблюдениями не обязательно должны быть одинаковыми. В выборке могут быть пропущенные наблюдения.

2. Коэффициенты при фиктивных переменных легко интерпретировать, они наглядно представляют структуру динамического процесса.

3. Для оценивания модели не приходится выходить за рамки классического метода наименьших квадратов.

5. Организационно-методический раздел

5.1. Цель, место и задачи курса

Цель курса

Конечной целью изучения дисциплины является формирование у будущих специалистов теоретических знаний и практических навыков по применению статистических методов для исследования и обобщения эмпирических зависимостей экономических переменных, а также построения надежных прогнозов в банковском деле, финансах, различных сферах предпринимательской деятельности с целью обоснования принимаемых решений.

Задачи курса

Основной задачей изучения дисциплины "Эконометрическое моделирование"является реализация требований, установленных Государственным стандартом высшего профессионального образования к подготовке специалистов в области экономических и бизнес-дисциплин. Данный курс рассчитан на студентов дневного отделения, обучающихся по специальности «Математические методы в экономике».

Место курса в системе дисциплин экономического образования

Эконометрика объединяет совокупность методов и моделей, позволяющих на базе экономической теории, экономической статистики и математико-статистического инструментария исследовать количественные выражения качественных зависимостей. При изучении дисциплины «Эконометрическое моделирование» предполагается, что студент владеет основами теории вероятностей, математической статистики, линейной алгебры и базовым уровнем математического анализа в объеме курса высшей математики для экономических специальностей.

Требования к уровню освоения содержания курса

В результате изучения курса «Эконометрическое моделирование» дипломированный специалист в области экономики должен знать и уметь:

- формировать концепцию эконометрической модели на основе качественного анализа объекта исследования;
- собирать и проводить статистическую обработку экономической информации с целью выявления основных характеристик числовой совокупности;
- проводить оценку взаимосвязей экономических показателей с помощью статистических методов, интерпретировать полученные результаты по оценке взаимосвязей с точки зрения экономической сущности явлений;
- строить эконометрические модели с использованием процедур регрессионного анализа и анализа временных рядов;
- оценивать качество построенных эконометрических моделей с точки зрения их адекватности фактическим данным;
- применять эконометрические модели в практике хозяйственного управления.

5.2. Содержание курса

Распределение часов по темам и видам учебной работы

Название и разделов и тем	Всего	Виды учебных занятий			
		Аудиторные занятия			Самостоятельная работа
		лекции	практические занятия, семинар	лабораторная работа	
Раздел 1. (Метод наименьших квадратов)					
1. Введение	2	2			
2. Простой МНК	6	2	2		2
3. Обобщенный МНК	6	2	2		2
Раздел 2. (Метод максимального правдоподобия)					
1. ММП в эконометрии	6	2			4
2. Свойства оценок ММП	12	2	4		6
3. Численные методы нахождения оценок ММП	6	2			4
Раздел 3. (Методы выбора модели)					
1. Функциональная форма регрессионной модели	8	2	2		4
2. Выбор между альтернативными формами	8	2	2		4
Раздел 4. (Модели временных рядов)					
1. Стационарные ряды	10	2	2		6
2. Модели ARMA	14	2	4		8
3. Процесс авторегрессии	12	2	4		6
4. Процесс скользящего среднего	10	2	2		6
5. Смешанный процесс авторегрессии	10	2	2		6
6. Подбор стационарной модели	16	4	4		8
Итого	126	30	30		66

Содержание курса

Раздел 1. Метод наименьших квадратов.

1. Введение. Основные задачи эконометрического исследования.

2. Простой МНК. Обзор результатов классической линейной модели регрессионного анализа. Метод наименьших квадратов (МНК). Свойства оценок МНК. Показатели качества регрессии. Построение доверительных интервалов и проверка линейных гипотез о параметрах.

3. Обобщенный МНК. Обобщенный метод наименьших квадратов (ОМНК). Свойства оценок ОМНК.

