МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ДАЛЬНЕВОСТОЧНАЯ ГОСУДАРСТВЕННАЯ АКАДЕМИЯ
ЭКОНОМИКИ И УПРАВЛЕНИЯ

Д.Б. Карп

Эконометрика: основные формулы с комментариями

Учебно-методическое пособие

Владивосток
Издательство ДВГАЭУ
2004

УДК 330.43+519.862

К26 **Карп Д.Б. Эконометрика: основные формулы с комментариями**: учебно-методическое пособие. - Владивосток: Изд-во ДВГАЭУ, 2004. - 50с.

Работа задумана как краткий справочник по основным формулам и алгоритмам, встречающимся в курсе эконометрики и не может заменить собой полноценного курса лекций по предмету. Тем не менее, каждый раздел содержит краткое изложение необходимой теории и подробные комментарии относительно применимости приводимых методов для обработки данных разной природы. Охват затронутых тем и разделов заметно превышает обычный семестровый курс и оставляет для преподавателя свободу выбора тем для практических занятий.

Во введении приведен краткий обзор типов данных и моделей, позволяющий читателю видеть излагаемые далее методы в более широком контексте. Внимательный читатель должен быть способен реализовать представленные методы на компьютере - вся необходимая для этого информация содержится в работе. Особенностью пособия является включение в него ряда новых формул и методов, некоторые из которых впервые публикуются на русском языке. Пособие предназначено для студентов специальностей 351400 - «Прикладная информатика в экономике» и 061700 - «Статистика», преподавателей соответствующих специальностей и работающих эконометристов.

Печатается по решению Учебно-методического совета в области междисциплинарных специальностей экономики и статистики ДВГАЭУ.

Рецензенты: Шмидт Ю.Д., д.э.н., зав. кафедрой математики и моделирования, проректор по научной работе ДВГАЭУ

Терешко Д.А., к.ф.м.н., старший научный сотрудник Института Прикладной Математики ДВО РАН

ISBN 5-93362-269-9

1 Введение

Термин «эконометрика» был введён в 1926 году норвежским экономистом Р. Фришем (Frisch) и в буквальном переводе означает «измерение экономики». Не существует общепринятого определения этой науки. С точки зрения математической статистики эконометрику можно считать одним из разделов многомерного статистического анализа. С точки зрения экономической теории эконометрика является методом (вернее набором методов), позволяющим придать точный количественный характер качественным экономическим зависимостям, постулируемым теорией. Таким образом можно сказать, что эконометрика представляет собой набор статистических методов, предназначенных для обработки экономических данных. Основная причина, по которой эти методы должны отличаться от методов статистической обработки данных других наук (медицины, биологии, психологии, и др.), состоит в том, что экономические данные не являются, как правило, экспериментальными - во-первых, мы не можем произвольно задать параметры (доходы, цены, ставки налогов, уровень инфляции или безработицы, и т. д.) и проследить за реакцией экономики (например увеличить вдвое цены на шоколад, оставив остальное без изменений); во-вторых, мы не можем проводить многократных экспериментов даже при постоянных значениях экономических показателей. Вместо дальнейших рассуждений приведём классический пример, демонстрирующий где пролегает граница между экономической теорией и эконометрикой.

Теория потребления Кейнса. В своей *Общей теории* (1936) Кейнс (Keynes) пишет:

> Мы определим, поэтому, то, что мы назовем склонностью к потреблению, как функциональное отношение f между данным уровнем дохода X и потреблением или объемом расходов C, соответствующим этому уровню дохода.
>
> Сумма, которую некоторое сообщество тратит на потребление, зависит (1) частично от его доходов, (2) частично от объективных сопутствующих обстоятельств и (3) от субъективных потребностей и психологических наклонностей и привычек индивидуумов его (сообщество) составляющих.
>
> Фундаментальный психологический закон, на которых мы можем с уверенностью опираться, как исходя из нашего знания человеческой природы, так и из подробных данных наблюдений, состоит в том, что люди предрасположены, как правило и в среднем, увеличивать свое потребление при увеличении дохода, но не

> на столько на сколько увеличился доход. То есть, производная dC/dX положительна и меньше единицы.
>
> Однако, если не учитывать краткосрочные изменения в уровне дохода, очевидно также, что более высокий абсолютный уровень дохода имеет тенденцию увеличивать разрыв между доходом и потреблением. По этой причине при увеличении дохода его часть, отводимая на сбережения, будет, как правило, также расти.

Теория, таким образом, постулирует зависимость между доходом и потреблением, $C = f(X)$ и утверждает, что мгновенная склонность к потреблению (MPC - marginal propensity to consume), равная dC/dX, лежит между 0 и 1. Последний параграф говорит, что средняя склонность к потреблению (APC - average propensity to consume), равная C/X, падает с увеличением дохода, то есть $\frac{d(C/X)}{dX} < 0$. По правилу дифференцирования дроби и учитывая, что $C = C(X)$, получим

$$\frac{d(C/X)}{dX} = \frac{C'X - C}{X^2} = \frac{C' - C/X}{X} = \frac{MPC - APC}{X} < 0.$$

Следовательно, $MPC < APC$.

Это рассуждение, типичное для экономической теории, служит основой для дальнейшего эконометрического исследования. Такое исследование, базируясь на данных наблюдений, призвано ответить на ряд вопросов, на которые не даёт ответа экономическая теория. Например, какова конкретная форма зависимости $C(X)$? Чему равны параметры этой зависимости для данного сообщества (например, если мы предположим, что зависимость линейная $C = aX + b$, то чему равны a и b)? Стабильны ли параметры этой зависимости во времени? Существуют ли систематические различия в виде этой зависимости между странами или регионами? Есть ли другие факторы, которые помогут нам лучше объяснить взаимосвязь между потреблением и доходом? Подтверждают ли данные положения теории?

Поведение экономики в целом складывается из поведения множества экономических агентов (фирм, финансовых институтов, семей, и.т.д.). Нет никакой надежды точно описать поведение каждого из этих агентов, находящихся в различных обстоятельствах и движимых различными мотивами. Поэтому эконометрические модели необходимо должны включать случайный, стохастический элемент, в отличие от детерминистских моделей экономической теории. [1].

[1] Впервые это убедительно показал норвежский экономист Тригве Хаавелмо (Trygve Haavelmo) в своей докторской диссертации 1944 года "The Probability Approach in Econometrics" (Вероятностный подход в эконометрике). Работа Хаавелмо оказала огромное влияние на дальнейшее развитие эконометрики и была удостоена Нобелевской премии по экономике 1987 года.

Наконец, экономическая теория относится к равновесному состоянию экономики или к так называемой «долгосрочной перспективе». Так, например, в равновесном состоянии потребность экономики в деньгах и предложение денег равны. В этом случае мы могли бы использовать данные о предложении денег вместо данных о потребности в деньгах. Однако, в действительности, денежный рынок почти никогда не находится в равновесии. При этом экономическая теория не говорит ничего о том, как экономика движется от одного равновесного состояния к другому. То есть, она не описывает процесс корректировки. К сожалению, реальные экономические данные обычно относятся к процессу корректировки, а не к последовательным положениям равновесия.

Обычно экономические данные можно отнести к одному из следующих трех типов:

- **Данные среза** (известные также как статические или пространственные данные) - это данные, относящиеся к одному моменту времени и дающие нечто вроде поперечного среза некоторой отрасли экономики (отсюда происходит термин cross-sectional data - дословно - данные поперечного сечения). К этому типу принадлежат, например, данные о ценах на автомобили или недвижимость в зависимости от их всевозможных характеристик и относящиеся к определенному моменту времени; данные опроса семей об их уровнях дохода, образования и потребления; или данные о курсах валют в в различных обменных пунктах города на какую-то фиксированную дату.

- **Временные ряды** - это наблюдения некоторых экономических показателей, относящиеся к последовательным моментам времени. Промежуток времени между наблюдениями чаще всего постоянный (ежедневные, ежемесячные, ежеквартальные или ежегодные данные), но может быть и переменным. К этому типу данных относятся, например, курс евро за последний месяц; ежеквартальные данные об уровне инфляции или безработицы в России за последние 5 лет; национальный доход или сальдо внешнеторгового баланса за последние 10 лет; уровень процентных ставок или индекс курса акций на бирже за последние 18 месяцев, и т.д.

- **Панельные данные** (или продольные данные) - это наблюдения за одной и той же группой экономических агентов, проведенные через определенные промежутки времени, то есть это набор срезов. Здесь мы имеем данные среза в динамике. Это могут быть данные ежегодных опросов избранной группы семей об их уровнях дохода и потребления; или ежеквартальный набор сведений (объем продаж, количество работников, прибыть, и т.д.) об избранной группе фирм.

Для описания экономических процессов эконометрика использует математические модели. Такие модели представляют собой уравнения или системы уравнений, призванные описать основные черты реального процесса, производящего наблюдаемые данные. Если модель адекватно представляет этот процесс, то её можно использовать для экономического анализа, для прогноза или для исследования реакции системы на изменение некоторых показателей. Можно выделить следующие основные типы моделей:

- **Статические регрессионные модели с одной зависимой переменной** - это модели вида

$$Y = f(X_1, X_2, \ldots, X_k, \varepsilon). \tag{1}$$

  Здесь переменные $X_1,\ldots, X_k$ принято называть *объясняющими* (поскольку мы пытаемся объяснить изменения зависимой переменной Y изменениями переменных $X_1,\ldots, X_k$), а ε - это случайная поправка, о которой говорилось выше. В типичном примере автомобильного рынка Y означает цену машины, X_1 - это год выпуска, X_2 - объем двигателя, X_3 - марка машины (которой нетрудно придать числовое значение просто кодируя различные марки: Toyota Corolla - 1, Toyota Corona - 2, Nissan Bluebird - 3, и т.д.), X_4 может быть тип трансмиссии (1 - коробка, 0 - автомат), а вид функции f может быть, например, следующим:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon. \tag{2}$$

  Здесь числа β_0, β_1, β_2, β_3, β_4 - это параметры модели, которые эконометрист должен оценить по имеющимся данным. Данные в нашем случае - это список выставленных на продажу машин вместе с интересующими нас характеристиками (ценами, объемами двигателей, и т.д.).

  Модели вида (1) подходят для моделирования данных среза или равновесных состояний экономики. Поскольку время явно не входит в уравнение (1), такая модель подразумевает мгновенную реакцию моделируемой системы на изменение значений объясняющих переменных.

- **Динамические модели с одной зависимой переменной** - это модели, в которые входят запаздывающие переменные, то есть значения экономических показателей, относящиеся к предыдущим моментам времени. Общий вид этих моделей таков:

$$
\begin{aligned}
Y_t =& f(Y_{t-1}, Y_{t-2}, \ldots, Y_{t-m}, \ldots, \\
& X_{1t}, X_{1(t-1)}, X_{1(t-2)}, \ldots, X_{1(t-m_1)}, \ldots, X_{kt}, X_{k(t-m_k)}, \varepsilon_t).
\end{aligned} \tag{3}
$$

Здесь $X_{n(t-i)}$ - значение переменной X_n в момент времени $t-i$, то есть запаздывающее на i периодов. Другой популярный термин для запаздывания это *лаг* (от англ. lag - задержка), i называют тогда длиной лага, а переменную $X_{n(t-i)}$ - переменной с лагом. Если в правую часть уравнения (3) не входят запаздывающие значения зависимой переменной, то есть

$$
Y_t = f(X_{1t}, X_{1(t-1)}, X_{1(t-2)}, \ldots, X_{1(t-m_1)}, \ldots, X_{kt}, \ldots, X_{k(t-m_k)}, \varepsilon_t), \tag{4}
$$

то модель называют моделью с *распределенным лагом.* Если же напротив запаздывающие значения зависимой переменной присутствуют, то говорят про *авторегрессивный лаг.*

Есть ряд причин, по которым запаздывающие значения переменных появляются в эконометрических уравнениях. Во-первых, они могут возникнуть по *технологическим* причинам. Например, при увеличении спроса на продукт производства некоторой фирмы, она не может мгновенно увеличить производство - требуется время на закупку материалов, а для существенного увеличения объёмов на установку дополнительного оборудования, и т.д. Во-вторых, *психологические факторы* приводят к задержкам. При изменении доходов, например, в силу потребительских привычек, расходы на потребления изменятся не сразу. В-третьих, запаздывание возникает в силу *несовершенной информации* - экономическим агентам требуется время для сбора информации, что задерживает принятие решений. В-четвертых, к запаздыванию могут приводить *институциональные факторы* - контрактные обязательства и т.п. Наконец, значение зависимой переменной в текущий момент времени очень часто зависит от *её значений в прошлом.*

Если наши данные имеют характер временного ряда, то нам необходимо использовать модель вида (3).

Если модель (1) описывает равновесное состояние экономики и часто непосредственно опирается на экономическую теорию, то модели (3), (4) описывают процесс корректировки. Если все значения объясняющих переменных долгое время не меняются, то система должна постепенно прийти в состояние равновесия. Пусть X_1 не меняется m_1

промежутков времени, то есть:

$$X_{1t} = X_{1(t-1)} = X_{1(t-2)} = \ldots = X_{1(t-m_1)} = X_1^{const}$$

Аналогично:

$$X_{2t} = X_{2(t-1)} = X_{2(t-2)} = \ldots = X_{2(t-m_2)} = X_2^{const},$$

и так далее. Тогда в состоянии равновесия модель (4) даст следующую связь между зависимой переменной и объясняющими переменными

$$Y^{const} = f(X_1^{const}, X_2^{const}, \ldots, X_k^{const}, \varepsilon).$$

Таким образом, мы пришли к статической модели вида (1). Эта статическая модель должна давать непротиворечащее экономической теории соотношение между переменными.

При наличии авторегрессивного лага связь между равновесной статической моделью и динамической моделью (3) проследить не так просто, поскольку все объясняющие переменные должны оставаться постоянными бесконечно долго, прежде чем Y_t также придет к равновесию.

Для примера рассмотрим две модели:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + \varepsilon_t \tag{5}$$

и

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + \varepsilon_t. \tag{6}$$

При изменении объясняющей переменной X на одну единицу изменение Y (или *отклик*) за один период равно в обеих моделях β_2. По этой причине, β_2 иногда называют *импульсным множителем.* В модели (5) Y измениться в следующем периоде еще на β_3 после чего будет оставаться постоянным до следующего изменения X. Таким образом, суммарное изменение Y при изменении X на одну единицу равно в модели с распределенным лагом (5) $\beta_2 + \beta_3$. Эту величину называют *долгосрочным или равновесным множителем.*

В модели с авторегрессивным лагом (6) Y_t измениться в общей сложности на величину

$$\beta_2 + \beta_2\beta_3 + \beta_2\beta_3^2 + \beta_2\beta_3^3 + \cdots$$

Если $|\beta_3| < 1$, то этот ряд сходиться и, по формуле для суммы бесконечной геометрической прогрессии, его сумма равна $\beta_2/(1 - \beta_3)$. Условие $|\beta_3| < 1$ является необходимым и достаточным для того, чтобы влияние единичного скачка X затухало во времени. Это условие *устойчивости* модели (6).

- **Системы одновременных уравнений.** Для описания некоторой отрасли экономики или, тем более, экономики в целом, одного уравнения, конечно, недостаточно. В этом случае модель состоит из набора уравнений, некоторые их которых могут быть регрессионными уравнениями вида (1) или (3) или тождествами (то есть не содержать случайной поправки, например Сбережения = Доходы - Расходы). Например, модель экономики США уортонской ассоциации содержала 207 уравнений. Аналогичный проект боннского университета для экономики Германии содержал 137 уравнений. Для исследования таких моделей требуется более сложный математический аппарат, чем для моделей первых двух типов и они не рассматриваются в настоящем пособии.

  В качестве примера рассмотрим следующую простейшую макроэкономическую модель :

$$\begin{aligned} C_t &= \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + \varepsilon_{1t} \\ I_t &= \beta_0 + \beta_1 R_t + \beta_2 (Y_t - Y_{t-1}) + \varepsilon_{2t} \\ Y_t &= C_t + I_t + G_t \end{aligned} \tag{7}$$

  Здесь потребление C_t, инвестиции I_t и суммарный спрос Y_t определяются из уравнений модели и называются *эндогенными* (внутренними, внутрисистемными) переменными. Уровень процентных ставок R_t и правительственные расходы, не связанные с заработной платой, G_t считаются заданными и называются *экзогенными* (внешними, внесистемными) переменными. Переменные C_{t-1} и Y_{t-1} к моменту времени t уже принимают некоторые значения и поэтому называются *предопределенными.*

Кроме перечисленных существуют и другие виды моделей, такие например, как модели для панельных данных, модели с дискретной зависимой переменной, модели с урезанными и цензурированными выборками и другие.

