

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ

Государственное образовательное учреждение высшего профессионального образования «Уральский государственный университет им. А.М. Горького»

ИОНЦ «Бизнес-информатика»

Экономический факультет

Кафедра экономического моделирования и информатики

ЭКОНОМЕТРИКА

Курс лекций

Подпись руководителя ИОНЦ

Дата

Екатеринбург
2007

1. 1. ЭКОНОМЕТРИЧЕСКОЕ МОДЕЛИРОВАНИЕ. ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ ЭКОНОМЕТРИКИ.	4
2. ПАРНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ.	18
3. МНОЖЕСТВЕННАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ	35
4. ОЦЕНКА КАЧЕСТВА ПОДГОНКИ ЛИНИИ РЕГРЕССИИ К ИМЕЮЩИМСЯ ДАННЫМ	44
5. НЕЛИНЕЙНЫЕ РЕГРЕССИОННЫЕ МОДЕЛИ.	48
6. СТАТИСТИЧЕСКИЕ СВОЙСТВА ОЦЕНОК КОЭФФИЦИЕНТОВ МЛРМ.	52
7. ПРОВЕРКА ГИПОТЕЗ ОТНОСИТЕЛЬНО КОЭФФИЦИЕНТОВ РЕГРЕССИИ.	66
8. МУЛЬТИКОЛЛИНЕАРНОСТЬ.	76
9. ОШИБКИ СПЕЦИФИКАЦИИ	81
10. ПРОЦЕДУРЫ ОТБОРА РЕГРЕССОРОВ	86
11. ГЕТЕРОСКЕДАСТИЧНОСТЬ	90
12. АВТОКОРРЕЛЯЦИЯ ВРЕМЕННЫХ РЯДОВ.	95
13. ОБОБЩЕННЫЙ МЕТОД НАИМЕНЬШИХ КВАДРАТОВ.	101
14. Прогнозирование при помощи регрессионных моделей.	107
15. Временные ряды	111
16. СИСТЕМЫ ОДНОВРЕМЕННЫХ УРАВНЕНИЙ.	136
17. Литература.	154

1.

1. ЭКОНОМЕТРИЧЕСКОЕ МОДЕЛИРОВАНИЕ. ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ ЭКОНОМЕТРИКИ.

Термин «эконометрика» был впервые введен бухгалтером П. Цьемпой (Австро-Венгрия, 1910 г.) («эконометрия» - у Цьемпы). Цьемпа считал, что если к данным бухгалтерского учета применить методы алгебры и геометрии, то будет получено новое, более глубокое представление о результатах хозяйственной деятельности. Это употребление термина, как и сама концепция, не прижилось, но название «эконометрика» оказалось весьма удачным для определения нового направления в экономической науке, которое выделилось в 1930 г.

Слово «эконометрика» представляет собой комбинацию двух слов: «экономика» и «метрика» (от греч. «метрон»). Таким образом, сам термин подчеркивает специфику, содержание эконометрики как науки: количественное выражение тех связей и соотношений, которые раскрыты и обоснованы экономической теорией. Й. Шумпетер (1883—1950), один из первых сторонников выделения этой новой дисциплины полагал, что в соответствии со своим назначением эта дисциплина должна называться «эконометрика».

Советский ученый А.Л. Вайнштейн (1892—1970) считал, что название настоящей науки основывается на греческом слове метрия (геометрия, планиметрия и т.д.), соответственно по аналогии – эконометрия. Однако в мировой науке общеупотребимым стал термин «эконометрика». В любом случае, какой бы мы термин ни выбрали, эконометрика является наукой об измерении и анализе экономических явлений.

Зарождение эконометрики является следствием междисциплинарного подхода к изучению экономики. Эта наука возникла в результате взаимодействия и объединения в особый «сплав» трех компонент: экономической теории, статистических и математических методов. Впоследствии к ним присоединилось развитие вычислительной техники как условие развития эконометрики.

В журнале «Эконометрика», основанном в 1933 г. Р. Фришем (1895—1973), он дал следующее определение эконометрики: «Эконометрика - это не то же самое, что экономическая статистика. Она не идентична и тому, что мы называем экономической теорией, хотя значительная часть этой теории носит количественный характер. Эконометрика не является синонимом приложений математики к экономике. Как показывает опыт, каждая из трех отправных точек - статистика, экономическая теория и математика — необходимое, но недостаточное условие для понимания количественных соотношений в современной экономической жизни. Это - единство всех трех составляющих. И это единство образует эконометрику».

Таким образом, эконометрика - это наука, которая дает количественное выражение взаимосвязей экономических явлений и процессов. Нельзя утверждать, что достигнуто однозначное определение эконометрики.

Так, Э. Маленво придерживался широкого понимания, интерпретируя эконометрику как «любое приложение математики или статистических методов к изучению экономических явлений».

О. Ланге (1904—1965) писал, что эконометрика занимается определением наблюдаемых в экономической жизни конкретных количественных закономерностей, применяя для этой цели статистические методы. Статистический подход к эконометрическим измерениям стал доминирующим.

Возникновение эконометрики как науки.

Каждая наука проходит сложный путь зарождения и выделения в самостоятельную область знания. Эконометрика — не исключение. Первоначальные попытки количественных исследований в экономике относятся к XVII в. «Политические арифметики» - В. Петти (1623-1667). Г. Кинг (1648-1712), Ч. Давенант (1656—1714) — вот первая когорта ученых, систематически использовавших цифры и факты в своих исследованиях, прежде всего в расчете национального дохода.

Круг их интересов был связан в основном с практическими вопросами: налогообложением, денежным обращением, международной торговлей и

финансам. Политическую арифметику можно назвать описательным политико-эконометрическим анализом. Это направление пробудило поиск законов в экономике.

Одним из первых был сформулирован так называемый «закон Книга», в котором на основе соотношения между урожаем зерновых и ценами на зерно была выявлена закономерность спроса. Исследователям хотелось достичь в экономике того, что И. Ньютон достиг в физике.

Неопределенная природа экономических закономерностей еще не была осознана. В этот же период все больше учетных данных становятся доступными, создавая основу для измерений.

Существенным толчком явилось развитие статистической теории в трудах Ф. Гальтона (1822-1911), К. Пирсона (1857-1936), Ф. Эджворта (1845—1926). Появились первые применения парной корреляции: при изучении связей между уровнем бедности и формами помощи бедным (Дж. Э. Юл, 1895, 1896); между уровнем брачности в Великобритании и благосостоянием (Г. Хукер, 1901), в котором использовалось несколько индикаторов благосостояния, к тому же исследовались временные ряды экономических переменных. Это были шаги по созданию современной эконометрики.

Параллельно происходил процесс создания маржиналистской (неоклассической) теории, зарождение которой можно датировать 60-ми годами XIX о. (появление работ С.Джепонса, Л.Вальраса, К.Менгера).