Раздел 2. Метод максимального правдоподобия.

1. ММП в эконометрии. Базовые понятия. Характеристика ММП. Связь ММП с МНК. Квази-МП методы.

2. Свойства оценок ММП. Связь гессиана и матрицы вкладов в градиент с информационной матрицей. Вычисление информационной матрицы. Распределение градиента и оценок максимального правдоподобия.

3. Численные методы нахождения оценок ММП. Численные методы нахождения оценок максимального правдоподобия, основанные на максимизации функции правдоподобия и методов построения градиента и гессиана целевой функции.

Раздел 3. Методы выбора модели.

1. Функциональная форма регрессионной модели. Тестирование правильности спецификации регрессионной модели. Линейные и нелинейные модели.

2. Выбор между альтернативными формами. Методы оценки качества построенной регерессионной модели с целью сравнения с другими формами и выбор наилучшей.

Раздел 4. Модели временных рядов.

1. Стационарные ряды. Общие понятия временного ряда, стационарность в узком и широком смысле. Процесс белого шума.

2. Модели ARMA. Общие понятия.

3. Процесс авторегрессии. AR(p)-модели, их свойства, области применения.

4. Процесс скользящего среднего. MA(q)-модели, их свойства, области применения.

5. Смешанный процесс авторегрессии. Смешанный процесс авторегрессии - скользящего среднего (процесс авторегрессии с остатками в виде скользящего среднего). ARMA(p,q)-модели, их свойства, области применения. Модели ARMA, учитывающие наличие сезонности.

6. Подбор стационарной модели. Подбор стационарной модели ARMA для ряда наблюдений. Идентификация стационарной модели ARMA, оценивание коэффициентов модели, диагностика оцененной модели.

Темы семинарских занятий

Раздел 1. Метод наименьших квадратов.

1. Простой МНК. Основные экономические показатели и простейшие задачи.

2. Обобщенный МНК. Метод наименьших квадратов (МНК).

Раздел 2. Метод максимального правдоподобия.

1. Свойства оценок ММП. Связь гессиана и матрицы вкладов в градиент с информационной матрицей. Вычисление информационной матрицы. Распределение градиента и оценок максимального правдоподобия.

Раздел 3. Методы выбора модели.

1. Функциональная форма регрессионной модели. Типовые зависимости и приведение их к линейному виду относительно параметров.

2. Выбор между альтернативными формами. Линейная регрессия с нормально распределенными ошибками.

Раздел 4. Модели временных рядов.

1. Стационарные ряды. Общие понятия временного ряда, стационарность в узком и широком смысле. Процесс белого шума.

2. Модели ARMA. Вывод некоторых частных случаев.

3. Процесс авторегрессии. Процесс авторегрессии. AR(p)-модели, коррелограмма как показатель.

4. Процесс скользящего среднего. Процесс скользящего среднего. MA(q)-модели, коррелограмма как показатель.

5. Смешанный процесс авторегрессии. Смешанный процесс авторегрессии - скользящего среднего (процесс авторегрессии с остатками в виде скользящего среднего). ARMA(p,q)-модели. Модели ARMA, учитывающие наличие сезонности.

6. Подбор стационарной модели. Построение временной модели на основе данных о годовом потреблении рыбных продуктов. Построение временной модели курса валюты.

Примерные вопросы к теме

1. Основные задачи эконометрического исследования.
2. Стационарность ряда в узком и широком смысле. Процесс белого шума.
3. Обзор результатов классической линейной модели регрессионного анализа. Метод наименьших квадратов (МНК).
4. Модели ARMA. Общие понятия.
5. Свойства оценок МНК. Показатели качества регрессии.
6. Процесс авторегрессии. AR(p)-модели, их свойства, области применения.
7. Обобщенный метод наименьших квадратов (ОМНК). Свойства оценок ОМНК.
8. Процесс скользящего среднего. MA(q)-модели, их свойства, области применения.
9. Метод максимального правдоподобия (ММП) в эконометрии. Характеристика ММП.
10. ARMA(p,q)-модели, их свойства, области применения.
11. Связь гессиана и матрицы вкладов в градиент с информационной матрицей. Вычисление информационной матрицы.
12. Модели ARMA, учитывающие наличие сезонности
13. Распределение градиента и оценок максимального правдоподобия.
14. Подбор стационарной модели ARMA для ряда наблюдений.