2 Множественная линейная регрессия

Простейшей и в тоже время важнейшей статической моделью является *множественная линейная регрессия.* Она задаётся уравнением:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_k X_k + \varepsilon, \tag{8}$$

где Y - зависимая переменная, X_1, ..., X_k - объясняющие переменные, ε - случайное возмущение. Пусть у нас имеются n наблюдений над всеми переменными. Сформируем из этих данных матрицу наблюдений $\mathbf{X}$ и вектор

$\mathbf{Y}$:

$$\mathbf{X} = \begin{bmatrix} 1 & X_{12} & X_{13} & \dots & X_{1k} \\ 1 & X_{22} & X_{23} & \dots & X_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n2} & X_{n3} & \dots & X_{nk} \end{bmatrix}, \qquad \mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}.$$

Через $\mathbf{x}_{\cdot m}$ мы будем обозначать m-ый столбец матрицы $\mathbf{X}$, то есть n-мерный вектор столбец наблюдений за переменной X_m; $\mathbf{x}_{i\cdot}$ будет обозначать i-ую строку матрицы $\mathbf{X}$, то есть k-мерный вектор-строку, содержащую i-ое наблюдение за всеми переменными. Согласно (8) $\mathbf{X}$ и $\mathbf{Y}$ связаны уравнением

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{9}$$

где $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)^{\mathrm{T}}$ и $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^{\mathrm{T}}$ - набор реализаций случайной величины ε. Классическая модель множественной линейной регрессии добавляет к уравнению (9) следующие предположения:

A1 Все столбцы матрицы $\mathbf{X}$ линейно независимы, то есть матрица $\mathbf{X}$ имеет ранг k. Это означает, что между объясняющими переменными не существует точной линейной зависимости. В противном случае, некоторые из переменных можно безболезненно удалить из модели.

A2 Случайные ошибки имеют нулевое математическое ожидание и некоррелированны с объясняющими переменными. Точнее предполагается, что при любой матрице наблюдений $\mathbf{X}$ условное математическое ожидание возмущений равно нулю: $E(\boldsymbol{\varepsilon}|\mathbf{X}) = \mathbf{0}$[2].

A3 Случайные возмущения для различных наблюдений не коррелируют друг с другом и имеют постоянную дисперсию для любой фиксированной матрицы наблюдений: $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}|\mathbf{X}) = \sigma^2\mathbf{I}$, где $\mathbf{I}$ - единичная матрица размера $n \times n$.

A4 Случайные возмущения распределены нормально со средним ноль и дисперсией σ^2: $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$.

Требуется оценить неизвестный вектор $\boldsymbol{\beta}$. Пусть у нас имеется некоторая оценка $\hat{\boldsymbol{\beta}}$ вектора $\boldsymbol{\beta}$. Тогда оценкой для возмущений служат остатки:

$$\hat{\varepsilon}_i = e_i = Y_i - \mathbf{x}_{i\cdot}\hat{\boldsymbol{\beta}}, \qquad i = 1, \dots, n, \tag{10}$$

которые образуют вектор

$$\mathbf{e} = (e_1, \dots, e_n)^{\mathrm{T}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{Y} - \hat{\mathbf{Y}}.$$

[2]Символ $E(\boldsymbol{\varepsilon}|\mathbf{X})$ означает условное математическое ожидание, то есть математическое ожидание в условном распределении $\boldsymbol{\varepsilon}$ при заданной матрице $\mathbf{X}$. На практике это условие означает, что значения объясняющих переменных не несут никакой информации о возмущениях.

Здесь $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ - прогноз значений $\mathbf{Y}$ по значениям $\mathbf{X}$. Метод наименьших квадратов состоит в таком выборе $\hat{\boldsymbol{\beta}}$, который минимизирует сумму квадратов остатков:

$$\sum_{i=1}^{n} e_i^2 = \mathbf{e}^{\mathrm{T}}\mathbf{e} \;\to\; \min .$$

Оценка методом наименьших квадратов имеет вид:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}. \tag{11}$$

Матрица $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ обратима благодаря предположению A1. Оценка методом наименьших квадратов будет несмещённой, если выполнено предположение A2. Эта оценка обладает наименьшей дисперсией среди всех линейных оценок вектора $\boldsymbol{\beta}$ если выполнены предположения A1-A3[3]. Наконец, оценка (11) эффективна, то есть обладает наименьшей дисперсией среди всех (а не только линейных) оценок вектора $\boldsymbol{\beta}$, если соблюдены все четыре предположения A1-A4. Условия состоятельности этой оценки приведены ниже в разделе 5. Теория статистического вывода, изложенная ниже в этом разделе и разделе 2, опирается на предположение A4 о нормальном распределении возмущений.

Для случая парной регрессии

$$Y = \alpha + \beta X + \varepsilon \tag{12}$$

формула (11) приводиться к виду:

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}, \qquad \hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}, \tag{13}$$

где $\overline{Y} = \left[\sum_{i=1}^{n} Y_i\right]/n$, $\overline{X} = \left[\sum_{i=1}^{n} X_i\right]/n$ - средние значения Y и X, соответственно.

Несмещенной оценкой для дисперсии σ^2 служит

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-k}\sum_{i=1}^{n} e_i^2 = \frac{\mathbf{e}^{\mathrm{T}}\mathbf{e}}{n-k}. \tag{14}$$

Матрица вариаций-ковариаций вектора $\hat{\boldsymbol{\beta}}$, равная $E(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^{\mathrm{T}})$, оценивается при помощи формулы:

$$\widehat{\mathbf{Var}}(\hat{\boldsymbol{\beta}}) = \widehat{E}(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^{\mathrm{T}}) = s^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}. \tag{15}$$

[3] Это утверждение составляет содержание **теоремы Гаусса-Маркова**.

Здесь $\widehat{E}$ - оценка математического ожидания. Диагональные элементы этой матрицы являются оценками для дисперсий отдельных коэффициентов, то есть:

$$s_{\hat{\beta}_j} = s\sqrt{(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}_{jj}}. \tag{16}$$

При помощи этих оценок строятся доверительные интервалы для индивидуальных коэффициентов. Зададим уровень доверия α. По заданному уровню доверия найдем критические значения t^{α}_{n-k} по таблице распределения Стьюдента с $n-k$ степенями свободы (в программе MS-Excel это можно сделать при помощи функции «=СТЬЮДРАСПОБР(α, $n-k$)»). Доверительный интервал для коэффициента β_j получаются по формуле:

$$\hat{\beta}_j - t^{\alpha}_{n-k} s_{\hat{\beta}_j} \leq \beta_j \leq \hat{\beta}_j + t^{\alpha}_{n-k} s_{\hat{\beta}_j}. \tag{17}$$

Статистика для проверки гипотезы $\beta_j = 0$ (t-тест) получается по формуле

$$t = \frac{\hat{\beta}_j}{s_{\hat{\beta}_j}} \tag{18}$$

и имеет распределение Стьюдента с $n-k$ степенями свободы. Соответственно, для проверки гипотезы $\beta_j = \beta^*_j$ можно использовать статистику

$$t^* = \frac{\hat{\beta}_j - \beta^*_j}{s_{\hat{\beta}_j}}, \tag{19}$$

имеющую такое же распределение. Доверительный интервал для дисперсии σ^2 находится по формуле

$$\frac{(n-k)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-k)s^2}{\chi^2_{1-\alpha/2}}, \tag{20}$$

где $\chi^2_{\alpha/2}$ и $\chi^2_{1-\alpha/2}$ - критические значения χ^2- распределения с $n-k$ степенями свободы.

Для исследования качества подгонки рассматривают соотношение между полной вариацией Y относительного своего среднего (SST=sum of squares total)

$$\mathrm{SST} = \sum_{i=1}^{n} (Y_i - \overline{Y})^2 = \mathbf{Y}^{\mathrm{T}}\mathbf{Y} - n\overline{Y}^2, \tag{21}$$

объясненной вариацией (SSE=sum of squares explained)

$$\mathrm{SSE} = \sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2 = \hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{Y} - n\overline{Y}^2 = \hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\hat{\boldsymbol{\beta}} - n\overline{Y}^2 \tag{22}$$

и остаточной вариацией (SSR=sum of squares residual)

$$\mathrm{SSR} = \mathbf{e}^{\mathrm{T}}\mathbf{e}. \tag{23}$$

Имеет место равенство:

$$\mathrm{SST} = \mathrm{SSE} + \mathrm{SSR}. \tag{24}$$

Коэффициент детерминации равен доли объясненной вариации Y в полной вариации:

$$R^2 = \frac{\mathrm{SSE}}{\mathrm{SST}} = 1 - \frac{\mathbf{e}^{\mathrm{T}}\mathbf{e}}{\mathbf{Y}^{\mathrm{T}}\mathbf{Y} - n\overline{Y}^2}. \tag{25}$$

Величина

$$s_y = \frac{\mathrm{SST}}{n-1} = \frac{\mathbf{Y}^{\mathrm{T}}\mathbf{Y} - n\overline{Y}^2}{n-1} \tag{26}$$

является несмещенной оценкой дисперсии Y. Коэффициент детерминации никогда не убывает при добавлении нового регрессора, даже не имеющего никакого отношения к объяснению движений зависимой переменной Y. Поэтому предпочтительней использовать *модифицированный коэффициент детерминации* $\tilde{R}^2$, в котором введен штраф за увеличение количества переменных:

$$\tilde{R}^2 = 1 - \frac{n-1}{n-k}(1 - R^2). \tag{27}$$

Коэффициенты детерминации (25), (27) имеют смысл *только если среди регрессоров есть константа (то есть, когда в матрице* $\mathbf{X}$ *есть столбец единиц)*. Для проверки значимости коэффициента детерминации (или, что то же самое, для проверки совместной значимости всех переменных) применяется F-статистика:

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}, \tag{28}$$

которая имеет F-распределение Фишера с $[k-1, n-k]$ степенями свободы в предположении о равенстве нулю всех коэффициентов при переменных: $\beta_2 = \beta_3 = \cdots = \beta_k = 0$. Следовательно, для проверки этой гипотезы на выбранном уровне значимости α нам необходимо сравнить наблюдаемое значение F-статистики с полученным по таблицам значением $F^{\alpha}_{k-1,n-k}$ (в MS-Excel это значение можно получить функцией «=FРАСПОБР(α, $k-1$, $n-k$)»). Высокие значения F-статистики свидетельствуют против гипотезы о равенстве нулю всех коэффициентов.

Необходимо иметь в виду, что для временных рядов характерны весьма высокие значения коэффициента детерминации (часто > 0.9). Такие значения могут быть результатом совместного тренда, если он присутствует как в зависимой переменной, так и в регрессорах, то есть являться следствием так называемой ложной корреляции (подробнее см. раздел 8 настоящего пособия). В то же время для данных среза значения $R^2 > 0.3$ могут быть вполне приемлемыми.

Для линейных комбинаций коэффициентов регрессии также можно строить доверительные интервалы. А именно, для $\gamma = \mathbf{w}^{\mathrm{T}}\boldsymbol{\beta}$, где $\mathbf{w}$ - заданный вектор, на уровне значимости α получаем доверительный интервал

$$\hat{\gamma} - t^{\alpha}_{n-k}s_{\hat{\gamma}} \leq \gamma \leq \hat{\gamma} + t^{\alpha}_{n-k}s_{\hat{\gamma}}, \tag{29}$$

где

$$\hat{\gamma} = \mathbf{w}^{\mathrm{T}}\hat{\boldsymbol{\beta}}, \qquad s^2_{\hat{\gamma}} = \mathbf{w}^{\mathrm{T}}s^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{w}.$$

Важным аспектом диагностики регрессионной модели (8) является проверка наличия *мультиколлинеарности* (то есть приближенной линейной зависимости между объясняющими переменными). Важность объясняется следующей формулой для дисперсии оценок коэффициентов:

$$\sigma^2_{\hat{\beta}_j} = \frac{\sigma^2}{(1 - R^2_j)\sum\limits_{i=1}^{n}(X_{ji} - \overline{\mathbf{x}}_{\cdot j})}, \tag{30}$$

где $\overline{\mathbf{x}}_{\cdot j}$ - среднее значение переменной X_j, то есть j-ого столбца матрицы $\mathbf{X}$, R^2_j - коэффициент детерминации при регрессии X_j по всем остальным объясняющим переменным. Эта формула показывает, что при значениях R^2_j близких к единице, дисперсия оценок $\hat{\beta}_j$ приближается к бесконечности, что приводит к размыванию доверительных интервалов (17) для коэффициентов регрессии и к низким значениям t-теста, при высокой совместной значимости переменных. Для проверки наличия мультиколлинеарности используется *коэффициент возрастания дисперсии* (VIF=Variance Inflation Factor):

$$\mathrm{VIF}_j = \frac{1}{1 - R^2_j}. \tag{31}$$

Еще один аспект диагностики - выявление выбросов. Для этого можно использовать оценку коэффициентов, полученную по всем наблюдениям, за исключением одного (i-ого):

$$\hat{\boldsymbol{\beta}}(i) = [\mathbf{X}(i)^{\mathrm{T}}\mathbf{X}(i)]^{-1}\mathbf{X}(i)^{\mathrm{T}}\mathbf{Y}(i), \tag{32}$$

где обозначение (i) показывает, что i-ое наблюдение было пропущено. Остаток, соответствующий пропущенному наблюдению, равен

$$e_i(i) = Y_i - \mathbf{x}_{i\cdot}\hat{\boldsymbol{\beta}}(i),$$

а его дисперсия оценивается по формуле:

$$\mathrm{Var}(e_i(i)) = s^2(i)(1 + \mathbf{x}_{i\cdot}[\mathbf{X}(i)^{\mathrm{T}}\mathbf{X}(i)]^{-1}\mathbf{x}_{i\cdot}^{\mathrm{T}}).$$

Нормированные остатки $e_i(i)/\sqrt{\mathrm{Var}(e_i(i))}$ должны иметь распределение близкое к стандартному нормальному распределению. Значения превышающие по модулю 2 указывают на наблюдения, требующие особого изучения (выбросы).

Приведем наконец, одну важную для вычислений формулу (формула обновления), которая применяется когда к имеющимся n наблюдениям добавляются еще m наблюдений и нам необходимо обновить оценку вектора коэффициентов $\boldsymbol{\beta}$. Пусть $\mathbf{X}_0(n \times k)$, $\mathbf{Y}_0(n \times 1)$ - первоначальные наблюдения и $\mathbf{A}_0 = (\mathbf{X}_0^{\mathrm{T}}\mathbf{X}_0)^{-1}$, $\mathbf{B}_0 = \mathbf{X}_0^{\mathrm{T}}\mathbf{Y}_0$. Пусть новые m наблюдений собраны в матрицу $\mathbf{X}_1(m \times k)$ и вектор $\mathbf{Y}_1(m \times 1)$. Обозначим через $\mathbf{X}((m+n) \times k)$, $\mathbf{Y}((m+n) \times 1)$ объединенные наблюдения. Имеют место формулы:

$$\begin{gathered}
\mathbf{A} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} = \mathbf{A}_0 - \mathbf{A}_0\mathbf{X}_1^{\mathrm{T}}[\mathbf{I} + \mathbf{X}_1\mathbf{A}_0\mathbf{X}_1^{\mathrm{T}}]^{-1}\mathbf{X}_1\mathbf{A}_0, \\
\mathbf{B} = \mathbf{X}^{\mathrm{T}}\mathbf{Y} = \mathbf{B}_0 + \mathbf{X}_1^{\mathrm{T}}\mathbf{Y}_1, \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y} = \mathbf{A}\mathbf{B}.
\end{gathered} \tag{33}$$

Общие сведения о множественной линейной регрессии завершим следующим замечанием. Для применения метода наименьших квадратов важна лишь линейность модели (8) по параметрам β_1, ..., β_k. Модель вида

$$g(Y) = \beta_1 + \beta_2 f_2(X_2, \ldots, X_k) + \cdots + \beta_k f_k(X_2, \ldots, X_k) + \varepsilon, \tag{34}$$

где g, f_2, ..., f_k - заданные функции, сводится к модели (8) при помощи введения новых переменных $Y^* = g(Y)$, $X_2^* = f_2(X_2, \ldots, X_k)$, ..., $X_k^* = f_k(X_2, \ldots, X_k)$. Кроме того, модель, зависящая от параметров θ_1, ..., θ_k называется *внутренне линейной*, если существует непрерывное взаимнооднозначное отображение набора параметров θ_1, ..., θ_k в набор параметров β_1, ..., β_k, относительно которых модель имеет вид (34).