С 30-х гг. XIX в. страны с наиболее высоким уровнем развития капитализма стали испытывать спорадические потрясения — упадок деловой активности, возникновение массовой безработицы. Эти явления не находили теоретического объяснения. Быстрая индустриализация выявила огромный диапазон социальных проблем, которые также не согласовывались с теорией.

Неоклассическая теория стала восприниматься как слишком удаленная от действительности. Для ее практического значения требовались количественные выражения базовых понятий, таких как «эластичность спроса» или «предельная полезность».

Теория спроса могла стать убедительной в том случае, если она смогла бы объяснить и оценить фактические кривые спроса и предложения, продемонстрировать формирование равновесных цен в конкретных условиях.

К этому же времени относится привлечение ученых-экономистов (А. Маршалла, С. Джепенса, К. Менгера) к парламентской деятельности, что подтолкнуло их к анализу макроэкономических проблем на основе временных рядов таких показателей, как, например, валютные курсы и т.п. Это также явилось важным шагом в подготовке развития эконометрики.

Многие исследователи признают первой работой, которая могла бы быть названа эконометрической, книгу американского ученого Г. Мура (1869—1958) «Законы заработной платы: эссе по статистической экономике» (1911). Г. Муром были проведены анализ рынка труда, статистическая проверка теории производительности Дж. Кларка, а также изложены основы стратегии объединения пролетариата и т. д.

В это время для США решение этих вопросов было безотлагательным: рабочий класс стремительно рос, возникали такие объединения, как «Индустриальные рабочие мира» и другие радикально настроенные организации. Г. Мур подошел к анализу поставленных проблем с позиций «высшей», как он называл, статистики, используя все достижения теории корреляции, регрессии, анализа динамических рядов. Он стремился показать, что сложные математические построения, наполненные фактическими данными, могли составить основу для разработки социальной стратегии.

К этому же периоду относится первое применение итальянским ученым Р. Бенини (1862-1956) метода множественной регрессии для оценки функции спроса. Значительным вкладом в становление эконометрики явились исследования по цикличности экономики.

К. Жюгляр (1819-1905), французский физик, ставший экономистом, первым занялся исследованием экономических временных рядов с целью выделения бизнес-циклоп. Им была обнаружена цикличность инвестиций (продолжительность цикла - 7—11 лет). Вслед за ним С. Китчин, С. Кузнец, Н.

Кондратьев, автономно занимаясь этой проблемой, выявили цикличность обновления оборотных средств (3 - 5 лет), циклы в строительстве (15 - 20 лет), долгосрочные волны, или «большие циклы» Кондратьева, продолжительностью 45—60 лет.

Значительной вехой в формировании эконометрики явилось построение экономических барометров, прежде всего так называемого гарвардского барометра. Большинство экономических барометров, включая названный, основано на следующей идее: в динамике различных элементов экономики существуют такие показатели, которые в своих изменениях идут впереди других, а потому могут служить предвестниками последних.

Гарвардский барометр был создан под руководством У. Персонса (1878—1937) и У. Митчелла (1874-1948). В течение 1903-1914 гг. он состоял из пяти групп показателей, которые в дальнейшем были сведены в три отдельные кривые: кривая А характеризовала фондовый рынок; кривая В — товарный рынок; кривая С — денежный рынок.

Каждая из этих кривых представляла среднюю арифметическую из рядов входящих в нее нескольких показателей. Эти ряды предварительно статистически обрабатывались путем исключения тенденции, сезонной волны и приведения колебаний отдельных кривых к сравнимому масштабу колеблемости.

В основу прогноза гарвардского барометра было положено свойство каждой отдельной кривой повторять движение остальных в определенной последовательности и с определенным отставанием.

Так, с 1903 г. и до первой мировой войны поворотные пункты кривой А предшествовали поворотным пунктам кривой В на 6—10 месяцев (в среднем — на 8 месяцев); поворотные пункты кривой В обгоняли аналогичные пункты кривой С на 2-8 месяцев (в среднем на 4 месяца); наконец, колебания кривой С предшествовали колебаниям кривой А следующего цикла на 6-12 месяцев.

Гарвардский барометр представлял собой описание подмеченных эмпирических закономерностей и экстраполяции последних на ближайшие

месяцы. Однако в построении гарвардского барометра можно обнаружить и некоторые теоретические предпосылки. Естественно, например, что изменение средних биржевых курсов и показателей фондового рынка (индекс спекуляции А) означало изменение спроса на товары, что влекло за собой, в свою очередь, изменение в том же направлении индекса оптовых цен, объема производства и товарооборота (индекс В). Возрастание, например, объема производства вызывало напряжение на денежном рынке, рост учетной ставки и падение курса ценных бумаг с фиксированным доходом (кривая С). Поэтому максимум кривой А обычно должен был совпадать с минимумом кривой С.

Успех гарвардского барометра породил буквально эпидемию таких построений в других странах (в частности, аналогичный барометр был построен в Великобритании). Несколько лет после первой мировой войны он еще удовлетворительно выполнял свое предназначение.

Но затем гарвардский барометр (приблизительно с 1925 г.) потерял чувствительность и сошел со сцены, пережив свою славу. Авторы гарвардского барометра объясняли его крах появлением мощного регулирующего фактора в экономике США. В этих условиях основным методом макроэкономического анализа становится метод «Затраты-выпуск» В.В. Леонтьева (1906-1999).

Что касается экономических барометров, то советский математик-статистик Е. Слуцкий (1880-1948) в работе «Сложение случайных причин как источник циклических процессов» (1927), взяв в качестве случайных рядов последние цифры номеров облигаций из тиражных таблиц выигрышного займа, блестяще доказал, что сложение случайных причин порождает волнообразные ряды, имеющие тенденцию на протяжении большего или меньшего числа волн имитировать гармонические ряды, сложенные из небольшого числа синусоид». Таким образом, никакой закономерности в любом экономическом барометре могло и не существовать.

В этот же период делались эконометрические построения, использующие методы гармонического анализа и периодограмм-анализа (Г.Мур в США,

Бэвэридж в Энстром в Швеции) Эти методы перенесены в экономику из области астрономии, метеорологии, физики.

К 30-м гг. сложились все предпосылки для выделения эконометрики в отдельную науку. Стало ясно, что специалисты, занимающиеся развитием эконометрической науки, должны использовать в той или иной степени математику и статистику. Возникла необходимость появления особого термина, объединяющего все исследования в этом направлении, подобно биометрике — науке, изучающей биологию статистическими методами.

В 1912г. И. Фишер попытался создать группу ученых для стимулирования развития экономической теории путем ее связи со статистикой и математикой. Но тогда эту группу создать не удалось. Тогда Р. Фриш и математик-экономист Ч. Рус обратились с идеей собрать специальный форум экономистов, готовых к использованию математики и статистики.

29 декабря 1930 г. по инициативе И. Фишера (1867—1947), Р. Фриша, Я. Тинбергена (1903-1995). И. Шумпетера, О. Андерсона (1887-1960) и других ученых на заседании Американской ассоциации развития науки (США, Кливленд, штат Огайо) было создано эконометрическое общество, на котором норвежский ученый Р. Фриш дал новой науке название — «эконометрика».