15. Численные методы нахождения оценок максимального правдоподобия.
16. Модели с дискретной зависимой переменной. Модели с бинарной зависимой переменной (логит и пробит).
17. Функциональная форма регрессионной модели. Тестирование правильности спецификации регрессионной модели.
18. Свойства оценок МНК. Построение доверительных интервалов и проверка линейных гипотез о параметрах.
19. Связь ММП с МНК. Квази-МП методы.
20. Идентификация стационарной модели ARMA, оценивание коэффициентов модели, диагностика оцененной модели.
21. Линейные и нелинейные модели. Выбор между альтернативными функциональными формами.
22. Коррелограмма, ее определение и свойства.

Примерные варианты заданий аудиторной работы

ВАРИАНТ 1

Задание 1. По данным $n = 15$ предприятий, каждое из которых характеризуется по трем показателям: x_1 - объем сменной выработки, x_2 - себестоимость продукции и x_3 - фондоотдача; получена матрица парных коэффициентов корреляции:

$$R = \begin{pmatrix} 1 & -0,6 & 0,8 \\ -0,6 & 1 & -0,6 \\ 0,8 & -0,6 & 1 \end{pmatrix}$$

Определите оценку частного коэффициента корреляции $r_{23/1}$.

Задание 2. По данным задания 1 проверить при $\alpha = 0,05$ значимость частного коэффициента корреляции $r_{23/1}$.

Задание 3. По данным задания 1 найти точечную оценку множественного коэффициента корреляции, характеризующего тесноту связи между себестоимостью и остальными переменными.

Задание 4. По данным заданий 1 и 3 при $\alpha = 0,05$ проверить значимость множественного коэффициента корреляции $r_{2/31}$.

Задание 5. По данным заданий 1 и 4 определите, какая доля дисперсии x_2 объясняется влиянием показателей x_1 и x_3.

Задание 6. Известно, что факторный признак x_3 усиливает связь между величинами x_1 и x_2. По результатам наблюдений получен частный коэффициент корреляции $r_{12/3} = -0,45$. Какое значение может принять парный коэффициент корреляции r_{12} :

а)-0,4; б) 0,344; в) - 0,8; г)1,2.

Задание 7. Предположим, Вы обследовали два динамических ряда А и В, каждый из которых содержит по 60 наблюдений. Данные ряды были исследованы на автокорреляцию уровней, в результате чего были получены следующие автокорреляционные функции:

Лаг(τ)	1	2	3	4	5	6	7
r_τ(А)	0,68	0,23	-0,055	-0,145	-0,106	-0,052	-0,135
r_τ(В)	-0,03	0,15	0,17	-0,08	-0,05	0,20	-0,01

Сделайте заключение о внутренней структуре временных рядов. Для описания какого из этих рядов лучше подойдет авторегрессионная модель?

Задание 8. При исследовании зависимости себестоимости продукции y от объема выпуска x_1 и производительности труда x_2 по данным $n = 20$ предприятий получено уравнение регрессии $\hat{y} = 2,88 - 0,72x_1 - 1,51x_2$. С доверительной вероятностью $p = 0,99$ определите, на какую величину максимально может измениться себестоимость продукции y, если объем производства x_1 увеличить на единицу (известно, что $S_{b1} = 0,052; \quad S_{b2} = 0,5)$:

а) -0,6; б) 0,72; в) -1,5; г) -0,83.