3 Линейные ограничения, стабильность параметров и фиктивные переменные

Пусть нам необходимо проверить J линейных ограничений на модель (9):

$$\begin{aligned}
&r_{11}\beta_1 + r_{12}\beta_2 + \cdots + r_{1k}\beta_k = q_1 \\
&r_{21}\beta_1 + r_{22}\beta_2 + \cdots + r_{2k}\beta_k = q_2 \\
&\vdots \\
&r_{J1}\beta_1 + r_{22}\beta_2 + \cdots + r_{Jk}\beta_k = q_J.
\end{aligned}$$

Запишем ограничения в векторной форме:

$$\mathbf{R}\boldsymbol{\beta} = \mathbf{q}, \tag{35}$$

где

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & \dots & r_{1k} \\ r_{21} & r_{22} & r_{23} & \dots & r_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{J1} & r_{J2} & r_{J3} & \dots & r_{Jk} \end{bmatrix}, \qquad \mathbf{q} = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_J \end{bmatrix}.$$

В предположении, что ограничения (35) выполнены F-статистика, равная

$$F = (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})^{\mathrm{T}}[s^2\mathbf{R}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{R}^{\mathrm{T}}]^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})/J, \tag{36}$$

имеет F-распределение Фишера с $[J, n-k]$ степенями свободы. F-статистика из формулы (28) получается как частный случай (36), когда ограничения имеют вид $\beta_2 = \beta_3 = \dots = \beta_k = 0$, то есть когда $\mathbf{R}$ - единичная матрица размера $(k-1) \times (k-1)$ с приписанным слева столбцом нулей, а $\mathbf{q} = 0$ (см. формулу (40)).

При помощи F-статистики мы можем строить доверительные области для группы коэффициентов. Пусть p - количество коэффициентов, для которых строится доверительная область. Эта область состоит из тех значений β'_{i_1}, β'_{i_2}, ... , β'_{i_p}, для которых гипотеза $\beta_{i_1} = \beta'_{i_1}$, $\beta_{i_2} = \beta'_{i_2}$, ..., $\beta_{i_p} = \beta'_{i_p}$ не может быть отвергнута F-статистикой на заданном уровне значимости α. Обозначим $\mathbf{i} = \{i_1, i_2, \dots, i_p\}$, $\boldsymbol{\beta}'_{\mathbf{i}} = (\beta'_{i_1}, \beta'_{i_2}, \dots, \beta'_{i_p})^{\mathrm{T}}$, $\hat{\boldsymbol{\beta}}_{\mathbf{i}} = (\hat{\beta}_{i_1}, \dots, \hat{\beta}_{i_p})^{\mathrm{T}}$. Найдем по таблице распределения Фишера с $[p, n-k]$ степенями свободы критическое значение $F^{\alpha}_{p,n-k}$. Доверительная область состоит из значений $\boldsymbol{\beta}'_{\mathbf{i}}$, удовлетворяющих неравенству

$$\frac{1}{p}(\hat{\boldsymbol{\beta}}_{\mathbf{i}} - \boldsymbol{\beta}'_{\mathbf{i}})^{\mathrm{T}}\widehat{\mathbf{Var}}_{\mathbf{i}}(\hat{\boldsymbol{\beta}})^{-1}(\hat{\boldsymbol{\beta}}_{\mathbf{i}} - \boldsymbol{\beta}'_{\mathbf{i}}) \le F^{\alpha}_{p,n-k}, \tag{37}$$

где $\widehat{\mathbf{Var}}_{\mathbf{i}}(\hat{\boldsymbol{\beta}})$ - матрица размера $p \times p$, состоящая из тех строк и столбцов матрицы (15), которые соответствуют ковариациям компонент $\hat{\boldsymbol{\beta}}_{\mathbf{i}}$. Доверительная область (37) имеет форму многомерного эллипсоида.

При наложении ограничений (35) на модель (9) метод наименьших квадратов с ограничениями приводит к следующему вектору оценок:

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{R}^{\mathrm{T}}[\mathbf{R}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{R}^{\mathrm{T}}]^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q}). \tag{38}$$

Матрица вариаций-ковариаций этого вектора оценок имеет вид:

$$\mathbf{Var}(\tilde{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} - \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{R}^{\mathrm{T}}[\mathbf{R}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{R}^{\mathrm{T}}]^{-1}\mathbf{R}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} \tag{39}$$

и может быть оценена заменой σ^2 на s^2. Эта матрица меньше матрицы (15) в том смысле, что получается из нее вычитанием положительно определённой матрицы.

Обозначим через $\tilde{\mathbf{e}}$ вектор остатков в регрессии с ограничениями (35):

$$\tilde{\mathbf{e}} = \mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\beta}}.$$

Тогда F-статистика (36) может быть также выражена формулой:

$$F = \frac{(\tilde{\mathbf{e}}^{\mathrm{T}}\tilde{\mathbf{e}} - \mathbf{e}^{\mathrm{T}}\mathbf{e})/J}{\mathbf{e}^{\mathrm{T}}\mathbf{e}/(n-k)} = \frac{(R^2 - R_1^2)/J}{(1-R^2)/(n-k)}, \tag{40}$$

где $\mathbf{e}$ и R^2 - вектор остатков и коэффициент детерминации в регрессии (9) без наложения ограничений (35), соответственно, R_1^2- коэффициент детерминации в регрессии (9) с ограничениями (35).

Одно из распространенных применений F-статистики это тестирование гипотезы о равенстве нулю некоторого подмножества коэффициентов. В этом случае регрессия с ограничениями - это регрессия в которой просто отсутствуют переменные, коэффициенты при которых предполагаются равными нулю. Вектор остатков этой регрессии и есть в этом случае $\tilde{\mathbf{e}}$ и можно воспользоваться формулой (40). Если вычисленная таким образом F-статистика превосходит критическое значение на заданном уровне значимости, то гипотеза о равенстве нулю соответствующего подмножества коэффициентов должна быть отвергнута.

Другое применение F-статистики - проверка стабильности параметров регрессии. Такая проверка применяется к временным рядам, когда имеется подозрение на структурный сдвиг, приведший к изменению значений параметров модели и произошедший в период наблюдений. Разобъем наши данные на два подмножества: $\mathbf{X}_1$, $\mathbf{Y}_1$ - до момента возможного структурного сдвига и $\mathbf{X}_2$, $\mathbf{Y}_2$ - после такого момента. Тогда в модели без ограничений вектор параметров может быть различным при $\mathbf{X}_1$ и при $\mathbf{X}_2$. Это соответствует тому, что мы оцениваем две различных модели - одну на данных до момента возможного структурного сдвига, вторую - на данных после такого момента:

$$\hat{\boldsymbol{\beta}}_1 = (\mathbf{X}_1^{\mathrm{T}}\mathbf{X}_1)^{-1}\mathbf{X}_1^{\mathrm{T}}\mathbf{Y}_1^{\mathrm{T}}, \quad \hat{\boldsymbol{\beta}}_2 = (\mathbf{X}_2^{\mathrm{T}}\mathbf{X}_2)^{-1}\mathbf{X}_2^{\mathrm{T}}\mathbf{Y}_2^{\mathrm{T}}. \tag{41}$$

Сумма квадратов остатков в регрессии без ограничений складывается из двух частей, соответствующих двум периодам:

$$\mathbf{e}^{\mathrm{T}}\mathbf{e} = \mathbf{e}_1^{\mathrm{T}}\mathbf{e}_1 + \mathbf{e}_2^{\mathrm{T}}\mathbf{e}_2. \tag{42}$$

Ограничение, соответствующее гипотезе об отсутствии структурного сдвига, имеет вид $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ или $\mathbf{R} = [\mathbf{I}, -\mathbf{I}]$, $\mathbf{q} = \mathbf{0}$, $\mathbf{R}\boldsymbol{\beta} = \mathbf{q}$, и можно применить формулу (36). Проще однако заметить, что регрессия с ограничениями в данном случае - это просто регрессия, в которой все данные объединены:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix}, \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}, \quad \tilde{\mathbf{e}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}. \tag{43}$$

Теперь можно применить формулу (40) с J равным количеству столбцов в $\mathbf{X}$ и $n - k = n_1 + n_2 - 2k$ (поскольку в регрессии без ограничений $n_1 + n_2$ наблюдений и $2k$ переменных).

Если необходимо проверить гипотезу о том, что структурный сдвиг повлиял не на все коэффициенты, а лишь на некоторое их подмножество (против альтернативы, что все коэффициенты изменились), то регрессия без ограничений снова состоит из двух независимых частей (41) с суммой остатков (42), а в регрессии с ограничениями матрица наблюдений имеет вид:

$$\mathbf{X}_R = \begin{bmatrix} \mathbf{Z}_{pre} & \mathbf{0} & \mathbf{W}_{pre} \\ \mathbf{0} & \mathbf{Z}_{post} & \mathbf{W}_{post} \end{bmatrix}, \tag{44}$$

где индексы pre и $post$ - обозначают наблюдения до и после момента структурного сдвига, соответственно; $\mathbf{Z}_{pre}$, $\mathbf{Z}_{post}$ - наблюдения за переменными, на коэффициенты при которых не налагается ограничений (то есть они могут быть различными до и после сдвига), $\mathbf{W}_{pre}$, $\mathbf{W}_{post}$ - наблюдения за переменными, коэффициенты при которых нам необходимо протестировать на неизменность. Если константа может быть различной до и после момента структурного сдвига, то столбец единиц $\mathbf{i}$ входит в группы $\mathbf{Z}_{pre}$, $\mathbf{Z}_{post}$ (если есть подозрение, что изменилась *только константа*, то $\mathbf{Z}_{pre} = \mathbf{i}_{n_1}$, $\mathbf{Z}_{post} = \mathbf{i}_{n_2}$); если необходимо протестировать осталась ли константа постоянной, то столбцы единиц подходящего размера входят в $\mathbf{W}_{pre}$, $\mathbf{W}_{post}$. Остатки от регрессии $\mathbf{Y}$ на матрицу $\mathbf{X}_R$ необходимо подставить в формулу (40) вместо $\tilde{\mathbf{e}}$. При этом J - это количество столбцов в $\mathbf{W}_{pre}$ (= количество столбцов в $\mathbf{W}_{post}$).

Если мы проверяем гипотезу о стабильности некоторой группы коэффициентов против альтернативы о стабильности всех коэффициентов, то (44) играет роль модели без ограничений, в то время как модель с ограничениями задается уравнением (43), в котором все параметры неизменны весь период наблюдений.

Наконец в том случае, когда наблюдений в одной из групп $\mathbf{X}_1$ или $\mathbf{X}_2$ (пусть для определенности в $\mathbf{X}_2$) недостаточно для расчета коэффициентов в модели без ограничений $(n_2 < k)$[4], применяется тест Чау на провал прогноза (Chow predictive failure test). Он состоит в следующем. Регрессия

[4]То есть ранг матрицы $\mathbf{X}_2^{\mathrm{T}}\mathbf{X}_2$ меньше k и она необратима.

с ограничениями проводится с использованием полного набора данных, то есть согласно (43). Остатки в этой регрессии обозначим $\tilde{\mathbf{e}}$. Модель без ограничений рассчитывается на данных из более длинной группы $\mathbf{X}_1$, то есть по первой из формул (41). Это делается потому, что на данных из второй группы мы можем выбрать $\hat{\boldsymbol{\beta}}_2$ так, чтобы остатки были равны нулю, поскольку оцениваемых параметров больше чем наблюдений. Обозначив остатки в этой регрессии через $\mathbf{e}$, вычислим F-статистику по формуле:

$$F = \frac{(\tilde{\mathbf{e}}^{\mathrm{T}}\tilde{\mathbf{e}} - \mathbf{e}^{\mathrm{T}}\mathbf{e})/n_1}{\mathbf{e}^{\mathrm{T}}\mathbf{e}/(n_2 - k)} \qquad (45)$$

и сравним наблюдаемое значение с критическим по таблице F-распределения с $[n_1, n_2 - k]$ степенями свободы.

Приведённые выше тесты применяются в тех случаях, когда момент возможного структурного сдвига известен заранее. Если это не так, можно использовать CUSUM (cumulative sum) тест Брауна-Дёрбина-Эванса (Brown-Durbin-Evans), основанный на *рекурсивных остатках*. Пусть $\hat{\boldsymbol{\beta}}_{t-1}$ - оценка вектора $\boldsymbol{\beta}$ методом наименьших квадратов, построенная по первым $t-1$ наблюдениям ($t-1 \geq k$). Тогда рекурсивный остаток e_t равен ошибке прогноза значения Y_t:

$$e_t = Y_t - \mathbf{x}_{t\cdot}\hat{\boldsymbol{\beta}}_{t-1},$$

где $\mathbf{x}_{t\cdot}$, как прежде, наблюдение с номером t, то есть t-ая строка матрицы $\mathbf{X}$. В предположении о постоянстве параметров модели весь период наблюдений нормированные рекурсивные остатки

$$w_t = \frac{e_t}{\sqrt{1 + \mathbf{x}_{t\cdot}^{\mathrm{T}}(\mathbf{X}_{t-1}^{\mathrm{T}}\mathbf{X}_{t-1})^{-1}\mathbf{x}_{t\cdot}}}$$

имеют нормальное распределение со средним 0 и дисперсией σ^2 и независимы. В последней формуле матрица $\mathbf{X}_{t-1}((t-1)\times k)$ - это первые $t-1$ строк матрицы $\mathbf{X}$. Оценим дисперсию выражением:

$$\hat{\sigma}^2 = \frac{1}{n-k-1}\sum_{t=k+1}^{n}(w_t - \overline{w})^2, \quad \overline{w} = \frac{1}{n-k}\sum_{t=k+1}^{n} w_t.$$

Тогда величины $w_t/\hat{\sigma}$ имеют распределение близкое с стандартному нормальному и некоррелированны. Можно использовать график $w_t/\hat{\sigma}$ в зависимости от t как одно из средств проверки нормальности. Выход за границы диапазона $(-2, 2)$ свидетельствует в пользу нарушения гипотезы о стабильности параметров модели. Кумулятивные суммы получаются по формуле

$$W_s = \sum_{t=k+1}^{s}\frac{w_t}{\hat{\sigma}}.$$

В предположении о постоянстве параметров модели весь период наблюдений величины W_s имеют распределение близкое к нормальному со средним ноль и дисперсией $s-k$. График W_s как функции s должен оставаться в коридоре, задаваемом прямыми, соединяющими точки $[k, \pm a(n-k)^{1/2}]$ с точками $[k, \pm 3a(n-k)^{1/2}]$, где значения a зависят от уровня значимости и равны 0.948 и 1.143 для уровней доверия 95% и 99%, соответственно.

Наконец Хансен (Hansen) предложил в 1992 году следующий тест на стабильность параметров модели (8). Пусть

$$\mathbf{f}_t = [\mathbf{x}_t, e_t^2 - \overline{e^2}]^{\mathrm{T}}, \qquad \overline{e^2} = \frac{1}{n}\sum_{t=1}^{n} e_t^2.$$

Тогда поскольку

$$\mathbf{X}^{\mathrm{T}}\mathbf{e} = \sum_{t=1}^{n} \mathbf{x}_{t\cdot} e_t = \mathbf{0}, \qquad \sum_{t=1}^{n}(e_t^2 - \overline{e^2}) = 0,$$

получаем

$$\sum_{t=1}^{n} \mathbf{f}_t = \mathbf{0}.$$

Обозначим:

$$\mathbf{s}_t = \sum_{r=1}^{t} \mathbf{f}_r, \qquad \mathbf{F} = \frac{1}{n}\sum_{t=1}^{n} \mathbf{f}_t\mathbf{f}_t^{\mathrm{T}}, \qquad \mathbf{S} = \sum_{t=1}^{n} \mathbf{s}_t\mathbf{s}_t^{\mathrm{T}}.$$

Статистика Хансена вычисляется по формуле[5]

$$H = \mathrm{tr}(\mathbf{F}^{-1}\mathbf{S}).$$

Высокие значения H свидетельствуют против нулевой гипотезы о стабильности параметров. Асимптотические критические значения на уровне доверия 95% следующие: при $k = 1$ - 1.01; при $k = 5$ - 1.68; $k = 15$ - 3.54; $k = 20$ - 4.52.