С самого начала эконометрическое общество было интернациональным. Уже в 1950 г. общество насчитывало почти 1000 членов. С 1933 г. под редакцией Р. Фриша стал издаваться журнал «Эконометрика» («Econometrica»), который и сейчас играет важную роль в развитии эконометрической науки. В 30—40-е гг. развитию эконометрики способствовала деятельность Департамента прикладной экономики под руководством Р. Стоуна (Великобритания). В 1941 г. появился первый учебник по эконометрике, который был создан Я. Тинбергеном (1913-1994).

В эти годы вплоть до 70-х гг. XX в. эконометрика понималась как эмпирическая оценка моделей, разработанных экономической теорией. Р. Фриш определял соотношение между теорией и данными наблюдений

следующим образом: теория, абстрактно формулирующая количественные соотношения, должна быть проверена множеством наблюдений.

Свежие статистические данные и другие факты должны предотвратить теорию от опасного догматизма. Под влиянием лидеров, таких как Р. Фриш, Т. Хаавелмо, Я. Тинберген, Л. Клейн, экономические модели, построенные в этом периоде, всегда были кейнсианскими.

Все изменилось в 70-е гг. В макроэкономике возникли противоречия между кейнсианцами, монетаристами и марксистами. Формальные методы стали использоваться для доказательства причинности при выборе теоретических концепций. Экономическая теория потеряла свое решающее значение.

Другим важным событием стало появление компьютеров с высоким быстродействием и мощной оперативной памятью. Существенное развитие получил статистический анализ временных рядов. Г. Бокс и Г. Дженкинс создали ARIMA-модель в 1970 г., а К. Симс и другие ученые — VAR-модели, ставшие популярными в начале 80-х гг. Вершиной этой стадии развития явился метод коинтеграции, развитый С. Йохансеном и др. (1990 г.).

В настоящее время эконометрика располагает огромным разнообразием типов моделей - от больших макроэкономических моделей, включающих несколько сот, а иногда и тысяч уравнений, до малых коинтеграционных моделей, предназначенных для решения специфических проблем.

Итак, современный взгляд на эконометрику отражен в следующем определении:

Эконометрика – научная дисциплина, объединяющая совокупность теоретических результатов, приемов, методов и моделей, предназначенных для того, чтобы на базе

1. экономической теории;
2. экономической статистики;
3. математико-статистического инструментария

придавать конкретное количественное выражение общим (качественным) закономерностям, обусловленным экономической теорией. (С. А. Айвазян, В. С. Мхитарян. Прикладная статистика и основы эконометрики.)

Иными словами, эконометрика позволяет на базе положений экономической теории и исходных данных экономической статистики, используя необходимый математико-статистический инструментарий, придавать конкретное количественное выражение общим (качественным) закономерностям.

Прикладные цели эконометрики.

- вывод экономических законов;
- формулировка экономических моделей, основываясь на экономической теории и эмпирических данных;
- оценка неизвестных величин (параметров) в этих моделях;
- прогнозирование и оценка точности прогноза;
- выработка рекомендаций по экономической политике.

Как же экономист добивается поставленных перед собой целей. В ходе эконометрического исследования экономист последовательно проходит несколько этапов. Этапы эконометрического моделирования:

1. Человек, начинающий изучать экономику, первым делом приходит к мысли, что в экономике некоторые переменные взаимосвязаны. Формирующийся на рынке спрос на товар рассматривается как функция его цены, затраты, связанные с изготовлением некоторого продукта предполагаются зависимыми от объема производства, потребительские расходы связаны с доходом и др. – примеры связей между двумя переменными, причем одна из переменных выступает в роли объясняемой переменной, другая в роли объясняющей. Для большей реалистичности приходится вводить в соотношение другие объясняющие переменные и случайный фактор. Спрос на товар – цена, потребительский доход, цены на конкурирующие, дополняющие и замещающие товары и др. Переменную, процесс формирования значений которой нас по каким-то причинам интересует, будем обозначать Y и называть

зависимой или объясняемой. Переменные, которые, как мы предполагаем, оказывают влияние на переменную Y , будем обозначать X_j и называть независимыми или объясняющими. Значения этих переменных являются для нее внешними, ничего о том, как формируются эти значения не указано.

На этом этапе процесс формирования значений объясняемой переменной можно представить в виде следующей схемы:

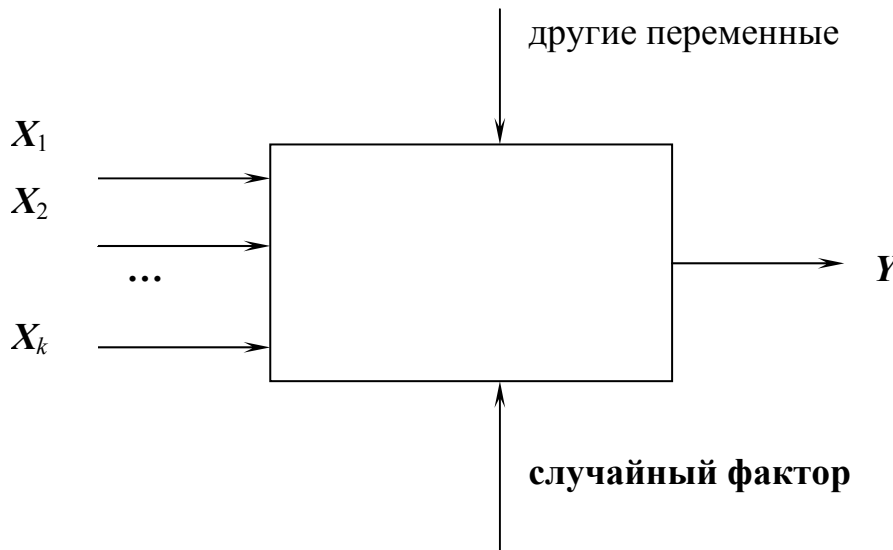


Рис. 1. Формирование значений переменной Y

X_1, \dots, X_k – выделенные переменные (наиболее существенно влияющие или представляющие для нас определенный интерес).

2. Группировка отдельных соотношений в модель – формулирование некоторых гипотез относительно того, как должны быть связаны переменные. Гипотезы эти возникают на основе теоретических экономических предпосылок, интуиции, опыта исследователя, его здравого смысла. Сразу же возникает вопрос, как проверить правильность модели? В физике, биологии все просто – проводим эксперимент и смотрим, подтверждают ли его результаты наши гипотезы. В экономике все сложнее. Как в экономике проводить эксперименты -? Мы можем только наблюдать за действительностью.

Таким образом, на этом этапе эконометрист занимается моделированием поведения экономических объектов. Моделирование – упрощение реальности объекта. Задача, искусство моделирования состоит в том, чтобы как можно

более лаконично и адекватно именно те стороны реальности, которые интересуют исследователя.