Задание 9. Для уравнения регрессии $\hat{y} = 2,88 - 0,72x_1 - 1,51x_2$ рассчитан множественный коэффициент корреляции $r_{y/x_1x_2} = 0,84$. Какая доля вариации результативного показателя y (в %) объясняется входящими в уравнение регрессии переменными x_1 и x_2 :

а) 70,6; б) 16,0; в) 84,0; г) 29,4.

Задание 10. Предположим, в результате Вашего исследования было получено два вида трендовых моделей, каждая из которых содержит по четыре объясняющих переменных. При этом было обследовано 35 объектов. Построенные модели имеют следующие характеристики:

Модель 1. $R^2 = 0,95$; $F = 0,5$; $DW = 3,5$;

Модель 2. $R^2 = 0,76$; $F = 1,85$; $DW = 2,1$.

Какая модель является более адекватной и почему?

ВАРИАНТ 2

Задание 1. На основании данных 5 стран о зависимости темпов роста ВНП (y) и темпов прироста промышленного производства (x) получена оценка уравнения регрессии $\hat{y} = 1,56 + 0,21x$ и оценка остаточной дисперсии, равная $0,04$. Матрица значений аргументов X имеет вид:

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 3 & 4 & 6 & 8 \end{pmatrix}^T$$

Проверить значимость коэффициента регрессии $H_0 : \beta_1 = 0$ при $\alpha = 0,05$.

Задание 2. По данным задания 1 при доверительной вероятности $p = 0,95$ определите интервальную оценку для β_0.

Задание 3. По данным задания 1, приняв $\bar{y} = 2,5$ и $\bar{x} = 3,5$ определите, на сколько процентов в среднем изменится y, если x увеличится на 1 %.

Задание 4. По данным задания 1 определите с доверительной вероятностью $p = 0,95$ интервальную оценку для β_1.

Задание 5. Множественный коэффициент корреляции $r_{1/23} = 0,8$. Определите, какой процент дисперсии величины x_1 объясняется влиянием величин x_2 и x_3 :

а) 28%; б) 32%; в) 64%; г) 80%.

Задание 6. По результатам 25 наблюдений получен парный коэффициент корреляции $r_{12} = 0,6$. Известно, что x_3 занижает связь между x_1 и x_2. Какое значение может принять частный коэффициент корреляции $r_{12/3}$:

а) -0,8; б) -0,2; в) 0,5; г) 0,8.

Задание 7. Предположим, Вы обследовали 66 наблюдений над некоторыми экономическими показателями X_1, X_2, X_3 и X_4. В результате получилась

следующая матрица парных коэффициентов корреляции:

$$R = \begin{pmatrix} 1 & 0,78 & 0,25 & -0.11 \\ 0,78 & 1 & 0,98 & 0.02 \\ 0,25 & 0,98 & 1 & 0,53 \\ -0,11 & 0,02 & 0,53 & 1 \end{pmatrix}^T .$$

Проверьте полученные парные коэффициенты корреляции на значимость ($\alpha = 0,05$) и определите, считая показатель X_1 откликом, а X_2, X_3 и X_4 - объясняющими переменными, какие из факторов, скорее всего, войдут в уравнение регрессии, а какие нет и почему?

Задание 8. Какие требования в модели регрессионного анализа предъявляются к распределению ошибок наблюдения ε_i , а именно к их математическому ожиданию и дисперсии $D\varepsilon_i$:

а) $M\varepsilon_i = 1$; $D\varepsilon_i = \sigma^2$; б) $M\varepsilon_i = 0$; $D\varepsilon_i = 1$;
в) $M\varepsilon_i = 0$; $D\varepsilon_i = \sigma^2$; г) $M\varepsilon_i = 1$; $D\varepsilon_i = 0$.

Задание 9. В результате Вашего исследования была построена следующая регрессионная модель: $Y_t = -43 + 0,39Y_{t-1} + 0,92X_t - 0,35X_{t-1} + \varepsilon_t$. Перечислите: а) объясняющие переменные; б) авторегрессионные составляющие; в) лаговые и трендовые компоненты.