Фиктивные переменные используются для описания качественных факторов в регрессионных моделях. Так например, для квартальных данных мы можем учесть влияние времени года на параметр β_1 в модели (8), добавив в число регрессоров переменную D_1, равную единице для всех наблюдений, относящихся к первому кварталу, и нулю для остальных кварталов, переменную D_2, равную единице для всех наблюдений, относящихся ко второму кварталу, и нулю для остальных кварталов и переменную D_3, равную единице для всех наблюдений, относящихся к третьему кварталу,

[5]Напомним, что $\mathrm{tr}(\mathbf{A})$ - след матрицы $\mathbf{A}$ - это сумма ее диагональных элементов

и нулю для остальных кварталов. *Мы не должны вводить такую переменную для четвёртого квартала, поскольку это привело бы к линейной зависимости столбцов матрицы* $\mathbf{X}$ *($D_1 + D_2 + D_3 + D_4 = 1$ для всех наблюдений), то есть к нарушению предположения A1 и необратимости матрицы* $(\mathbf{X}^{\mathrm{T}}\mathbf{X})$! Мы можем проверить существенность сезонных эффектов по значимости коэффициентов при сезонных фиктивных переменных. При помощи F-статистики можно проверить гипотезу об отсутствии сезонных эффектов, то есть о равенстве нулю всех коэффициентов при сезонных переменных одновременно. Если мы хотим учесть влияние качественного фактора не только на сдвиг β_1, но и на некоторый наклон β_j, мы можем добавить в число регрессоров DX_j, где D - мультипликативная фиктивная переменная, равная 1 для наблюдений с выделенным значением качественного фактора и 0 для остальных наблюдений.

4 Инструментальные переменные

Важнейшим предположением, благодаря которому оценка МНК (11) обладает ключевыми свойствами состоятельности и несмещенности является предположение A2 о некоррелированности объясняющих переменных X_i и возмущения ε[6]. Часто однако, это предположение не выполняется. Наиболее распространенными причинами являются ошибки измерения и принадлежность рассматриваемого регрессионного уравнения к системе одновременных уравнений. Для оценивания параметров в таких случаях применяется *метод инструментальных переменных.* Он состоит в следующем. Пусть у нас есть $l \geq k$ переменных Z_i, $i = 1, \ldots, l$, некоррелированных с ε, но коррелированных с X_i. Для тех переменных X_j, которые некоррелированы с ε можно взять $Z_j = X_j$. Для других переменных в качестве инструментов часто выбирают их значения в предыдущем периоде: $Z_{it} = X_{i(t-1)}$. Пусть $\mathbf{Z}(n \times l)$ - матрица наблюдений над инструментальными переменными. Предположим сначала $l = k$. Тогда оценка вектора параметров $\boldsymbol{\beta}$ модели (8) методом инструментальных переменных равна:

$$\hat{\boldsymbol{\beta}}_{\mathrm{IV}} = (\mathbf{Z}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{Y}. \tag{46}$$

Состоятельная оценка дисперсии возмущений σ^2 может быть, как прежде, основана на сумме квадратов остатков:

$$\hat{\sigma}^2 = \frac{1}{n}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{IV}})^{\mathrm{T}}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{IV}}). \tag{47}$$

[6]Точнее предполагается, что условное математическое ожидание $E(\varepsilon|\mathbf{X})$ равно нулю для любых наблюдений $\mathbf{X}$. Некоррелированность является следствием этого предположения.

Коррекция знаменателя для учета степеней свободы здесь не требуется, поскольку оценка в любом случае будет смещённой и необходимо опираться лишь на её асимптотические свойства. Асимптотическая матрица вариаций-ковариаций вектора $\hat{\boldsymbol{\beta}}_{\mathrm{IV}}$ оценивается формулой:

$$\widehat{\mathbf{AVar}}(\hat{\boldsymbol{\beta}}_{\mathrm{IV}}) = \hat{\sigma}^2(\mathbf{Z}^{\mathrm{T}}\mathbf{X})^{-1}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})(\mathbf{X}^{\mathrm{T}}\mathbf{Z})^{-1}. \tag{48}$$

Эту формулу можно использовать для построения доверительных интервалов для отдельных компонент вектора $\boldsymbol{\beta}$, для этого матрица (48) используется вместо матрицы (15) в формулах (16) и (17). Для построения доверительных областей для группы компонент необходимо в формуле (37) заменить матрицу $\widehat{\mathbf{Var_i}}(\hat{\boldsymbol{\beta}})$ на соответствующую подматрицу матрицы (48).

Пусть теперь инструментальных переменных больше чем первоначальных регрессоров: $l > k$. В этом случае новый набор инструментов, содержащий ровно k регрессоров получается по формуле:

$$\hat{\mathbf{X}} = \mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{X}. \tag{49}$$

В этом выражении $\hat{\mathbf{X}}$ можно интерпретировать как набор прогнозов значений столбцов $\mathbf{X}$ (то есть значений переменных X_i, $i = 2, \ldots, k$) при регрессии каждого столбца $\mathbf{X}$ на все переменные из матрицы $\mathbf{Z}$. Если какие-то переменные из $\mathbf{X}$ содержатся в $\mathbf{Z}$ (то есть некоторое подмножество столбцов $\mathbf{X}$ повторяется в $\mathbf{Z}$), то эти переменные будут без изменения воспроизведены в $\hat{\mathbf{X}}$. Оценка $\hat{\boldsymbol{\beta}}_{\mathrm{IV}}$ вычисляется теперь по любой из трех эквивалентных формул:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\mathrm{IV}} =&(\hat{\mathbf{X}}^{\mathrm{T}}\mathbf{X})^{-1}\hat{\mathbf{X}}^{\mathrm{T}}\mathbf{Y} = \\ =&[\mathbf{X}^{\mathrm{T}}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{Y} \\ &= (\hat{\mathbf{X}}^{\mathrm{T}}\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}^{\mathrm{T}}\mathbf{Y}.\end{aligned} \tag{50}$$

Асимптотическая матрица вариаций-ковариаций вектора $\hat{\boldsymbol{\beta}}_{\mathrm{IV}}$ снова вычисляется по формуле (48). Асимптотическое распределение вектора $\hat{\boldsymbol{\beta}}_{\mathrm{IV}}$ является нормальным со средним $\boldsymbol{\beta}$ и матрицей вариаций-ковариаций (48). Следовательно, для построения приближенных доверительных областей можно пользоваться таблицами многомерного нормального распределения.

В тех случаях, когда наличие корреляции между объясняющими переменными и возмущениями не очевидно из свойств модели, можно использовать следующий тест Ву (Wu, 1973) на наличие такой корреляции. Пусть $\mathbf{X}^*$ - наблюдения над k^* переменными подозрительными на корреляцию с возмущениями (то есть $\mathbf{X}^*$ некоторое подмножество столбцов матрицы $\mathbf{X}$). Пусть

$$\hat{\mathbf{X}}^* = \mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{X}^*$$

- прогнозируемые значения $\mathbf{X}^*$ при регрессии на $\mathbf{Z}$. Составим расширенную регрессию:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \hat{\mathbf{X}}^*\boldsymbol{\gamma} + \boldsymbol{\varepsilon}^*. \qquad (51)$$

Совместная значимость всех компонент вектора $\boldsymbol{\gamma}$ свидетельствует в пользу корреляции между переменными из $\mathbf{X}^*$ и первоначальными возмущениями $\boldsymbol{\varepsilon}$. Для проверки значимости используется F-статистика (40) с $[k^*, n-k-k^*]$ степениями свободы.

5 Несферические возмущения - общий случай

При рассмотрении регрессионой модели (8) мы предполагали, что возмущения, соответствующие различным наблюдениям, обладают одинаковой дисперсией и некоррелированны друг с другом, то есть $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}|\,\mathbf{X}) = \sigma^2\mathbf{I}$ для любой матрицы наблюдений $\mathbf{X}$ (предположение A3). Отказ от этого требования приводит к модели (8) с дополнительным условием:

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}|\,\mathbf{X}) = \sigma^2\boldsymbol{\Omega}, \qquad (52)$$

где $\boldsymbol{\Omega}$ - некоторая положительно-определённая матрица[7]. Модель (8) с дополнительным условием (52) называется *обобщенной линейной регрессией*. Два важнейших частных случая этой модели это **гетероскедастичность** и **автокорреляция** возмущений. Возмущения гетероскедастичны, если матрица $\sigma^2\boldsymbol{\Omega}$ имеет вид:

$$\sigma^2\boldsymbol{\Omega} = \sigma^2 \begin{bmatrix} \omega_1 & 0 & \dots & 0 \\ 0 & \omega_2 & \dots & 0 \\ & & \vdots & \\ 0 & 0 & \dots & \omega_n \end{bmatrix}, \qquad (53)$$

то есть возмущения обладают различными дисперсиями $\sigma_i^2 = \sigma^2\omega_i$, но не коррелированны друг с другом. Для удобства веса ω_i и среднюю дисперсию σ^2 часто выбирают так, чтобы

$$\mathrm{tr}(\boldsymbol{\Omega}) = \sum_{i=1}^{n} \omega_i = n. \qquad (54)$$

При такой нормировке классической гомоскедастичной модели соответствуют $\omega_i = 1$, $i = 1, \dots, n$. Подробнее этот случай рассмотрен ниже.

[7]Напомним,что матрица $\mathbf{A}$ называется положительно-определенной если $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} > 0$ для любого вектора $\mathbf{x}$.

Автокорреляция возмущений встречается в основном во временных рядах. Наиболее часто встречается вид автокорреляции возмущений, задаваемый матрицей

$$\sigma^2\mathbf{\Omega} = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{n-1} \\ \rho_1 & 1 & \dots & \rho_{n-1} \\ & & \vdots & \\ \rho_{n-1} & \rho_{n-2} & \dots & 1 \end{bmatrix}. \tag{55}$$

Такой вид матрица вариаций-ковариаций возмущений будет иметь всегда, когда корреляция между возмущениями зависит лишь от временного промежутка между ними, но не от самого момента времени. В этом случае ρ_i - это корреляция между возмущениями, отстоящими друг от друга на i промежутков времени. Большинство реальных процессов обладает затухающей памятью, что приводит к убыванию значений ρ_i при увеличении i.

Одновременно автокорреляция и гетероскедастчность встречается при исследовании панельных данных - то есть данных среза в динамике.

Оценка (11) вектора $\boldsymbol{\beta}$ методом наименьших квадратов в обобщенной модели (8)+(52) сохраняет свойство несмещенности. Будет ли она состоятельной зависит от выполнения ряда условий. В частности, если

$$\text{plim}(\mathbf{X}^{\mathrm{T}}\mathbf{X}/n) = \mathbf{Q}_1$$

и

$$\text{plim}(\mathbf{X}^{\mathrm{T}}\mathbf{\Omega}\mathbf{X}/n) = \mathbf{Q}_2$$

- конечные положительно- определенные матрицы, то $\hat{\boldsymbol{\beta}}$ - состоятельная оценка вектора $\boldsymbol{\beta}$[8]. Альтернативный набор условий состоятельности следующий:

а) наименьшее (по модулю) собственное значение матрицы $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ неограниченно возрастает при $n \to \infty$; и

б) наибольшее (по модулю) собственное значение $\mathbf{\Omega}$ конечно при всех n.

Первое условие будет выполняться, если с ростом n сумма квадратов значений наблюдений над каждой переменной $\sum_{i=1}^{n} X_{ji}^2$ неограниченно возрастает, при этом ни одно наблюдение не доминирует над остальным:

$$X_{jm}/\sum_{i=1}^{n} X_{ji}^2 \to 0$$

при $n \to \infty$ для каждого j и m. Второе условие означает конечность дисперсий в гетероскедастичной модели (53) и достаточно быстрое убывание автокорреляций при удалении от главной диагонали в модели (55).

[8]Обозначение plim означает предел по вероятности.

Даже если оценка методом наименьших квадратов $\hat{\boldsymbol{\beta}}$ состоятельна для модели (8)+(52), она уже не будет обладать наименьшей дисперсией ни среди всех, ни даже среди линейных оценок вектора $\boldsymbol{\beta}$. Еще более серьезное последствие замены предположения A3 на $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}|\,\mathbf{X}) = \sigma^2\boldsymbol{\Omega}$ в том, что перестают выполняться оценка (15) матрицы вариаций-ковариаций $\hat{\boldsymbol{\beta}}$ и оценка (14) для дисперсии возмущений. Следовательно, все использованные нами в разделах 1 и 2 процедуры статистического вывода становятся неверными.

Матрица вариаций-ковариаций оценки МНК $\hat{\boldsymbol{\beta}}$ для модели (8)+(52) равна

$$\mathbf{V}_{\mathrm{OLS}} = \frac{1}{n}\left(\frac{1}{n}\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\left(\frac{1}{n}\mathbf{X}^{\mathrm{T}}[\sigma^2\boldsymbol{\Omega}]\mathbf{X}\right)\left(\frac{1}{n}\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}. \tag{56}$$

Для эффективного оценивания обобщённой регрессионой модели применим спектральное разложение положительно-определенной матрицы $\boldsymbol{\Omega}$:

$$\boldsymbol{\Omega} = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}^{\mathrm{T}},$$

где столбцы матрицы $\mathbf{C}$ - собственные векторы $\boldsymbol{\Omega}$, а $\boldsymbol{\Lambda}$ - диагональная матрица с собственными значениями $\boldsymbol{\Omega}$ на главной диагонали. Эти собственные значения положительны и поэтому определена диагональная матрица $\boldsymbol{\Lambda}^{-1/2}$, состоящая из элементов матрицы $\boldsymbol{\Lambda}$ в степени $-1/2$. Обозначим

$$\mathbf{P}^{\mathrm{T}} = \mathbf{C}\boldsymbol{\Lambda}^{-1/2}. \tag{57}$$

Тогда $\boldsymbol{\Omega}^{-1} = \mathbf{P}^{\mathrm{T}}\mathbf{P}$. Умножим (8) на $\mathbf{P}$:

$$\mathbf{PY} = \mathbf{PX}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{\varepsilon}$$

или, вводя обозначения $\mathbf{Y}_* = \mathbf{PY}$, $\mathbf{X}_* = \mathbf{PX}$, $\boldsymbol{\varepsilon}_* = \mathbf{P}\boldsymbol{\varepsilon}$:

$$\mathbf{Y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_*. \tag{58}$$

Матрица вариаций-ковариаций возмущений в преобразованной модели (58) равна:

$$E(\boldsymbol{\varepsilon}_*\boldsymbol{\varepsilon}_*^{\mathrm{T}}) = \mathbf{P}\sigma^2\boldsymbol{\Omega}\mathbf{P}^{\mathrm{T}} = \sigma^2\mathbf{I},$$

то есть возмущения $\boldsymbol{\varepsilon}_*$ из (58) удовлетворяют требованиям A2, A3 классической регрессионой модели. Более того, если $\boldsymbol{\varepsilon}$ имел многомерное нормальное распределение с матрицей вариаций-ковариаций $\boldsymbol{\Omega}$, то $\boldsymbol{\varepsilon}_*$ имеет стандартное нормальное распределение, то есть удовлетворяет также требованию A4. Поэтому эффективная оценка вектора $\boldsymbol{\beta}$ получается по формуле:

$$\boldsymbol{\beta}_* = (\mathbf{X}_*^{\mathrm{T}}\mathbf{X}_*)^{-1}\mathbf{X}_*^{\mathrm{T}}\mathbf{Y}_* = (\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{Y}. \tag{59}$$

Она называется оценкой *обобщённым методом наименьших квадратов* (ОМНК) или оценкой Айткена (Aitken, 1935). Оценка (59) может быть получена минимизацией обобщённой суммы квадратов остатков:

$$\mathbf{e}_*^{\mathrm{T}}\mathbf{e}_* = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\Omega}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \ \rightarrow \ \min .$$

Матрица вариаций-ковариаций этой оценки находится по формуле

$$\mathbf{Var}(\boldsymbol{\beta}^*) = \sigma^2(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}, \tag{60}$$

а дисперсия σ^2 может быть оценена выражением:

$$s_*^2 = \frac{\mathbf{e}_*^{\mathrm{T}}\mathbf{e}_*}{n-k} = \frac{(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_*)^{\mathrm{T}}\boldsymbol{\Omega}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_*)}{n-k}. \tag{61}$$

Формулы (36) и (40) для F-статистики для тестирования линейных ограничений в обобщенной модели примут вид:

$$F = (\mathbf{R}\boldsymbol{\beta}_* - \mathbf{q})^{\mathrm{T}}[\mathbf{R}s_*^2(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{R}^{\mathrm{T}}]^{-1}(\mathbf{R}\boldsymbol{\beta}_* - \mathbf{q})/J = \frac{(\tilde{\mathbf{e}}_*^{\mathrm{T}}\tilde{\mathbf{e}}_* - \mathbf{e}_*^{\mathrm{T}}\mathbf{e}_*)/J}{\mathbf{e}_*^{\mathrm{T}}\mathbf{e}_*/(n-k)}, \tag{62}$$

где $\tilde{\mathbf{e}}_* = \mathbf{Y}_* - \mathbf{X}_*\tilde{\boldsymbol{\beta}}_*$ - остатки в модели с ограничениями, полученные при помощи оценки

$$\tilde{\boldsymbol{\beta}}_* = \boldsymbol{\beta}_* - (\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{R}^{\mathrm{T}}[\mathbf{R}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{R}^{\mathrm{T}}]^{-1}(\mathbf{R}\boldsymbol{\beta}_* - \mathbf{q})$$

вектора параметров $\boldsymbol{\beta}$ в модели (8)+(52) с ограничениями (35) (ср. с формулой (38)). Наблюдаемое значения F-статистики по-прежнему необходимо сравнить с критическим значением по таблице F-распределения Фишера с $[J, n-k]$ степенями свободы на заданном уровне значимости, где J -количество ограничений. Высокие значения F-статистики свидетельствуют против ограничений.