Математическая модель схемы:

$$Y = f(X_1, \dots, X_k) + \varepsilon \quad (1.1).$$

Если $M(\varepsilon | X_1, \dots, X_k) = \mathbf{0}$, то уравнение (1.1) называют *уравнение регрессии* Y на X_1, \dots, X_k . Функция f – *регрессионная функция*, линия, которую эта функция описывает в пространстве – *линия регрессии*.

Первая задача – перевести эти предположения на математический язык. К сожалению, проделать это единственным образом нельзя. Что означает возрастающая функция. Есть много функций, которые являются возрастающим функциями своих аргументов. Линейные, нелинейные, разные.

Выход – первоначально сформулировать самую простую модель. Введем следующие обозначения для наблюдаемых экономических параметров:

Уравнение поведения здесь имеют форму точных функциональных зависимостей. Однако, как мы увидим позднее, это нереалистично и нельзя приступать к эконометрической разработке, не пользуясь некоторыми дополнительными стохастическими спецификациями. Мы добавляем в поведенческие уравнения стохастический член. Поскольку ни для каких реальных экономических данных нельзя обеспечить постоянное соблюдение простых соотношений. Кроме того, из всех возможных объясняющих переменных в спецификацию включается лишь небольшое их подмножество, т. е. мы можем говорить только об аппроксимации моделью некоторых, по-видимому достаточно сложных, но неизвестных нам взаимосвязей. Чтобы обеспечить равенство между правой и левой частью, в каждое соотношение приходится вводить случайную ошибку.

В нашей модели рассматриваются зависимости между некоторыми переменными. Переменные, значения которых объясняются в рамках нашей модели, называются *эндогенными*. Переменные, значения которых нашей моделью не объясняются, являются для нее внешними, ничего о том, как формируются эти значения, мы не знаем, называются *экзогенными*. Еще одна

переменная – лаговое значение эндогенной переменной. Она тоже для нашей модели внешняя. Экзогенные переменные и лаговые значения эндогенных переменных – предопределенные переменные.

В ходе курса мы столкнемся с несколькими типами эконометрических моделей, которые применяются для анализа и прогноза:

а) модели временных рядов. Такие модели объясняют поведение переменной, меняющейся с течением времени, исходя только из его предыдущих значений. К этому классу относятся модели тренда, сезонности, тренда и сезонности (аддитивная и мультипликативная формы) и др.

б) регрессионные модели с одним уравнением. В таких моделях зависимая (объясняемая) переменная представляется в виде функции от независимых (объясняющих) переменных и параметров. В зависимости от вида функции модели бывают линейными и нелинейными. Будем изучать именно их.

в) Системы одновременных уравнений. Ситуация экономическая, поведение экономического объекта описывается системой уравнений (наш пример). Системы состоят из уравнений и тождеств, которые могут содержать в себе объясняемые переменные из других уравнений (поэтому вводят понятия экзогенных и эндогенных переменных).

Пункт 2 носит название спецификация модели. Необходимо:

- а) определить цели моделирования;
- б) определения списка экзогенных и эндогенных переменных;
- в) определение форм зависимостей между переменными;
- г) формулировка априорных ограничений на случайный член, что важно для свойств оценок и выбора метода оценивания, и некоторые коэффициенты

3. Теперь необходимо модель проверить. Как это сделать, если мы не физики и не биологи? Методы эконометрии, позволяющие проводить эмпирическую проверку теоретических утверждений и моделей, выступают мощным инструментом развития самой экономической теории. С их помощью отвергаются теоретические концепции и принимаются новые, более полезные

гипотезы. Теоретик, не привлекающий эмпирический материал для проверки своих гипотез и не использующий для этого эконометрические методы, рискует оказаться в мире своих фантазий. Собрать данные – необходимый статистический материал. Здесь нам на помощь приходят методы экономической статистики и статистики вообще.

Типы данных, с которыми эконометристу приходится сталкиваться при моделировании экономических процессов:

а) cross-sectional data – пространственные данные – набор сведений по разным экономическим объектам в один и тот же момент времени;

б) time-series data – временные ряды – наблюдение одного экономического параметра в разные периоды или моменты времени. Эти данные естественным образом упорядочены во времени. Инфляция, денежная эмиссия (годовые), курс доллара США (ежедневные);

в) panel data – панельные данные – набор сведений по разным экономическим объектам за несколько периодов времени (данные переписи населения).

4. Идентификация модели – статистический анализ модели и, прежде всего – статистическое оценивание параметров. Выбор метода оценивания сюда тоже входит. Зависит от особенностей модели.

5. Верификация модели – сопоставление реальных и модельных данных, проверка оцененной модели с тем, чтобы прийти к выводу о достаточной реалистичности получаемой с ее помощью картины объекта, либо признать необходимость оценки другой спецификации модели.