Задание 10. Какое значение может принимать коэффициент детерминации: а) -0,5; б) -0,2; в) 0,4; г) 1,2.

ВАРИАНТ 3

Задание 1. Что минимизируется согласно методу наименьших квадратов:
а) $\sum_{i=1}^{n} (y_i - \bar{y})^2$; б) $\sum_{i=1}^{n} |y_i - \hat{y}_i|$; в) $\sum_{i=1}^{n} (y_i - \hat{y}_i)$; г) $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$.

Задание 2. По данным $n = 15$ фирм исследована зависимость прибыли y от числа работающих x вида: $\hat{y} = b_0 + b_1 x$. Была получена оценка остаточной дисперсии $S^2_{OCT} = 2,2$ и обратная матрица:

$$\left(X^T X\right)^{-1} = \begin{pmatrix} 0,31 & -0,03 \\ -0,03 & 0,05 \end{pmatrix}$$

Определить, чему равна ошибка оценки параметра b_1 :
а) 1,500; б) 0,332; в) 0,682; г) 0,242.

Задание 3. Предположим, Вы обследовали два динамических ряда А и В, каждый из которых содержит по 40 наблюдений. Данные ряды были исследованы на автокорреляцию уровней, в результате чего были получены следующие автокорреляционные функции:

Лаг(τ)	1	2	3	4	5	6	7
r_τ(А)	0,67	-0,21	-0,28	0,73	-0,05	-0,13	0,78
r_τ(В)	0,80	0,75	0,70	0,61	0,53	0,44	0,35

Сделайте заключение о внутренней структуре временных рядов. Для описания какого из этих рядов лучше подойдет трендовая модель?

Задание 4. Какое значение может принять множественный коэффициент корреляции:

а) –1; б) –0,5; в) 0; г) 1,2.

Задание 5. Дайте определение и укажите различие парного и частного коэффициентов корреляции.

Задание 6. Известно, что при фиксированном значении X_3 между величинами X_1 и X_2 существует положительная связь. Какое значение может принять частный коэффициент корреляции $r_{12/3}$:

а) –0,8; б) 0; в) 0,4; г) 1,3.

Задание 7. На основании данных о динамике процента хронических больных на 1000 жителей, приведенных в таблице, а также предположения, что генеральное уравнение регрессии имеет вид: $\hat{y} = \beta_0 + \beta_1 x$ определить оценки b_0 и b_1 параметров уравнения регрессии

Годы (X)	0	1	2	3	4
Доля хронических больных на 1000 жителей в % (Y)	10	8	5	3	4

Задание 8. По данным задания 7 определить величину остаточной дисперсии $S^2_{ОСТ}$, объясненной дисперсии $S^2_{\textit{объясн}}$ и общей дисперсии $S^2_{\textit{общ}}$. Проверить значимость уровня регрессии при $\alpha = 0,05$.

Задание 9. Статистический анализ экономических показателей производственного объединения показал, что зависимость объема выпускаемой продукции (Y) от производительности труда (X) оценивается двумя уравнениями регрессии: 1) $Y = 2 + 4X$; 2) $Y = 2,5 + 4,5 \ln X$.

Какое уравнение регрессии точнее описывает зависимость объема выпускаемой продукции (Y) от производительности труда (X), если известно, что в первом уравнении 80% общей дисперсии объема выпускаемой продукции определяет влияние неучтенных факторов, а во втором уравнении 60% общей дисперсии объема выпускаемой продукции обусловлено влиянием производительности труда?

Задание 10. По результатам 20 наблюдений найден множественный коэффициент корреляции $r_{1/23} = 0,8$. Проверьте значимость множественного коэффициента корреляции $H_0 : r_{1/23} = 0$ при $\alpha = 0,05$.