Таким образом, с преобразованной моделью (58) можно работать как с классической моделью применяя методы и формулы разделов 1 и 2 с единственной, но важной, оговоркой. Поскольку преобразованная матрица наблюдений $\mathbf{X}_*$ не содержит столбца единиц, модель (58) не содержит свободного члена. По этой причине сумма остатков $\sum \tilde{e}_{i*}$ *не равна нулю*. Это приводит к тому, что коэффициент детерминации для преобразованной модели не ограничен рамками $[0, 1]$ и поэтому не поддается интерпретации как процент объясненных движениями переменных $\mathbf{X}_*$ движений $\mathbf{Y}_*$.

Конечно, практическое применение формул (59)-(62) невозможно без знания матрицы $\boldsymbol{\Omega}$. Эта матрица в наиболее общем случае содержит $n(n+1)/2$ различных элементов и не может быть оценена при помощи n наблюдений. Чтобы сделать оценивание возможным, предполагают, что все элементы матрицы $\mathbf{\Omega}$ зависят от небольшого числа параметров $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$: $\mathbf{\Omega} = \mathbf{\Omega}(\boldsymbol{\theta})$. Количество параметров m чаще всего находится в пределах от одного до пяти. Тогда, если $\hat{\boldsymbol{\theta}}$ - состоятельная оценка вектора параметров $\boldsymbol{\theta}$, мы можем оценить $\mathbf{\Omega}$ формулой $\hat{\mathbf{\Omega}} = \mathbf{\Omega}(\hat{\boldsymbol{\theta}})$. Подстановка этой оценки в формулу (59) приводит к следующей оценке вектора $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}}_* = (\mathbf{X}^{\mathrm{T}}\hat{\mathbf{\Omega}}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\hat{\mathbf{\Omega}}^{-1}\mathbf{Y}. \tag{63}$$

Эта оценка *осуществимым обобщённым методом наименьших квадратов*. Для асимптотической эффективности этой оценки достаточно, чтобы оценка $\hat{\boldsymbol{\theta}}$ была состоятельной.

6 Гетероскедастичность

Гетероскедастичность типична для данных среза, когда исследуются объекты разного масштаба. Другой распространенный случай возникает при работе с данными в форме средних по группам - дисперсия тогда обратно пропорциональна количеству наблюдений, по которым проводилось усреднение. Для временных рядов (в особенности для данных финансовых рынков) характерно чередование периодов с высокой и низкой дисперсией возмущений - то есть формирование кластеров из наблюдений с высокой дисперсией возмущений (соответствующих периодам нестабильности или высокой волатильности[9] рынка) и кластеров из наблюдений с низкой дисперсией возмущений (соответствующих периодам относительной стабильности или низкой волатильности рынка)[10].

Как указано выше, при наличии гетероскедастичности оценка (11) методом наименьших квадратов теряет эффективность. Процедуры статистического вывода, использованные в разделах 1 и 2 перестают быть верными в конечных выборках. Тем не менее, *в случае, если гетероскедастичность не коррелирована с переменными модели, формулы и статистические выводы метода наименьших квадратов остаются асимптотически верными и могут быть использованы в больших выборках.*

В случае, когда трудно оценить последствия гетероскедастичности для метода наименьших квадратов, необходимо использовать формулу (56) для матрицы вариаций-ковариаций оценок. Чтобы оценить эту матрицу Уайт (White, 1980) предложил следующую состоятельную оценку матрицы

[9] От англ. volatile - летучий, непостоянный, изменчивый.

[10] Для моделирования этого явления применяются ARCH и GARCH модели ([Generalized] autoregressive conditional heteroscedasticity - [обобщенная] обусловленная авторегрессивная гетероскедастичность). Рассмотрение этих моделей выходит за рамки настоящего пособия.

$\mathbf{X}^{\mathrm{T}}[\sigma^2\mathbf{\Omega}]\mathbf{X}/n$:

$$\mathbf{S}_0 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \mathbf{x}_{i\cdot}\mathbf{x}_{i\cdot}^{\mathrm{T}}. \tag{64}$$

На основе этой формулы для матрицы (56) получается **устойчивая к гетероскедастичности оценка Уайта**:

$$\begin{aligned}\hat{\mathbf{V}}_{\mathrm{OLS}} =& \frac{1}{n}\left(\frac{1}{n}\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} e_i^2\mathbf{x}_{i\cdot}\mathbf{x}_{i\cdot}^{\mathrm{T}}\right)\left(\frac{1}{n}\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\\ &= n(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{S}_0(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}.\end{aligned} \tag{65}$$

Для нормально распределенных возмущений мы можем использовать многомерное нормальное распределение со средним $\boldsymbol{\beta}$ и матрицей вариаций-ковариаций (65) для построения доверительных областей и тестирования значимости компонент вектора $\boldsymbol{\beta}$. Эксперименты методом Монте-Карло показали, что оценка Уайта дает несколько заниженные значения дисперсий и ковариаций оценок. По этой причине, Дэвидсон и МакКиннон (Davidson-Mackinnon, 1993) предложили модифицировать формулу (64):

$$\mathbf{S}_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{e_i^2}{1-\mathbf{x}_{i\cdot}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{x}_{i\cdot}}\mathbf{x}_{i\cdot}\mathbf{x}_{i\cdot}^{\mathrm{T}}. \tag{66}$$

Подстановка этой матрицы в (65) вместо $\mathbf{S}_0$ приводит к оценке Дэвидсона-Маккиннона для матрицы вариаций-ковариаций вектора $\hat{\boldsymbol{\beta}}$.

Для тестирования на наличие гетероскедастичности в данных применяют следующие тесты:

1. **Тест Уайта (White,1980)**. Это наиболее общий тест, который следует применять когда о форме гетероскедастичности ничего неизвестно. Тестируется гипотеза

$$\mathrm{H}_0: \ \sigma_i^2 = \sigma^2, \quad i = 1, \ldots, n,$$

против наиболее общей альтернативы, состоящей в том, что H_0 не выполняется. Для этого из всевозможных поэлементных произведений столбцов матрицы $\mathbf{X}$ (включая квадраты) - $X_{mi}X_{li}$, $m, l = 1, \ldots, k$, $i = 1, \ldots, n$ - и столбца единиц строится матрица наблюдений размера $n \times (k(k+1)/2 + 1)$. В качестве зависимой переменной выбирают квадраты остатков e_i^2 в первоначальной регрессии $\mathbf{Y}$ на $\mathbf{X}$ методом наименьших квадратов. Обозначим через R^2 коэффициент детерминации в регрессии этой зависимой переменной на построенную матрицу наблюдений. Статистика Уайта равна nR^2 и имеет асимптотическое χ^2 распределение с $k(k+1)/2$ степенями свободы. При превышении критического значения на заданном уровне значимости по таблице распределения χ^2 гипотеза H_0 о гомоскедастичности должна быть отвергнута.

2. **Тест Гольдфельда-Квандта (Goldfeld-Quandt, 1965)**. Для применения этого теста все наблюдения необходимо разбить на две группы. При соблюдении нулевой гипотезы дисперсии возмущений для обеих групп равны, тогда как альтернативная гипотеза состоит в том, что между дисперсиями существует систематическое различие. Если имеется предположение о зависимости дисперсии возмущений от значений одной из переменных, наблюдения упорядочивают по возрастанию значений этой переменной. В первую группу включают наблюдения с предположительно высокой дисперсией, во вторую - с предположительно низкой. Для увеличения мощности теста рекомендуют также опустить некоторое количество наблюдений из центральной части последовательности. Количество пропущенных наблюдений не должно превышать одной третьей от их общего числа. После разбиения на группы регрессия проводится отдельно по каждой из групп. Тестовая статистика вычисляется по формуле:
$$GQ = \frac{\mathbf{e}_1^{\mathrm{T}}\mathbf{e}_1/(n_1-k)}{\mathbf{e}_2^{\mathrm{T}}\mathbf{e}_2/(n_2-k)}, \tag{67}$$
где n_1, n_2 - количество наблюдений в группах с предположительно большей и меньшей дисперсиями, соответственно. Если соблюдается нулевая гипотеза о гомоскедастичности и возмущения нормально распределены, то эта статистика имеет F- распределение Фишера с $[n_1-k, n_2-k]$ степенями свободы. Таким образом, нулевая гипотеза должна быть отвергнута, если наблюдаемое значение этой статистики превышает критическое на заданном уровне значимости.

3. **Тест Бройша-Пагана (Breusch-Pagan, 1979)**. Дисперсия возмущений может зависеть от нескольких переменных или от неизвестной заранее переменной. Если это происходит, то мы не можем применить тест Гольдфельда-Квандта. Для этого случая Бройш и Паган разработали тест множителей Лагранжа для тестирования гипотезы
$$\sigma_i^2 = \sigma^2 f(\alpha_0 + \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{z}_{i\cdot}),$$
где $\mathbf{z}_{i\cdot}$ - i-ое наблюдение над некоторым подмножеством объясняющих переменных (в него могут входить также все объясняющие переменные). Модель гомоскедастична, если $\boldsymbol{\alpha} = \mathbf{0}$. Пусть
$$\mathbf{g} = (g_1, g_2, \ldots, g_n)^{\mathrm{T}}, \quad g_i = e_i^2/(\mathbf{e}^{\mathrm{T}}\mathbf{e}/n) - 1.$$

Объяснённая сумма квадратов отклонений g_i от нуля (легко проверить, что $\overline{g} = \overline{e} = 0$) при регрессии на наблюдения $(1, \mathbf{z}_{i\cdot})$, $i = 1, \ldots, n$, из которых составлена матрица $\mathbf{Z}_{n\times(p+1)}$, равна $\mathbf{g}^{\mathrm{T}}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{g}$ (см. формулу (22)). Тест множителей Лагранжа Бройша-Пагана равен половине этой величины:

$$\mathrm{LM} = [\mathbf{g}^{\mathrm{T}}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{g}]/2. \tag{68}$$

При условии справедливости нулевой гипотезы о гомоскедастичности, LM имеет асимптотическое χ^2 распределение с p степенями свободы (p - это количество переменных в $\mathbf{Z}$).

4. **Тест Кёнкера-Бассетта (Koeonker-Basset, 1982)**. Тест Бройша-Пагана весьма чувствителен к предположению о нормальности возмущений. Кёнкер и Бассетт предложили более устойчивую модификацию этого теста. Пусть $\mathbf{u} = (e_1^2, e_2^2, \ldots, e_n^2)$ и $\mathbf{i}$ - столбец единиц размером $n \times 1$. Тогда $\overline{u} = \mathbf{e}^{\mathrm{T}}\mathbf{e}/n$, а оценкой для σ^2 может служить

$$V = \frac{1}{n}\sum_{i=1}^{n}\left[e_i^2 - \frac{\mathbf{e}^{\mathrm{T}}\mathbf{e}}{n}\right]^2 = (\mathbf{u} - \overline{u}\mathbf{i})^{\mathrm{T}}(\mathbf{u} - \overline{u}\mathbf{i})/n.$$

Тестовая статистика имеет вид:

$$\mathrm{LM} = \frac{1}{V}(\mathbf{u} - \overline{u}\mathbf{i})^{\mathrm{T}}\mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}(\mathbf{u} - \overline{u}\mathbf{i}). \tag{69}$$

При условии справедливости нулевой гипотезы о гомоскедастичности и для нормально распределенных возмущений, она имеет такое же асимптотическое распределение как статистика (68).

5. **Тест на межгрупповую гетероскедастичность**. Пусть все данные разбиты на G групп:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_G \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_G \end{bmatrix}. \tag{70}$$

Обозначим через n_g количество наблюдений в группе с номером g, $n = \sum_{g=1}^{G} n_g$. Мы хотим проверить гипотезу о постоянстве дисперсии возмущений во всех группах против альтернативной гипотезы о том, что дисперсии различны. Оценка максимального правдоподобия для дисперсии возмущений, при условии справедливости нулевой гипотезы, равна

$$s^2 = \frac{\mathbf{e}^{\mathrm{T}}\mathbf{e}}{n},$$

где $\mathbf{e}$ - вектор остатков при оценивании объединённых данных обычным методом наименьших квадратов: $\mathbf{e} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$. При условии справедливости альтернативной гипотезы оценка максимального правдоподобия для дисперсии возмущений в группе с номером g равна

$$s_g^2 = \frac{\mathbf{e}_g^{\mathrm{T}}\mathbf{e}_g}{n_g}.$$

Здесь вектор остатков $\mathbf{e}_*$ разбит на подвекторы $\mathbf{e}_g$, $g = 1, \dots, G$, соответствующие группам наблюдений. Вектор $\mathbf{e}_* = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_*$, где $\hat{\boldsymbol{\beta}}_*$ - оценка максимального правдоподобия, полученная по формуле (75) методом изложенным в пункте (в) ниже. Критерий отношения правдоподобия для проверки гипотезы о постоянстве дисперсий во всех группах равен:

$$\mathrm{LR} = n \ln s^2 - \sum_{g=1}^{G} n_g \ln s_g^2.$$

Этот критерий имеет асимптотическое χ^2 распределение с $G-1$ степенями свободы. Превышение наблюдаемого значения над критическим, взятым из таблицы распределения χ^2_{G-1}, свидетельствует против постоянства дисперсий в различных группах.

При применении этих тестов необходимо иметь в виду, что отрицание гипотезы о гомоскедастичности может, в первую очередь, свидетельствовать о неверной спецификации модели, а вовсе не о наличии истинной гетероскедастичности в данных. Требуется тщательное дополнительное исследование, связанное с содержательным характером обрабатываемых данных, прежде чем гипотеза о наличии гетероскедастичности может быть принята.

Если в результате тестирования принято решение, что в модели присутствует истинная гетероскедастичность, то для эффективного оценивания можно применить метод обобщенных наименьших квадратов, то есть формулу (59). Матрица $\boldsymbol{\Omega}$ имеет вид (53) с нормировкой (54). Следовательно, матрица $\mathbf{P}$ из (57) диагональна с числами $1/\sqrt{\omega_i}$ на главной диагонали, поэтому трансформированная модель (58) получается делением i-ой строки $\mathbf{X}$ и $\mathbf{Y}$ на $\sqrt{\omega_i}$. Оценка методом обобщённых наименьших квадратов принимает вид

$$\boldsymbol{\beta}_* = \left[\sum_{i=1}^{n} \frac{1}{\omega_i} \mathbf{x}_{i\cdot}\mathbf{x}_{i\cdot}^{\mathrm{T}}\right]^{-1} \left[\sum_{i=1}^{n} \frac{1}{\omega_i} \mathbf{x}_{i\cdot} Y_i\right]. \tag{71}$$

Согласно этой формуле, наблюдения с меньшей дисперсией возмущений получают больший вес и наоборот. По этой причине оценку (71) называют еще оценкой *взвешенным методом наименьших квадратов*.