Схема эконометрического исследования

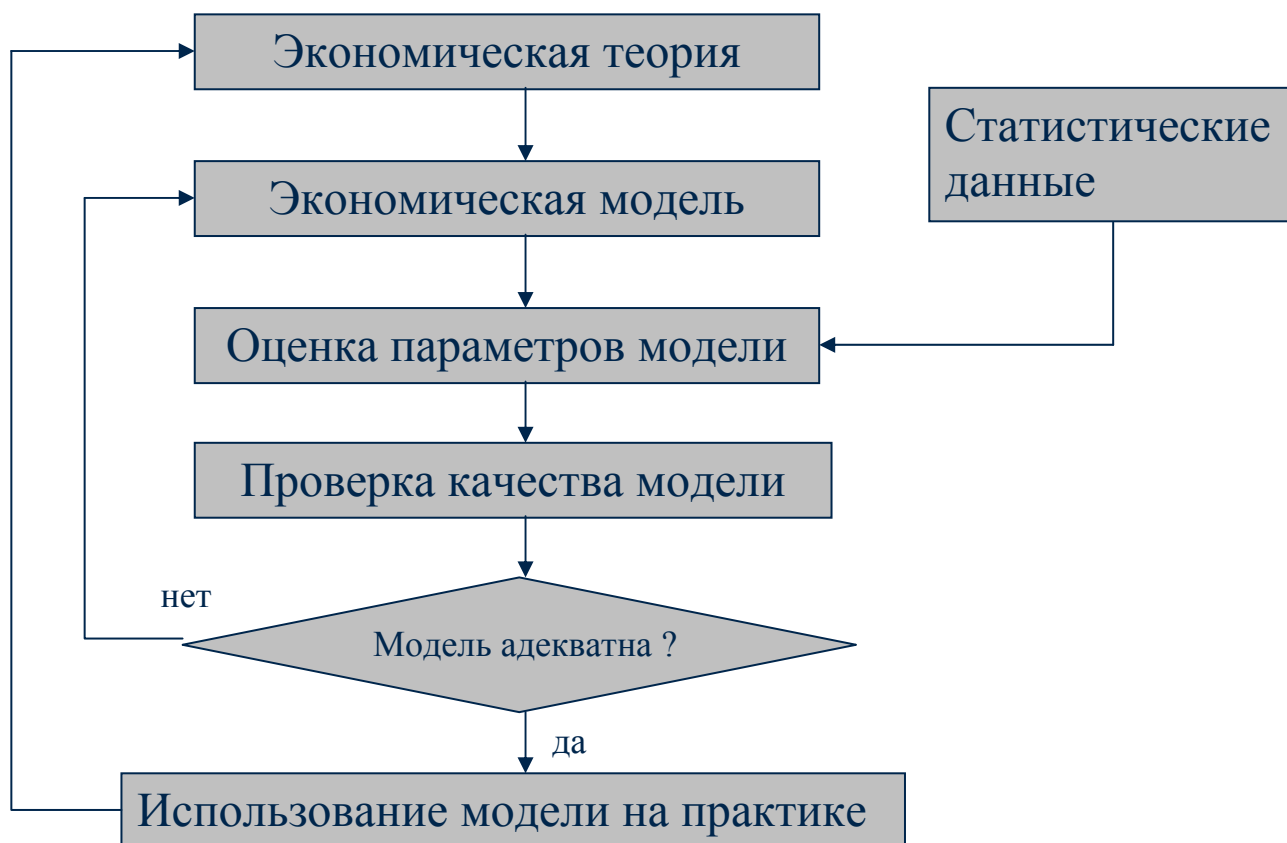


Рис. 2. Схема эконометрического исследования

Итак, эконометрические методы разработаны, в основном, для оценивания параметров экономических моделей. Каждая модель содержит, как правило, несколько уравнений, а в уравнение входит несколько переменных. Начнем с самого простого – парной линейной регрессионной модели.

2. ПАРНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ.

Эконометрические методы разработаны, в основном для оценивания параметров экономических моделей. Каждая модель содержит, как правило, несколько уравнений, а в каждое уравнение входит несколько переменных. Чтобы понять техническую основу, на которой возникают эти достаточно сложные методы, мы начнем с рассмотрения самого простого случая, рассмотрев одно уравнение, которое содержит только две переменных.

Итак, у нас есть переменная Y , зависимая или эндогенная, характеризующая результат или эффективность функционирования экономической системы, которую мы анализируем. Ее значения формируются в процессе и внутри функционирования этой системы под воздействием ряда других переменных и факторов. По своему характеру переменная Y всегда случайна. Есть набор объясняющих переменных, экзогенных, характеризующих состояние экономической системы. Эти переменные в существенной степени объясняют процесс формирования эндогенных переменных. Эти переменные, как правило, поддаются хотя бы частичному регулированию и управлению. По своей природе они могут быть как случайными, так и детерминированными.

Две переменные могут быть связаны либо функциональной зависимостью (т.е. существует функция f что $Y = f(X)$, значения переменной Y полностью определяются значениями переменной X), либо статистической, либо быть независимыми.

Определения.

1. Если при изменении X меняется закон распределения случайной величины Y , то говорят, что величины (X, Y) связаны статистической зависимостью.

2. Статистическая зависимость называется корреляционной, если при изменении X меняется Математическое ожидание случайной величины Y .

Приведем пример случайной величины Y , которая не связана с величиной X функционально, а связана корреляционно. Пусть Y – урожай зерна, а X – количество удобрений. С одинаковых по площади участков земли при равных количествах внесенных удобрений снимают различный урожай, т. е. Y не является функцией от X . Это объясняется влиянием случайных факторов (осадки, температура воздуха и др.). Вместе с тем, как показывает опыт, средний урожай является функцией от количества удобрений, т. е. Y связан с X корреляционной зависимостью.

3. Условным математическим ожиданием $M(Y|X) = M(Y|X = x)$ – условное мат. ожидание переменной Y при фиксированном значении X .

Если каждому значению величины X соответствует свое значение $M(Y|X)$, то говорят, что существует регрессионная функция

$$M(Y|X) = f(X)$$

Это уравнение называют уравнением регрессии Y на X . Т. о. $f(X) = M(Y|X=x)$ – описывает изменение условного среднего значения результирующей переменной в зависимости от изменения значений X объясняющих переменных.

Функциональная зависимость наблюдается крайне редко. Тем не мене, большая часть традиционных экономических теорий, в которой связи между экономическими категориями отражаются с помощью формул, имеют дело с точными алгебраическими соотношениями. Однако если мы посмотрим на отдельные наблюдения переменных, фигурирующих в этих законах, то мы увидим, что они не будут точно соответствовать этим соотношениям. (Функция Кобба-Дугласа, например). Кроме того, они почти никогда не будут соответствовать любому другому гладкому соотношению. В учебниках по экономической теории эта проблема решается обычно следующим образом: соотношение приводится, как если бы оно было точным, а читателя предупреждают, что это только аппроксимация. Но нас с вами такой подход устраивать не должен. В математической статистике факт точности соотношения признается путем включения в уравнение случайного фактора, описываемого случайным остаточным членом. В простейшей модели

$$Y = \alpha + \beta X.$$

Величина Y , рассматриваемая как зависимая переменная, состоит из двух составляющих:

- 1) неслучайной (детерминированной) составляющей $\alpha + \beta X$, где X выступает как объясняющая (независимая) переменная;
- 2) случайного члена ε .

Откуда берется этот случайный член. Причин может быть несколько и основная:

- 1) невключение объясняющих переменных в уравнение. На самом деле на переменную Y влияет не только переменная X , но и ряд других переменных, которые не учтены в нашей модели по следующим причинам:

- a) мы знаем, что другая переменная влияет, но не можем ее учесть, потому как не знаем, как измерить (психологический фактор, например);

- b) существуют факторы, которые мы знаем, как измерить, но влияние их на Y так слабо, что их не стоит учитывать;

- c) существенные переменные, но из-за отсутствия опыта или знаний мы их таковыми не считаем.

Если бы мы точно знали, какие переменные сюда входят и как их надо измерять и имели бы возможность точно их измерить, мы бы могли включить их в уравнение, исключив тем самым соответствующий элемент из случайного члена. Проблема состоит в том, что мы никогда не можем быть уверены, что входит в данную совокупность, а что нет. Даже если бы мы включили все эти факторы в уравнение, то мы бы могли оказаться в ситуации, когда число факторов превысило бы число наблюдений так, что любое статистическое усреднение потеряло бы всякий смысл. Итак, мы можем сказать, что вместо зависимости $Y = f(X_1, \dots, X_n)$, где n слишком велико для практических целей, мы рассматриваем зависимость с меньшим числом наиболее важных переменных или переменных, которые представляют для нас наибольший интерес.

2) Неправильная функциональная спецификация. Функциональное соотношение между Y и X может быть определено неправильно. Например, мы предположили линейную зависимость, а она может быть более сложной.

3) Ошибки наблюдений (занижение реального уровня доходов). В этом случае наблюдаемые значения не будут соответствовать точному соотношению, и существующее расхождение будет вносить свой вклад в остаточный член.