ВАРИАНТ 4

Задание 1. На основании данных о темпе прироста (%) ВНП (Y) и промышленного производства (X) десяти развитых стран мира за 1992 г., приведенных в таблице, и предположения, что генеральное уравнение регрессии имеет вид $\hat{y} = \beta_0 + \beta_1 x$, определить оценки параметров уравнения регрессии.

СТРАНЫ	У	Х
Япония	3,5	4,3
США	3,1	4,6
Германия	2,2	2,0
Франция	2,7	3,1
Италия	2,7	3,0
Великобритания	1,6	1,4
Канада	3,1	3,4
Австралия	1,8	2,6
Бельгия	2,3	2,6
Нидерланды	2,3	2,4

Задание 2. По данным задания 1 определить величины остаточной, объясненной и общей дисперсии.

Задание 3. По данным задания 1 и результатам расчетов в задании 2 при $\alpha = 0,05$ проверить значимость уравнения регрессии.

Задание 4. По данным задания 1 и результатам расчетов в задании 2 при $\alpha = 0,05$ проверить значимость коэффициентов уравнения регрессии.

Задание 5. Предположим, Вы исследовали экономическую природу некоторого показателя Y. В результате на основании 25 наблюдений было построено уравнение регрессии от двух факторов X_1 и X_2 следующего вида: $Y = -18,53 + 2,38X_1 - 0,76X_2$. Величина остаточной дисперсии составляет 6,79; величина объясненной дисперсии равна 15,75. Определите стандартную ошибку оценки по регрессии (среднеквадратическое отклонение от линии регрессии).

Задание 6. По данным задания 5 определите коэффициент множественной корреляции r_{Y/X_1X_2} и коэффициент детерминации.

Задание 7. По данным задания 5 определите, является ли уравнение регрессии значимым по критерию Фишера при $\alpha = 0,05$.

Задание 8. В результате исследования экономической природы выпуска некоторого продукта было построено уравнение регрессии от двух факторов L (труд) и K (капитал) на основе обследования 20 предприятий некоторой отрасли. Полученное уравнение регрессии имеет следующий вид: $Y = 5,03K^{0,3}L^{0,7}$. Остаточная дисперсия составляет 9,18; объясненная дисперсия равна 15,32. Определите стандартную ошибку оценки по регрессии (среднеквадратическое отклонение от линии регрессии).

Задание 9. По данным задания 8 определите коэффициент множественной корреляции $r_{Y/KL}$ и коэффициент детерминации.

Задание 10. По данным задания 8 определите, является ли уравнение регрессии значимым по критерию Фишера при $\alpha = 0,05$.

Основная литература

[1] Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. Учебник для вузов. - М.: ЮНИТИ, 1998. - 1022 с.

[2] Айвазян С.А., Мхитарян В.С. Практикум по прикладной статистике и эконометрике (учебное пособие). - М.: МЭСИ, 1998. - 158 с.

[3] Анатольев С. Эконометрика для продолжающих: Курс лекций. - М.: РЭШ, 2002. - 60 с.

[4] Замков О.О. Эконометрические методы в макроэкономическом анализе: Курс лекций. - М.: ГУ ВШЭ, 2001. - 122 с.

Дополнительная литература

[5] Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Исследование зависимостей: Справ. Изд. - М.: Финансы и статистика, 1985. - 487 с.

[6] Джонстон Дж. Эконометрические методы. - М.: Статистика, 1980. - 446 с.

[7] Макаров В.Л., Айвазян С.А., Борисова С.В., Лакалин Э.А. Эконометрическое моделирование. Выпуск 2: Модель экономики России для целей краткосрочного прогноза и сценарного анализа (учебное пособие). - М.: МЭСИ, 2002. - 34 с.

[8] E.E.Leamer, "Let's Take the Con out of Econometrics", American Economic Review, 73 (1983), 31-43 с.

[9] Frish,R. "Editorial,"Econometrica, 1 (1933), 1-4 с.