Для практического применения формулы (71) нам необходимо оценить дисперсии возмущений σ_i^2. Если обозначить такие оценки через $\hat{\sigma}_i^2$, то:

$$\hat{\sigma}^2 = \sum_{i=1}^{n} \hat{\sigma}_i^2 / n, \quad \hat{\omega}_i = \hat{\sigma}_i^2 / \hat{\sigma}^2. \tag{72}$$

Для получения оценок $\hat{\sigma}_i^2$ было предложено несколько способов:

(а) дисперсии предполагают пропорциональными заданной функции от одной из объясняющих переменных, например $\sigma_i^2 = \sigma^2 X_{im}^2$, $\sigma_i^2 = \sigma^2 \ln(X_{im})$ или $\sigma_i^2 = \sigma^2 (X_{im})^{-1}$. Выбранная переменная X_m отвечает за масштаб рассматриваемого объекта - доход семьи, количество работников фирмы или валовый внутренний продукт страны, и т.п.;

(б) дисперсии предполагаются зависящими линейно[11] от некоторого набора переменных $\mathbf{z}$: $\sigma_i^2 = \mathbf{z}_i \boldsymbol{\alpha}$, где $\mathbf{z}_i$ - i-ое наблюдение за $\mathbf{z}$. В этот набор могут быть включены как переменные взятые из матрицы $\mathbf{X}$, так и любые другие переменные, от которых предположительно зависит дисперсия возмущений. Для оценки вектора параметров этой зависимости $\boldsymbol{\alpha}$ используют тот факт, что остатки $e_i = Y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}$ при оценивании обычным методом наименьших квадратов являются состоятельными оценками возмущений ε_i даже при наличии гетероскедастичности и $\varepsilon_i^2 = \sigma_i^2 + \nu_i$, где ν_i - некоторая случайная поправка. Таким образом, формулируется регрессионная модель:

$$e_i^2 = \mathbf{z}_i \boldsymbol{\alpha} + \nu_i^*, \qquad i = 1, \ldots, n, \tag{73}$$

в которой возмущения ν_i^* гетероскедастичны и автокоррелированны. Оценка этой модели обычным методом наименьших квадратов[12]

$$\hat{\boldsymbol{\alpha}} = (\mathbf{Z}^{\mathrm{T}} \mathbf{Z})^{-1} \mathbf{Z}^{\mathrm{T}} \mathbf{u}$$

состоятельна, хотя и крайне неэффективна. Эта оценка используется для состоятельного оценивания дисперсии возмущений:

$$\hat{\sigma}_i^2 = \mathbf{z}_i \hat{\boldsymbol{\alpha}}.$$

Такой метод называют *двухшаговым оцениванием.*

(в) оценку предыдущего пункта можно существенно улучшить для модели мультипликативной гетероскедастичности Харвея (Harvey, 1976):

$$\sigma_i^2 = \exp(\mathbf{q}_i \boldsymbol{\gamma}) \quad \Leftrightarrow \quad \sigma_i^2 = \sigma^2 \exp(\mathbf{z}_i \boldsymbol{\alpha}),$$

[11]Эта зависимость может быть нелинейна по переменным и даже по параметрам, если она линеаризуема. Указанный вид имеет тогда уже линеаризованная модель. Подробнее см. замечание в конце второго раздела на странице 15.

[12]Напомним, что $\mathbf{u} = (e_1^2, e_2^2, \ldots, e_n^2)$.

где $\boldsymbol{\gamma} = \begin{bmatrix} \ln \sigma^2 \\ \boldsymbol{\alpha} \end{bmatrix}$, $\mathbf{q}_i = [1, \mathbf{z}_i]$, а $\mathbf{z}_i$ имеет тот же смысл, что в пункте (б). Следующий алгоритм в случае сходимости даёт оценку максимального правдоподобия для вектора $\boldsymbol{\alpha}$ и асимптотически эффективную оценку осуществимым методом обобщенных наименьших квадратов для $\boldsymbol{\beta}$:

1. Объединить данные и оценить $\boldsymbol{\beta}$ обычным методом наименьших квадратов по формуле (11), вычислить остатки $\mathbf{e} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$;

2. Регрессией $\ln(e_i^2)$ по $\mathbf{q}_i$, $i = 1, \ldots, n$, обычным методом наименьших квадратов вычислить $\hat{\boldsymbol{\gamma}}$. К полученной оценке $\hat{\gamma}_1$ для первой компоненты вектора $\hat{\boldsymbol{\gamma}}$ добавить 1.2704[13].

3. Оценить дисперсии возмущений формулой $\hat{\sigma}_i^2 = \exp(\mathbf{q}_i\hat{\boldsymbol{\gamma}})$.

4. Вычислить оценки весов по формулам (72) и оценить $\hat{\boldsymbol{\beta}}_*$ взвешенным методом наименьших квадратов (71). Посчитать остатки $\mathbf{e}_* = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_*$.

5. Обновить оценку вектора $\boldsymbol{\gamma}$ регрессией $[e_{*i}^2 \exp(-\hat{\boldsymbol{\gamma}}\mathbf{q}_i) - 1]$ по $\mathbf{q}_i$:
$$\hat{\boldsymbol{\gamma}} := \hat{\boldsymbol{\gamma}} + (\mathbf{Q}^{\mathrm{T}}\mathbf{Q})^{-1} \left[\sum_{i=1}^{n} \mathbf{q}_i^{\mathrm{T}} (e_{*i}^2 \exp(-\hat{\boldsymbol{\gamma}}\mathbf{q}_i) - 1) \right]. \tag{74}$$

6. Сравнить объединённый вектор $[\hat{\boldsymbol{\beta}}_*, \hat{\boldsymbol{\gamma}}]$ с вычисленным на предыдущей итерации. Если разница мала (не превышает нескольких процентов), то сходимость достигнута и текущее значения $\hat{\boldsymbol{\beta}}_*$ даёт оценку вектора $\boldsymbol{\beta}$ с ковариационной матрицей примерно раной $(\mathbf{X}^{\mathrm{T}}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X})^{-1}$, где $\hat{\boldsymbol{\Omega}}$ - диагональная матрица, сформированная из текущих значений $\hat{\sigma}_i^2$. Иначе вернуться к шагу 3.

(г) для данных, разбитых на G групп (см. (70)), таким образом, что внутри каждой группы дисперсии возмущений одинаковы, но отличаются между группами, следующая процедура Оберхофера-Кмента (Oberhofer-Kmenta, 1974) приводит к оценке максимального правдоподобия для вектора параметров $\boldsymbol{\beta}$:

1. Объединить данные и оценить $\boldsymbol{\beta}$ обычным методом наименьших квадратов по формуле (11);

2. Используя оценку $\hat{\boldsymbol{\beta}}_*$, полученную на предыдущем шаге, вычислить остатки отдельно для каждой группы наблюдений: $\mathbf{e}_g = \mathbf{Y}_g - \mathbf{X}_g\hat{\boldsymbol{\beta}}_*$, $g = 1, \ldots, G$, и оценки дисперсий по группам $\hat{\sigma}_g^2 = \mathbf{e}_g^{\mathrm{T}}\mathbf{e}_g/n_g$, где n_g - количество наблюдений в g-ой группе;

[13] Харвей показал, что $\hat{\gamma}_1$ даёт несостоятельную оценку γ_1, однако $\hat{\gamma}_1' = \hat{\gamma}_1 + 1.2704$ уже будет состоятельной.

3. Пересчитать оценку вектора параметров по формуле (71), которая в этом случае принимает вид:
$$\hat{\boldsymbol{\beta}}_* = \left[\sum_{g=1}^{G} \frac{1}{\hat{\sigma}_g^2}\mathbf{X}_g^{\mathrm{T}}\mathbf{X}_g\right]^{-1} \left[\sum_{g=1}^{G} \frac{1}{\hat{\sigma}_g^2}\mathbf{X}_g^{\mathrm{T}}\mathbf{Y}_g\right]; \tag{75}$$
4. Если новая оценка мало отличается от полученной в предыдущей итерации (скажем на 1-2%), то сходимость достигнута и $\hat{\boldsymbol{\beta}}_*$ можно считать оценкой максимального правдоподобия для $\boldsymbol{\beta}$; иначе вернуться к шагу 2.

При любом выборе диагональной матрицы весов $\hat{\boldsymbol{\Omega}}$ оценка
$$\hat{\boldsymbol{\beta}}_* = [\mathbf{X}^{\mathrm{T}}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Y} \tag{76}$$
взвешенным методом наименьших квадратов будет состоятельной. Однако, она будет тем менее эффективной, чем сильнее оцененная матрица $\hat{\boldsymbol{\Omega}}$ отличается от истинной матрицы $\boldsymbol{\Omega}$. Кроме того, матрица вариаций-ковариаций оценки (76) (а следовательно и дисперсии отдельных коэффициентов) не будет иметь вид (60). Правильная асимптотическая матрица вариаций-ковариаций оценки (76) имеет вид:
$$\mathbf{Var}\hat{\boldsymbol{\beta}}_* = \sigma^2[\mathbf{X}^{\mathrm{T}}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Omega}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X}[\mathbf{X}^{\mathrm{T}}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X}]^{-1}. \tag{77}$$
Грин (Greene, 1997) предложил использовать формулу (77) совместно с оценкой Уайта (64). Это приводит к следующей оценке матрицы вариаций-ковариаций оценки $\hat{\boldsymbol{\beta}}_*$ взвешенным методом наименьших квадратов
$$\widehat{\mathbf{Var}}\hat{\boldsymbol{\beta}}_* = [\mathbf{X}^{\mathrm{T}}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X}]^{-1}\left[\sum_{i=1}^{n} \frac{e_{*i}^2}{\hat{\omega}_i^2}\mathbf{x}_{i\cdot}^{\mathrm{T}}\mathbf{x}_{i\cdot}\right][\mathbf{X}^{\mathrm{T}}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X}]^{-1}. \tag{78}$$

7 Автокорреляция возмущений

Автокорреляция возмущений характерна, главным образом, для временных рядов. В этом случае векторы наблюдений над объясняющими переменными $\mathbf{x}_{\cdot j}$, $j = 2, \ldots, k$, вектор наблюдений над зависимой переменной $\mathbf{Y}$ и вектор возмущений $\boldsymbol{\varepsilon}$ для целей анализа и моделирования считают реализациями *случайных процессов*[14]. Для статистического оценивания параметров случайного процесса необходимо иметь набор его реализаций. Экономические данные, однако, представляют собой единственную реализацию.

[14]**Случайным процессом** с дискретным временем называют последовательность случайных величин.

Для того, чтобы по единственной реализации процесса можно было достоверно оценить его параметры, необходимо потребовать, чтобы вероятностные характеристики процесса оставались неизменны с течением времени. Такое свойство случайного процесса называют стационарностью[15]. Более точное определение следующее. Случайный процесс ε_t называют **слабо стационарным**, если его математическое ожидание, дисперсия и автоковариация не зависят от времени:

$$\begin{aligned} &E(\varepsilon_t) = \mu = \text{const}, \\ &E(\varepsilon_t - \mu)^2 = \gamma_0 = \sigma^2 = \text{const}, \\ &E((\varepsilon_t - \mu)(\varepsilon_{t-s} - \mu)) = \gamma_s = \text{const}. \end{aligned} \tag{79}$$

Таким образом, мы рассмотрим модель (9), в которой возмущения $\boldsymbol{\varepsilon}$ являются реализацией стационарного случайного процесса. Благодаря третьему из условий (79) матрица $\mathbf{\Omega}$ имеет вид (55) с $\rho_s = \gamma_s/\gamma_0$. Любой слабо стационарный случайный процесс может быть сколь угодно точно приближен **процессом авторегрессии - скользящего среднего** конечного порядка, то есть процессом вида[16]:

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \ldots + \phi_p \varepsilon_{t-p} + \eta_t + \theta_1 \eta_{t-1} + \ldots + \theta_q \eta_{t-q}, \tag{80}$$

где η_t - последовательность независимых нормально распределенных возмущений со средним ноль и дисперсией σ^2. Следующие два частных случая этого процесса наиболее важны:

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \ldots + \phi_p \varepsilon_{t-p} + \eta_t \tag{81}$$

называют процессом авторегрессии (или просто авторегрессией) порядка p и обозначают $\varepsilon_t \sim \mathrm{AR}(p)$;

$$\varepsilon_t = \eta_t + \theta_1 \eta_{t-1} + \ldots + \theta_q \eta_{t-q} \tag{82}$$

называют процессом скользящего среднего (или просто скользящим средним) порядка q и обозначают $\varepsilon_t \sim \mathrm{MA}(q)$. Смешанный процесс (80) обозначают $\varepsilon_t \sim \mathrm{ARMA}(p,q)$. Процессы (81) и (80) слабо стационарны тогда и только тогда, когда все корни характеристического уравнения

$$\lambda^p - \phi_1 \lambda^{p-1} - \ldots - \phi_p = 0 \tag{83}$$

[15]Точнее для состоятельности оценок, полученных по одной реализации требуется **эргодичность**. Это понятие не рассматривается в настоящем пособии, но для всех встречающихся здесь процессов эргодичность следует из стационарности.

[16]Это следует из **теоремы Вольда о декомпозиции**, которая позволяет записать любой слабо стационарный случайный процесс в виде суммы бесконечного скользящего среднего (недетерминисткой части) и конечной линейной комбинации начальных значений процесса (детерминистской части).

по модулю строго меньше единицы. Процесс (82) слабо стационарен при любых значениях параметров.

Чаще всего для моделирования автокорреляции возмущений используется процесс авторегрессии первого порядка ($|\rho| < 1$):

$$\varepsilon_t = \rho\varepsilon_{t-1} + \eta_t. \tag{84}$$

В этом случае

$$\gamma_0 = \sigma_\eta^2/(1-\rho^2), \quad \gamma_s = \rho^s,$$

где σ_η^2 - дисперсия η_t и матрица $\mathbf{\Omega}$ имеет вид:

$$\sigma^2\mathbf{\Omega} = \frac{\sigma_\eta^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \dots & \rho^{n-1} \\ \rho & 1 & \dots & \rho^{n-1} \\ & & \vdots & \\ \rho^{n-1} & \rho^{n-2} & \dots & 1 \end{bmatrix}. \tag{85}$$

Другой часто используемый вид автокорреляции возмущений - процесс скользящего среднего первого порядка

$$\varepsilon_t = \eta_t + \theta\eta_{t-1} \tag{86}$$

приводит к значениям:

$$\gamma_0 = \sigma_\eta^2(1+\theta^2), \qquad \gamma_1 = \sigma_\eta^2\theta, \qquad \gamma_s = 0, \ s > 1,$$

и следующему виду матрицы $\mathbf{\Omega}$:

$$\sigma^2\mathbf{\Omega} = \sigma_\eta^2(1+\theta^2)\begin{bmatrix} 1 & \frac{\theta}{1+\theta^2} & 0 & \dots & 0 \\ \frac{\theta}{1+\theta^2} & 1 & \frac{\theta}{1+\theta^2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & \frac{\theta}{1+\theta^2} \\ 0 & 0 & \dots & \frac{\theta}{1+\theta^2} & 1 \end{bmatrix}. \tag{87}$$

Оценка методом наименьших квадратов модели в автокоррелированными возмущениями сохранят свойство несмещенности. Её состоятельность зависит от поведения матрицы $\mathbf{X}^{\mathrm{T}}\mathbf{\Omega}\mathbf{X}/n$ с ростом n. Если предел по вероятности существует и является положительно определенной матрицей, то оценка МНК (11) остаётся состоятельной. Это, в частности, верно для возмущений видов (84) и (86). В любом случае оценка методом наименьших квадратов не будет эффективной. Наконец в случае, *когда среди объясняющих переменных есть запаздывающие значения зависимой переменной (то есть $X_{tm} = Y_{t-1}$ для некоторого m), оценка методом наименьших квадратов несостоятельна при наличии автокорреляции возмущений.* Процедуры статистического вывода, приведенные в разделах 1 и 2 должны быть модифицированы как указано в разделе 5. Для оценивания матрицы $\mathbf{X}^{\mathrm{T}}[\sigma^2\mathbf{\Omega}]\mathbf{X}/n$ Ньюэй и Вест (Newey-West, 1987) предложили оценку:

$$\hat{\mathbf{Q}} = \mathbf{S}_0 + \frac{1}{n}\sum_{l=1}^{L}\sum_{t=l+1}^{n} w_l e_t e_{t-l}(\mathbf{x}_{t\cdot}\mathbf{x}_{t-l\cdot}^{\mathrm{T}} + \mathbf{x}_{t-l\cdot}\mathbf{x}_{t\cdot}^{\mathrm{T}}), \quad w_l = \frac{l}{L+1}, \tag{88}$$

где $\mathbf{S}_0$ - оценка Уайта (64). Величина L должна выбираться заранее - она равна максимальному промежутку времени (то есть количеству периодов времени), при превышении которого мы можем пренебречь автокорреляцией возмущений. Эта оценка будет состоятельной, если автокорреляции возмущений достаточно быстро убывают при удалении от главной диагонали матрицы $\mathbf{\Omega}$.