Остаточный член является суммарным проявлением всех факторов. Если бы он отсутствовал, мы бы знали, что каждое изменение Y от наблюдения к наблюдению вызвано изменением X и смогли бы точно вычислить коэффициенты. А так каждое изменение Y вызвано изменением X и ε , поэтому ε иногда называют шумом.

Итак, мы предполагаем, что значения результирующей переменной Y выступают в роли функции, значения которой определяются с некоторой погрешностью значениями объясняющей переменной X , выступающих в роли аргументов этой функции. Математически это может быть выражено в виде уравнения регрессионной связи:

$$\begin{cases} Y(X) = f(X) + \varepsilon(X) \\ M[\varepsilon(X) | X] = 0 \end{cases}$$

где $f(x) = M[Y | X]$, $\varepsilon(X) = Y(X) - f(X)$

Последнее соотношение в следует из смысла функции регрессии, действительно, поскольку $M[Y(X)|X] = M[f(X)|X] + M[\varepsilon | X]$, а $M(Y(X)|X)=f(X)$ по определению, $M[f(X)|X]=f(X)$, поскольку величина $f(X)$ при фиксированных значениях параметра X не является случайной.

Содержательные соображения должны подсказать нам форму $f(X)$ – теория, интуиция, опыт, анализ эмпирических данных. Выбор вида функции $f(X)$ – спецификация модели. Одним и тем же условиям могут удовлетворять несколько различных функций, поэтому нам придется обратиться к статистическому анализу и с его помощью осуществить выбор из возможных

альтернативных вариантов. Начинают, обычно, с самого простого соотношения между двумя переменными – линейного.

Возможны и другие формы зависимости (примеры).

Выбор формы зависимости можно осуществить при помощи графического анализа материала наблюдений. В парном случае материал наблюдений представляет собой набор пар чисел: $(X_i, Y_i) \quad i = 1 \dots N$. На плоскости каждому такому наблюдению соответствует точка:

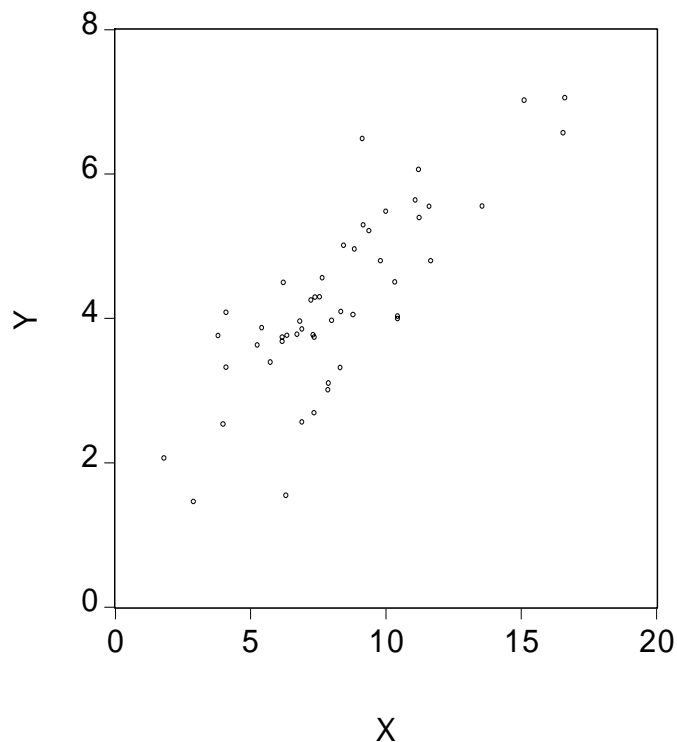


Рис. 1. Линейная зависимость $Y = \alpha + \beta X + \varepsilon$.

Полученный график называют *облако наблюдений*, *поле корреляции* или *диаграмма рассеяния*. По виду облака наблюдений можно определить вид регрессионной функции. На приведенном графике выше – линейная. Другие примеры:

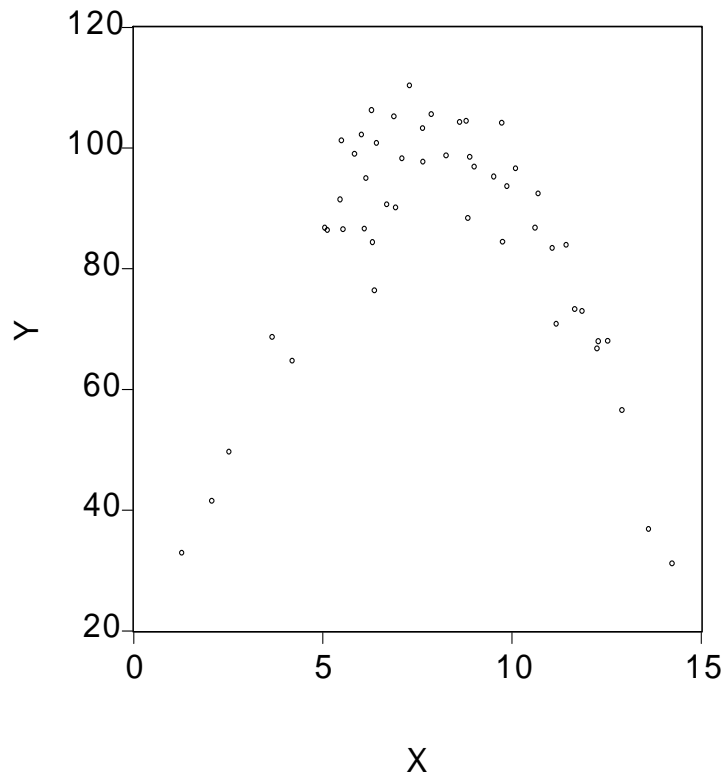


Рис. 2. Квадратичная зависимость: $Y = \alpha + \beta X + \gamma X^2 + \varepsilon$