Значения возмущений ε_t ненаблюдаемы, поэтому для тестирования на наличие автокорреляции в возмущениях мы вынуждены использовать остатки $e_t = Y_t - \mathbf{x}_{t\cdot}\hat{\boldsymbol{\beta}}$ в качестве оценок для возмущений[17]. Таким образом, мы фактически тестирует автокоррелированность остатков, а не возмущений! Автокоррелированность же остатков, в свою очередь, может быть вызвана как автокорреляцией возмущений так и рядом других причин. Первой в ряду этих причин стоит неверная спецификация модели. Более того, неверная спецификация модели является намного более частой причиной автокорреляции остатков, чем автокорреляция возмущений. Следовательно, приведённые ниже тесты можно рассматривать как тесты на неверную спецификацию модели.

1. **Тест Дёрбина-Уатсона (Durbin-Watson, 1950).** Тестовая статистика вычисляется по формуле:

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} = 2(1-r) - \frac{e_1^2 + e_n^2}{\sum_{t=1}^{n} e_t^2} \approx 2(1-r), \tag{89}$$

где r - выборочная автокорреляция:

$$r = \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}. \tag{90}$$

[17]Остатки МНК будут состоятельными оценками возмущений при наличии автокорреляции, *но лишь с случае, когда среди регрессоров нет задаздывающих значений зависимой переменной.*

Если справедлива нулевая гипотеза об отсутствии автокорреляции $H_0: \ \rho = 0$, где ρ - истинный коэффициент корреляции, следует ожидать, что $r \approx 0$ и $d \approx 2$. Тогда мы должны отвергнуть H_0 против альтернативы $\rho > 0$, если $d < d_*$, где d_* - некоторое критическое значение на заданном уровне значимости. Для вычисления этого критического значения нам необходимо знать распределение вероятностей d. Это распределение вероятностей не зависит от $\boldsymbol{\beta}$ и σ, но зависит от n, k и матрицы наблюдений $\mathbf{X}$, поэтому не может быть затабулировано[18]. Дёрбин и Уатсон доказали, что распределение вероятностей d лежит между распределениями вероятностей двух граничных статистик, обозначаемых d_u (upper - верхняя граница) и d_l (lower -нижняя граница). Распределения вероятностей d_u и d_l зависят только от числа наблюдений n и количества регрессоров k. Таким образом, на заданном уровне значимости α по таблице находят критические значения d_u^α и d_l^α, поле чего сравнивают их с наблюдаемым значением d. Гипотеза об отсутствии автокорреляции отвергается против альтернативы о наличии положительной автокорреляции если $d < d_l^\alpha$ и не может быть отвергнута, если $d > d_u^\alpha$. Если $d_l^\alpha < d < d_u^\alpha$, то статистика оставляет вопрос открытым. Точно также гипотеза об отсутствии автокорреляции отвергается против альтернативы о наличии отрицательной автокорреляции если $d > 4 - d_l^\alpha$ и не может быть отвергнута если $d < 4 - d_u^\alpha$; промежуток $(4 - d_l^\alpha, 4 - d_u^\alpha)$ остаётся зоной неопределенности. Аналогичный тест для квартальных данных предложен Валлисом (Wallis, 1972). Соответствующая статистика имеет вид

$$d_4 = \frac{\sum\limits_{t=5}^{n}(e_t - e_{t-4})^2}{\sum\limits_{t=1}^{n} e_t^2}. \tag{91}$$

Валлис затабулировал критические значения для этой статистики.

При применении теста Дёрбина-Уатсона важно иметь в виду, что *при наличии запаздывающих значений зависимой переменной среди регрессоров значения d смещены к 2 и тест слишком часто оказывается не способен обнаружить автокорреляцию, то есть на его результаты нельзя полагаться.*

2. **Тест Бройша-Годфри (Breusch-Godfrey, 1978).** Это тест гипоте-

[18] Некоторые современные программные продукты, тем не менее, способны вычислять точные критические значения статистики Дёрбина-Уатсона для нормально распределенных возмущений и заданной матрицы $\mathbf{X}$.

зы H_0 об отсутствии автокорреляции против альтернативы

$$H_1 : \varepsilon \sim \mathrm{AR}(p) \text{ или } \varepsilon \sim \mathrm{MA}(p).$$

Для проведения теста надо вычислить коэффициент детерминации R^2 (см. формулу (25)) в регрессии $\mathbf{e}$ (в качестве $\mathbf{Y}$) на матрицу $\mathbf{Z}_{n\times(k+p)}$ вида

$$\mathbf{Z} = \begin{bmatrix} 1 & X_{12} & X_{13} & \dots & X_{1k} & 0 & 0 & \dots & 0 \\ 1 & X_{22} & X_{23} & \dots & X_{2k} & e_1 & 0 & \dots & 0 \\ 1 & X_{32} & X_{33} & \dots & X_{3k} & e_2 & e_1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ 1 & X_{t2} & X_{t3} & \dots & X_{tk} & e_{t-1} & e_{t-2} & \dots & e_{t-p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ 1 & X_{n2} & X_{n3} & \dots & X_{nk} & e_{n-1} & e_{n-2} & \dots & e_{n-p} \end{bmatrix}.$$

Тестовая статистика LM (Lagrange multiplier - множитель Лагранжа) равна nR^2 и имеет, если справедлива нулевая гипотеза, χ^2_p распределение. Данный тест работает также при наличии запаздывающих значений зависимой переменной среди регрессоров. Отрицание нулевой гипотезы означает наличие автокорреляции остатков, отстоящих друг от друга на расстояние от 1 до p наблюдений.

3. **Q тест Вокса-Пирса (Box-Pierce, 1970)**. Тестовая статистика вычисляется по формуле:

$$Q = n \sum_{j=1}^{p} r_j^2, \tag{92}$$

где выборочные коэффициенты автокорреляции равны

$$r_j = \frac{\sum\limits_{t=j+1}^{n} e_t e_{t-j}}{\sum\limits_{t=1}^{n} e_t^2}.$$

Этот тест асимптотически эквивалентен предыдущему и наблюдаемое значение Q также сверяют с таблицей распределения χ^2 с p степенями свободы. Высокие значения Q свидетельствуют против гипотезы об отсутствии автокорреляции.

4. **Тест Льюнга-Бокса (Ljung-Box, 1979)**. Это модификация предыдущего теста. Тестовая статистика равна

$$Q' = n(n+2) \sum_{j=1}^{p} \frac{r_j^2}{n-j} \tag{93}$$

и также имеет асимптотическое распределение χ^2_p. Для конечных выборок, однако, распределение Q' ближе к χ^2_p, чем распределение Q.

5. **Тест Дёрбина (Durbin, 1970)**. Эта модификация теста Дёрбина-Уатсона специально предназначена для случая, когда среди регрессоров имеются запаздывающие значения зависимой переменной. Тестовая статистика
$$h = \left(1 - \frac{d}{2}\right)\sqrt{\frac{n}{1 - ns^2_{y_{t-1}}}} \tag{94}$$
сравнивается с критическим значением по таблице стандартного нормального распределения на заданном уровне значимости. Высокие значения h свидетельствуют против гипотезы об отсутствии автокорреляции. В формуле (94) d - это статистика Дёрбина-Уатсона (89), $s^2_{y_{t-1}}$ - оценка дисперсии коэффициента при y_{t-1}, вычисляемая по формуле (16). Недостаток теста в невозможности вычислить h, когда $s^2_{y_{t-1}} > 1/n$.

Если в результате тестирования и содержательного исследования данных было принято решение о наличии истинной автокорреляции возмущений, то для эффективного оценивания модели применяют осуществимый метод наименьших квадратов (63) или метод максимального правдоподобия.

Пусть сначала возмущения порождены процессом авторегрессии первого порядка (84), $\varepsilon \sim \mathrm{AR}(1)$, и ρ известно. Тогда мы можем использовать формулу (59) с $\mathbf{\Omega}$ из (85) для эффективного оценивания $\boldsymbol{\beta}$. Матрица $\mathbf{P}$ из (57) оказывается при этом равной:

$$(1-\rho^2)^{1/2}\mathbf{P} = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & 0 & \dots & -\rho & 1 \end{bmatrix} \tag{95}$$

и применение (59) эквивалентно применению обычного метода наименьших квадратов к трансформированной модели (58), которая примет вид:

$$\mathbf{X}_* = \begin{bmatrix} \sqrt{1-\rho^2}\mathbf{x}_{1\cdot} \\ \mathbf{x}_{2\cdot} - \rho\mathbf{x}_{1\cdot} \\ \mathbf{x}_{3\cdot} - \rho\mathbf{x}_{2\cdot} \\ \vdots \\ \mathbf{x}_{n\cdot} - \rho\mathbf{x}_{n-1\cdot} \end{bmatrix}, \qquad \mathbf{Y}_* = \begin{bmatrix} \sqrt{1-\rho^2}Y_1 \\ Y_2 - \rho Y_1 \\ Y_3 - \rho Y_2 \\ \vdots \\ Y_n - \rho Y_{n-1} \end{bmatrix}. \tag{96}$$

Для применения этих формул неизвестное значение ρ заменяют одной из оценок. Наиболее часто используются следующие оценки:

- В качестве $\hat{\rho}$ используют выборочный коэффициент корреляции (90).
- Оценка Тейла r^* (Theil, 1971):

$$r^* = \left[\frac{n-k}{n-1}\right] r. \tag{97}$$

- Оценка Тейла r^{**} (Theil, 1971):

$$r^{**} = 1 - \frac{d}{2}, \tag{98}$$

  где d - статистика Дёрбина-Уатсона (89).

- Оценка Дёрбина (Durbin, 1970) равна оценке методом наименьших квадратов коэффициента при Y_{t-1} в регрессии

$$Y_t = \rho Y_{t-1} + \mathbf{x}_{t\cdot}\boldsymbol{\beta} - \mathbf{x}_{t-1\cdot}(\rho\boldsymbol{\beta}) + \varepsilon_t, \quad t = 2, \ldots, n. \tag{99}$$

- Оценка Хильдерта-Лу (Hilderth-Lu, 1960) получается последовательным перебором значений ρ от -1 до 1 с некоторым шагом. Выбирается значение ρ, при котором минимальна сумма квадратов остатков при оценивании трансформированной модели (96) методом наименьших квадратов.

Все приведенные оценки ρ состоятельны, но их свойства в конечных выборках различны. Выбранную оценку ρ подставляют в формулу (96) после чего трансформированную модель оценивают методом наименьших квадратов. Полученную таким образом оценку $\boldsymbol{\beta}$ называют *оценкой Прайса-Уинстена* (Prais-Winsten, 1954). Иногда для простоты опускают первое наблюдение (поскольку формула для его преобразования отличается от формулы преобразования остальных наблюдений). Получающаяся при этом оценка была предложена Кохрейном и Оркаттом (Cochrane-Orcutt, 1949). Исследования показали, однако, что оценка Прайса-Уинстена всегда предпочтительней.

Метод максимального правдоподобия состоит в выборе значений ρ, $\boldsymbol{\beta}$ и σ_η^2 которые максимизируют логарифмическую функцию правдоподобия

$$\ln L = -\frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma_\eta^2}\sum_{t=1}^{n} e_{*t}^2 + \frac{1}{2}\ln(1-\rho^2) - \frac{n}{2}\ln\sigma_\eta^2, \tag{100}$$

где

$$e_{*1} = \sqrt{1-\rho^2}(Y_1 - \mathbf{x}_{1\cdot}\boldsymbol{\beta}),$$
$$e_{*t} = (Y_t - \rho Y_{t-1}) - (\mathbf{x}_{t\cdot} - \rho \mathbf{x}_{t-1\cdot})\boldsymbol{\beta}, \quad t = 2, \ldots, n.$$

При заданном значении ρ оценками максимального правдоподобия для $\boldsymbol{\beta}$ и σ_η^2 являются оценка обобщенным методом наименьших квадратов (то есть оценка трансформированной модели (96) обычным методом наименьших квадратов) и оценка

$$\hat{\sigma}_\eta^2 = \frac{1}{n}\sum_{t=1}^{n} e_{*t}^2, \tag{101}$$

соответственно. Это подсказывает следующий алгоритм построения оценки максимального правдоподобия:

(а) положим $\hat{\rho} = -1$;

(б) увеличим текущее значение $\hat{\rho}$ на шаг δ: $\hat{\rho} := \hat{\rho} + \delta$;

(в) вычислим значения оценок $\hat{\boldsymbol{\beta}}_*$ методом обобщенных наименьших квадратов и $\hat{\sigma}_\eta^2$ по формуле (101);

(г) вычислим значение функции правдоподобия (100) при текущих значениях $\hat{\rho}$, $\hat{\boldsymbol{\beta}}_*$ и $\hat{\sigma}_\eta^2$ и запомним вычисленные значение $\ln L$, $\hat{\rho}$, $\hat{\boldsymbol{\beta}}_*$ и $\hat{\sigma}_\eta^2$;

(д) Если $\hat{\rho} < 1$ переход к шагу (б), иначе

(е) выберем значения $\hat{\rho}$, $\hat{\boldsymbol{\beta}}_*$ и $\hat{\sigma}_\eta^2$ соответствующие максимуму функции правдоподобия.

Если возмущения порождены процессом авторегрессии 2-го порядка

$$\varepsilon_t = \phi_1\varepsilon_{t-1} + \phi_2\varepsilon_{t-2} + \eta_t, \tag{102}$$

то трансформированная модель (58) получается по формулам:

$$\begin{aligned}
\mathbf{z}_{*1} &= \left[\frac{(1+\phi_2)[(1-\phi_2^2)-\phi_1^2]}{1-\phi_2}\right]^{1/2}\mathbf{z}_1 \\
\mathbf{z}_{*2} &= (1-\phi_2^2)^{1/2}\mathbf{z}_2 - \frac{\phi_1(1-\phi_1^2)^{1/2}}{1-\phi_2}\mathbf{z}_1 \\
\mathbf{z}_{*t} &= \mathbf{z}_t - \phi_1\mathbf{z}_{t-1} - \phi_2\mathbf{z}_{t-2}, \quad t > 2.
\end{aligned} \tag{103}$$

Здесь $\mathbf{z}_t$ обозначает Y_t или $\mathbf{x}_{t\cdot}$. Оценки для ϕ_1, ϕ_2 могут быть получены применением метода наименьших квадратов к модели (102), в которой возмущения ε_t заменены остатками e_t.