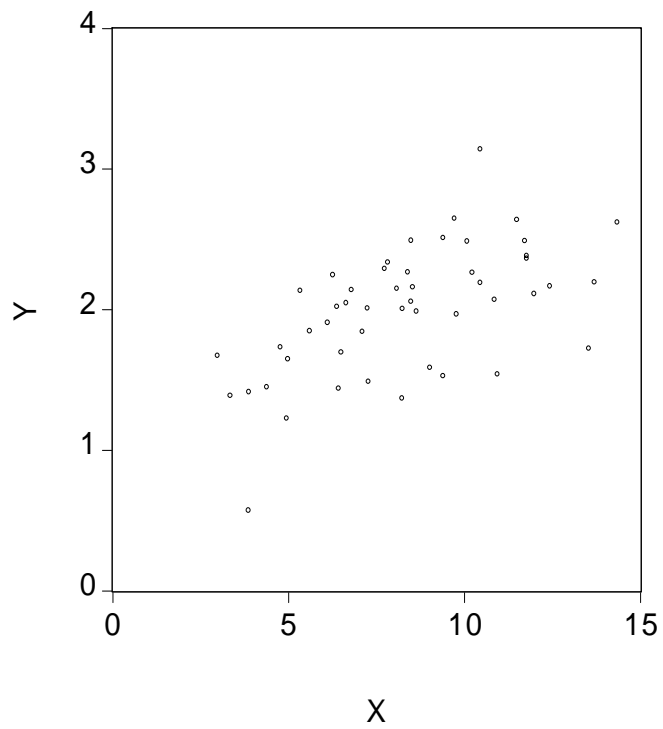


Рис. 3. Показательная зависимость $Y = AX^\alpha \varepsilon$

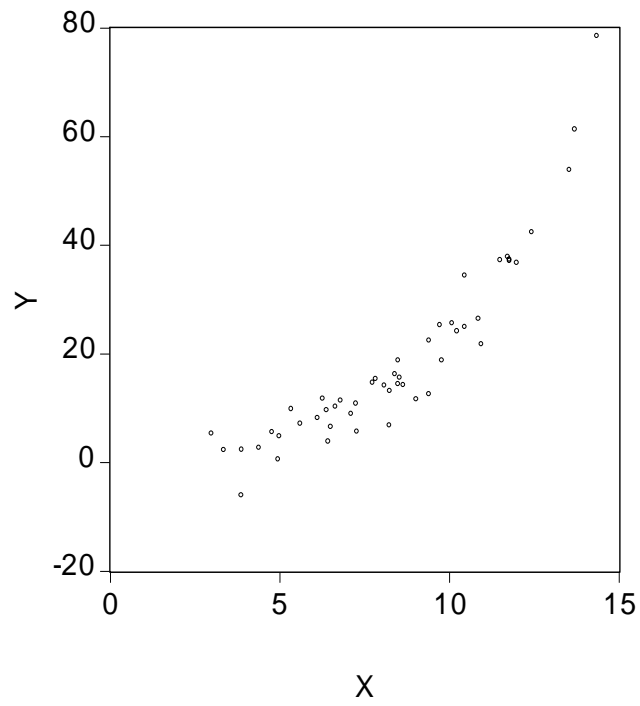


Рис. 4. Степенная зависимость $Y = Ae^{\beta X} \varepsilon$

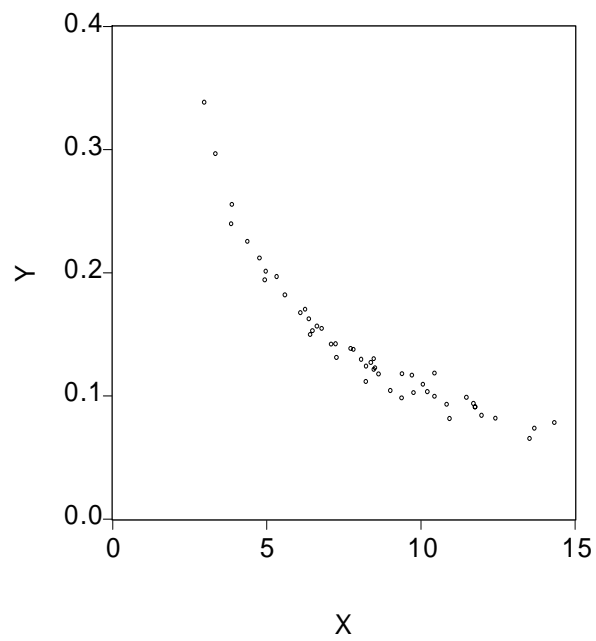


Рис. 5. Гиперболическая зависимость $Y = \alpha + \frac{\beta}{X} + \varepsilon$

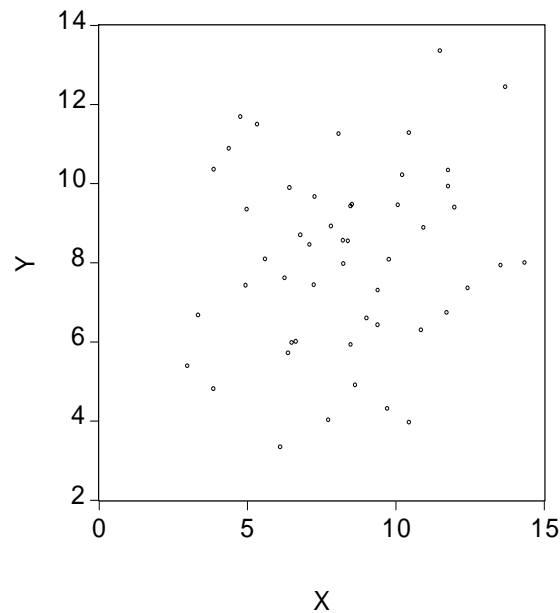


Рис. 6. X и Y независимы.

Нас интересуют только те формы зависимости, которые путем преобразования переменных и параметров можно свести к линейным. Т. е. после преобразования переменных и коэффициентов новые переменные и ошибка будут связаны линейным соотношением. Для нелинейных соотношений так же разработан метод оценивания – нелинейный МНК, метод максимального правдоподобия.

Рассмотрим *парную линейную модель*:

$$Y = \alpha + \beta X + \varepsilon.$$

Для оценки коэффициентов этого уравнения у нас есть набор наблюдений переменной X и соответствующий набор наблюдений переменной Y . Всего у нас N пар чисел (X_i, Y_i) . Этот набор наблюдений называется выборкой. Расположим их на плоскости. Если бы соотношение между Y и X было бы точным, то соответствующие значения Y лежали бы на прямой. Наличие случайного члена приводит к тому, что в действительности значения Y на прямой не лежат.