Если среди регрессоров есть запаздывающие значения зависимой переменной, все оценки коэффициента автокорреляции ρ, основанные на использовании на МНК-остатков $\mathbf{e}$, перестают быть состоятельными. Следовательно, и оценка Прайса-Винстена, полученная с применением неверной оценки для ρ, также не будет состоятельной. В этом случае можно применить следующий метод Хатанака (Hatanaka, 1974) для модели:

$$\begin{aligned} Y_t &= \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + \gamma Y_{t-1} + \varepsilon_t \\ \varepsilon_t &= \rho\varepsilon_{t-1} + \eta_t. \end{aligned} \tag{104}$$

Для получения первой оценки $\boldsymbol{\beta}$ используется метод инструментальной переменной. В качестве инструмента для Y_{t-1} используются прогнозные значения Y_t в регрессии Y_t на $\mathbf{x}_{t\cdot}$ и $\mathbf{x}_{t-1\cdot}$, то есть

$$\hat{Y}_t = \mathbf{x}_{t\cdot}\hat{\boldsymbol{\alpha}}_1 + \mathbf{x}_{t-1\cdot}\hat{\boldsymbol{\alpha}}_2, \qquad t = 2, \ldots, n,$$

где $\hat{\boldsymbol{\alpha}} = \begin{bmatrix} \hat{\boldsymbol{\alpha}}_1 \\ \hat{\boldsymbol{\alpha}}_2 \end{bmatrix}$ - оценка методом наименьших квадратов вектора $\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{bmatrix}$ в регрессии

$$Y_t = \mathbf{x}_{t\cdot}\boldsymbol{\alpha}_1 + \mathbf{x}_{t-1\cdot}\boldsymbol{\alpha}_2, \qquad t = 2, \ldots, n.$$

Матрица $\mathbf{Z}$ получается добавлением к матрице $\mathbf{X}$ (без первого наблюдения) столбца $[\hat{Y}_2, \hat{Y}_3, \ldots, \hat{Y}_n]^{\mathrm{T}}$. Оценка вектора $\boldsymbol{\beta}$ методом инструментальной переменной рассчитывается по формуле (46). Эта оценка используется для вычисления остатков $\mathbf{e} = \mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}}_{\mathrm{IV}}$. На их основе получается состоятельная оценка для ρ:

$$\hat{\rho} = \frac{\sum\limits_{t=3}^{n} e_t e_{t-1}}{\sum\limits_{t=3}^{n} e_t^2}.$$

Для эффективного оценивания вектора $\boldsymbol{\beta}$ можно применить теперь доступный обобщенный метод наименьших квадратов. Для этого вычисляется оценка МНК вектора параметров в регрессии

$$Y_{*t} = Y_t - \hat{\rho}Y_{t-1}$$

на переменные

$$\mathbf{x}_{*t\cdot} = \mathbf{x}_{t\cdot} - \hat{\rho}\mathbf{x}_{t-1\cdot}, \qquad Y_{*t-1} = Y_{t-1} - \hat{\rho}Y_{t-2}, \qquad e_{t-1} = Y_{t-1} - \mathbf{z}_{t-1\cdot}\hat{\boldsymbol{\beta}}_{\mathrm{IV}}.$$

Если через b обозначить оценку коэффициента при e_{t-1} в этой регрессии, то эффективной оценкой ρ будет:

$$\hat{\hat{\rho}} = \hat{\rho} + b.$$

8 Нестационарность и коинтеграция

Временные ряды экономических показателей обычно нестационарны. Валовые показатели всех экономически развитых стран в послевоенное время имели устойчивую тенденцию к росту. Это нарушает первое из условий стационарности (79). Гренджер и Ньюболд (Granger, Newbold, 1974) показали, что при исследовании зависимости между нестационарными временными рядами методами разделов 1 и 2 мы придем к неверным выводам. Другими словами, распределения вероятностей t-статистики (18) не является распределением Стьдента, F-статистики (28) не является распределением Фишера, распределения всех других статистик из разделов 2 и 3 также не будут классическими. Более того, Филипс (Philips, 1986) доказал в числе прочего, что t-статистика в этом случае не имеет предельного распределения и расходиться при $n \to \infty$. Поэтому, чем больше выборка, тем больше шансов прийти к ложному заключению о наличии связи между несвязанными переменными. *Классическая теория статистического вывода неприменима, если переменные нестационарны.* Такая ситуация известна как «мнимая регрессия». На практике признаками мнимой регрессии могут являться высокие значения коэффициента детерминации (25) при малых значениях статистики Дёрбина-Уатсона (89).

Популярные методы избавления от нестационарности - взятие первых разностей (или разностей логарифмов) и удаление детерминистского тренда в большинстве случаев не согласуются с экономической теорией. Лучший выход был предложен Кливом Гренджером[19]. Пусть у нас есть нестационарный ряд X_t. Возьмём первые разности, то есть рассмотрим ряд $\Delta X_t = X_t - X_{t-1}$. Если ряд ΔX_t оказывается стационарным, то исходный ряд X_t называется *интегрированным порядка* 1. Этот факт обозначается $X_t \sim I(1)$. Стационарный ряд в этой терминологии интегрирован порядка 0: $\Delta X_t \sim I(0)$. Вообще, если временной ряд становится стационарным после взятия k-той разности[20] при том, что $k-1$-ая разность нестационарна, то ряд называется *интегрированным порядка* k, $X_t \sim I(k)$.

Рассмотрим теперь два нестационарных ряда $X_t \sim I(1)$ и $Y_t \sim I(1)$. Их линейная комбинация $Y_t - \beta X_t$, вообще говоря, также является нестационарной и интегрированной порядка 1. Однако, если между экономическими показателями X_t и Y_t имеется долгосрочное экономическое соотношение вида

$$Y_t = \alpha + \gamma X_t + \varepsilon_t, \tag{105}$$

то разность $Y_t - \gamma X_t$ должна колебаться вокруг α, то есть быть стационарной. Следовательно, при значении $\beta = \gamma$ линейная комбинация $Y_t - \beta X_t$ должна быть стационарной. Если такое значение β существует, то ряды X_t и Y_t называют *коинтегрированными*, а вектор $(1, -\beta)^{\mathrm{T}}$ называют *ко-*

[19]За эти результаты Гренджер был удостоен Нобелевской премии по экономике за 2003 год.

[20]Разность порядка k определяется формулой $\Delta^k X_t = \Delta^{k-1} X_t - \Delta^{k-1} X_{t-1}$. Например $\Delta^2 X_t = \Delta X_t - \Delta X_{t-1} = X_t - 2X_{t-1} - X_{t-2}$.

интегрирующим вектором. Вообще, если некоторая линейная комбинация переменных $I(1)$ оказывается стационарной, то эти переменные коинтегрированны, а коэффициенты этой линейной комбинации называют коинтегрирующим вектором. При наличии коинтеграции оказывается возможным получить состоятельную оценку $\hat{\gamma}$ в уравнении (105) обычным методом наименьших квадратов. Более того, как показал Сток (Stock, 1987), $\hat{\gamma}$ сходится к истинному значению γ со скоростью порядка $1/n$ (n - количество наблюдений)[21]. По этой причине такая оценка называется *сверхсостоятельной.*

Важность коинтеграции для моделирования нестационарных экономических временных рядов кроется в теореме Гренджера о представлении (Granger-Weiss, 1983). Продемонстрируем этот результат на следующей системе двух авторегрессивных уравнений:

$$\begin{aligned} X_t &= \nu_1 + \sum_{j=1}^{p} \gamma_{1j} X_{t-j} + \sum_{j=1}^{p} \delta_{1j} Y_{t-j} + \varepsilon_{1t}, \\ Y_t &= \nu_2 + \sum_{j=1}^{p} \gamma_{2j} X_{t-j} + \sum_{j=1}^{p} \delta_{2j} Y_{t-j} + \varepsilon_{2t}, \end{aligned} \tag{106}$$

где Y_t и X_t интегрированы порядка $I(1)$ и коинтегрированы, а ε_{1t}, ε_{2t} независимы и одинаково распределены со средним 0 и постоянной дисперсией. Тогда теорема Гренджера утверждает, что эту систему можно записать в виде:

$$\begin{aligned} \Delta X_t &= \alpha_1 (Y_{t-1} - \nu - \beta X_{t-1}) + \sum_{j=1}^{p-1} \gamma_{1j}^* \Delta X_{t-j} + \sum_{j=1}^{p} \delta_{1j}^* \Delta Y_{t-j} + \varepsilon_{1t}, \\ \Delta Y_t &= \alpha_2 (Y_{t-1} - \nu - \beta X_{t-1}) + \sum_{j=1}^{p-1} \gamma_{2j}^* \Delta X_{t-j} + \sum_{j=1}^{p-1} \delta_{2j}^* \Delta Y_{t-j} + \varepsilon_{2t}, \end{aligned} \tag{107}$$

где по меньшей мере одно из чисел α_1, α_2 отлично от нуля. Оба уравнения системы «сбалансированы» в том смысле, что обе части имеют одинаковый порядок интегрированности, поскольку $Y_{t-1} - \nu - \beta X_{t-1} \sim I(0)$.

Пусть уравнение $Y_t = \nu + \beta X_t$ определяет долгосрочное (или равновесное) соотношение между переменными X_t и Y_t. Тогда $Y_{t-1} - \nu - \beta X_{t-1}$ показывает степень отклонения от состояния равновесия (эту величину иногда называют ошибкой неравновесности). Коэффициенты α_1, α_2 отражают силу стремления рассматриваемой экономической системы к состоянию равновесия. Таким образом, согласно уравнениям (107), изменение

[21] В классическом случае статической модели сходимость имеет порядок $1/\sqrt{n}$. Более точно, в классическом случае $\sqrt{n}(\hat{\gamma} - \gamma)$ имеет предельное нормальное распределение, а в данном случае $n(\hat{\gamma} - \gamma)$ имеет некоторое предельное распределение.

значений переменных за данный период состоит из реакции системы на изменения значений переменных в прошлом, сила которой характеризуется коэффициентами γ_{1j}^*, γ_{2j}^*, δ_{1j}^*, δ_{2j}^* и коррекции к положению равновесия, сила которой характеризуется коэффициентами α_1, α_2. Система в форме (107) называется *моделью коррекции ошибок* (или ECM - error correction model).

Тест для проверки временного ряда на стационарность был разработан Дики и Фуллером (Dickey-Fuller, 1979). При этом они учитывали, что в экономических данных нестационарность чаще всего имеет характер *случайного блуждания*:

$$Y_t = Y_{t-1} + \varepsilon_t, \tag{108}$$

случайного блуждания с дрейфом:

$$Y_t = \alpha + Y_{t-1} + \varepsilon_t, \tag{109}$$

или комбинации случайного блуждания с линейным временным трендом:

$$Y_t = \alpha + Y_{t-1} + \mu t + \varepsilon_t. \tag{110}$$

Здесь ε_t это белый шум[22]. Уравнение (108) есть ни что иное как авторегрессия первого порядка (81) с $\phi_1 = 1$. Таким образом, нам необходимо проверить гипотезу $\phi = 1$ в следующих моделях:

$$Y_t = \phi Y_{t-1} + \varepsilon_t, \tag{111}$$

$$Y_t = \alpha + \phi Y_{t-1} + \varepsilon_t, \tag{112}$$

$$Y_t = \alpha + \phi Y_{t-1} + \mu t + \varepsilon_t. \tag{113}$$

Для проверки используется обычная t-статистика (19), равная в данном случае $(\hat{\phi} - 1)/s_{\hat{\phi}}$, где $\hat{\phi}$ - оценка ϕ методом наименьших квадратов, а $s_{\hat{\phi}}$ находится по формуле (16). Однако, как показали Дики и Фуллер, при условии $\phi = 1$ эта статистика не распределена по закону Стьюдента и ее распределение не стремиться к нормальному с ростом числа наблюдений. Критические значения этой статистики, полученные методом Монте-Карло, представлены в Таблице 1. Если наблюдаемое значение $(\hat{\phi} - 1)/s_{\hat{\phi}}$ ниже указанного в таблице на заданном уровне доверия, то мы должны отвергнуть гипотезу о нестационарности (то есть гипотезу $\phi = 1$). В противном случае данные не противоречат предположению о нестационарности. Интересно отметить, что указанные в Таблице 1 критические значения

[22]**Белым шумом** называют последовательность независимых одинаково распределённых случайных величин со средним ноль и конечной дисперсией. Если распределение этих случайных величин - нормальное, то говорят про белый шум в узком смысле, или гауссовский белый шум.

Таблица 1: Критические значения статистики Дики-Фуллера (DF)

Уровень доверия	Размер выборки			
	25	50	100	∞
AR(1) модель (111)				
0.010	-2.66	-2.62	-2.60	-2.58
0.025	-2.26	-2.25	-2.24	-2.23
0.050	-1.95	-1.95	-1.95	-1.95
AR(1) модель с константой (112)				
0.010	-3.75	-3.58	-3.51	-3.43
0.025	-3.33	-3.22	-3.17	-3.12
0.050	-3.00	-2.93	-2.89	-2.86
AR(1) модель с константой и трендом (113)				
0.010	-4.38	-4.15	-4.04	-3.96
0.025	-3.95	-3.8	-3.69	-3.66
0.050	-3.6	-3.5	-3.45	-3.41

остаются верными, если в правые части уравнений (111)-(113) включить запаздывающие значения скачков Y_t, то есть ΔY_{t-1}, ΔY_{t-2}, и т.д. Получающийся тест называется расширенным тестом Дики-Фуллера (ADF - augmented Dickey-Fuller). Если порядок процесса авторегрессии не известен заранее, то рекомендуется включать возможно большее количество запаздывающих значений скачков Y_t, чтобы избежать автокорреляции возмущений, поскольку приведенные критические значения получены в предположении, что возмущения являются белым шумом. С другой стороны мощность теста уменьшается с увеличением количества переменных.

Для тестирования на наличие коинтеграции между переменными необходимо оценить методом наименьших квадратов предполагаемое равновесное соотношение между ними. После этого на стационарность тестируются остатки этой регрессии. Нулевая гипотеза состоит в отсутствии коинтеграции, то есть в нестационарности остатков. Однако, для этой проверки нельзя использовать DF и ADF тесты. Причина этого кроется в том, что МНК минимизирует среднеквадратическое отклонение остатков от нуля и делает их похожими на стационарные. Поэтому для тестирования остатков на стационарность необходимо использовать другие критические значения для t-статистики. Эти критические значения были получены Дэвидсономи МакКинноном (MacKinnon, 1991; Davidson and MacKinnon, 1993) и приведены в Таблице 2 для случая бесконечного количества наблюдений.

Таким образом, если нам необходимо работать с нестационарными переменными, между которыми экономическая теория предсказывает долгосрочное равновесное соотношение, можно использовать следующий алго-

Таблица 2: Критические значения теста на коинтеграцию Дэвидсона и МакКиннона

Количество переменных	Тип регрессии	Уровень доверия		
		0.01	0.05	0.10
2	константа	-3.90	-3.34	-3.04
	константа и тренд	-4.32	-3.78	-3.50
3	константа	-4.29	-3.74	-3.45
	константа и тренд	-4.66	-4.12	-3.84
4	константа	-4.64	-4.10	-3.81
	константа и тренд	-4.97	-4.43	-4.15
5	константа	-4.96	-4.42	-4.13
	константа и тренд	-5.25	-4.72	-4.43
6	константа	-5.25	-4.71	-4.42
	константа и тренд	-5.52	-4.98	-4.70

ритм:

- Проверить переменные на стационарность при помощи DF или ADF статистики Дики-Фуллера[23]. Если выявлена нестационарность, проверить на стационарность первые разности. В случае из нестационарности проверить также вторые разности[24]. В результате этого шага находим порядок интегрированности имеющихся временных рядов.

- Сформулировать модель предполагаемого равновесного соотношения между переменными. Проверить наличие коинтеграции между переменными как описано выше.

- При наличии коинтеграции оценить коинтегрирующий вектор методом наименьших квадратов. Сверхсостоятельность позволяет рассчитывать на высокое качество этой оценки.

- Сформулировать модель коррекции ошибок, согласованную с равновесным соотношением, предсказываемым экономической теорией.

- Вместо коинтегрирующего вектора подставить его оценки МНК. Оценить оставшиеся параметры модели методом наименьших квадратов или его модификацией.

[23]Точнее на наличие единичного корня. См. описание теста Дики-Фуллера выше.

[24]Можно проверять также разности логарифмов или другие функции от исходных данных, зависящие от предположительного вида модели.

Список литературы

[1] Greene W.H. Econometric Analysis. Fourth Edition. Prentice-Hall International Inc., 2000.

[2] Thomas R.L. Modern Econometrics. An Introduction. Addison Wesley Longman, 1997.

[3] Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс. М.:Дело, 2001.

[4] Магнус Я.Р., Нейдеккер Х. Матричное дифференциальное исчисление с приложениями к статистике и эконометрике. М:Физматлит, 2002.

Содержание

1 Введение **3**

2 Множественная линейная регрессия **9**

3 Линейные ограничения, стабильность параметров и фиктивные переменные **15**

4 Инструментальные переменные **21**

5 Несферические возмущения - общий случай **23**

6 Гетероскедастичность **27**

7 Автокорреляция возмущений **34**

8 Нестационарность и коинтеграция **43**

Карп Дмитрий Борисович

ЭКОНОМЕТРИКА: ОСНОВНЫЕ ФОРМУЛЫ С КОММЕНТАРИЯМИ

Учебно-методическое пособие

Отпечатано по оригинал-макету,
подготовленному автором,
минуя редподготовку.
Вне плана.

Подписано в печать 26.05.04. Формат 60×84/16
Усл. печ. л. 2,8. Уч.-изд. л. 3,0

Тираж 150 экз. Заказ №

Издательство
Дальневосточной государственной академии
экономики и управления
Участок оперативной полиграфии

690950, Владивосток, Океанский пр-т., 19

40-66-35. E-mail:rio@mail.fesaem.ru