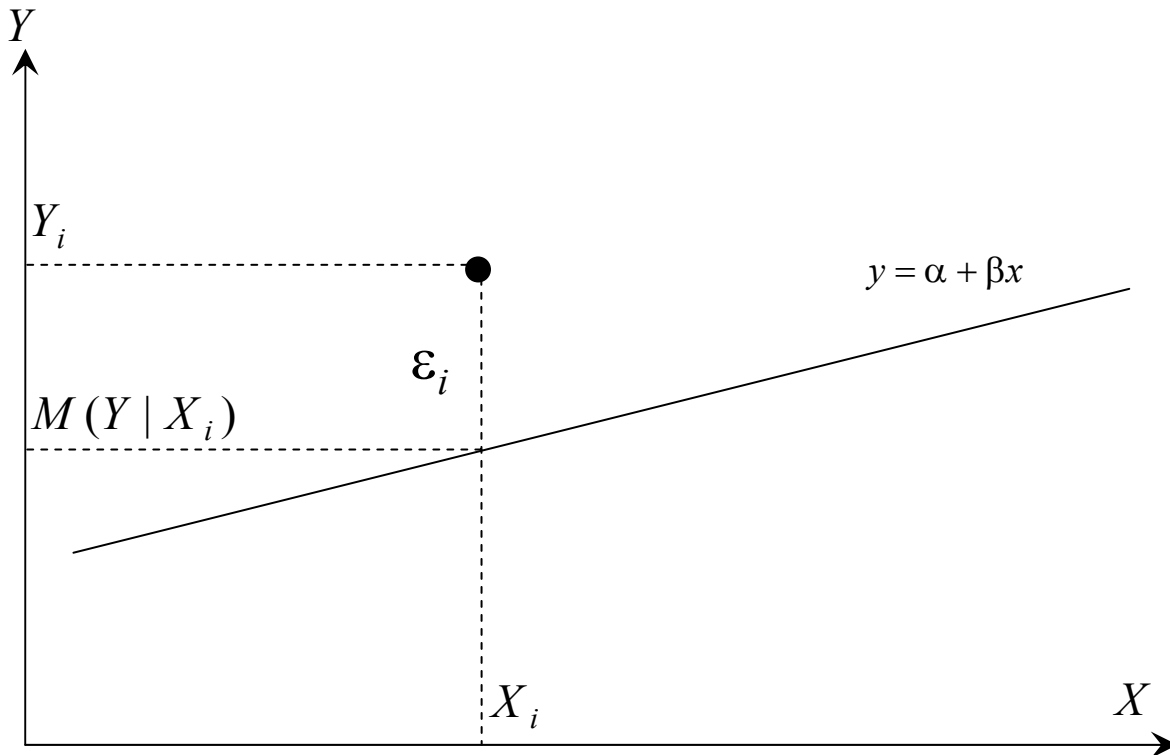


Рис. 7. Ошибки на графике

$Y_i = \alpha + \beta X_i + \varepsilon_i$ – выполняется для каждого наблюдения. α , β и ε_i нам неизвестны и никогда не будут известны. Мы сможем получить только оценки, хорошие или плохие. Они могут случайным образом совпасть с реальными значениями, но мы этого никогда не узнаем.

Каким образом получить эти оценки? Мы предположили, что переменные Y и X связаны линейной зависимостью, т.е. эта зависимость описывается прямой линией. И теперь наша задача – построить прямую. Из всех возможных прямых мы хотим выбрать ту, чтобы она «наилучшим образом» подходила к нашим данным, т. е. отражала бы линейную зависимость Y от X . Иными словами, мы хотим чтобы каждое Y_i лежало бы как можно ближе к прямой.

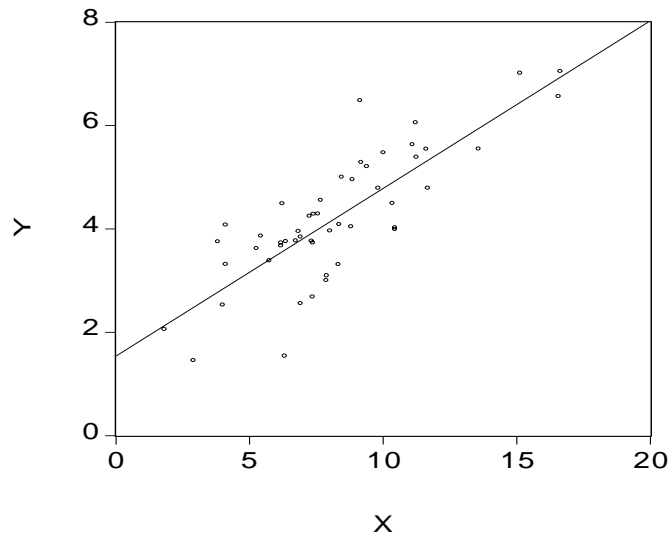


Рис. 8. Облако наблюдений и линия регрессии.

В качестве меры близости точек к прямой мы введем разность

$Y_i - a - bX_i = Y_i - \hat{Y}_i = e_i$ - остаток или невязка регрессии.

\hat{Y}_i - прогнозное значение переменной Y в i -м наблюдении.

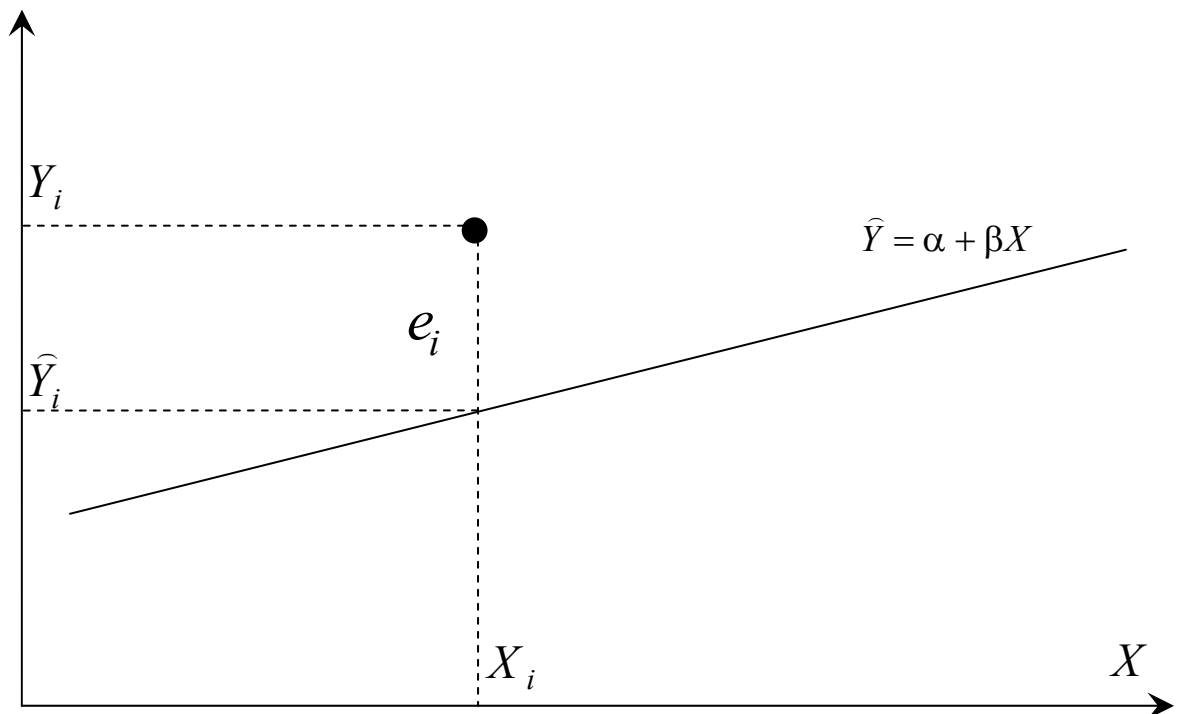


Рис. 10. Остатки на графике.

Можно сказать, мы хотим, чтобы желаемая прямая *была бы в центре скопления наших данных*, т. е. чтобы все Y_i как можно ближе лежали к нашей прямой.

Очевидно, значения a и b надо подбирать таким образом, чтобы минимизировать некоторую интегральную (т. е. по всем имеющимся наблюдениям) характеристику невязок или остатков:

1. $\sum_{i=1}^N e_i^2 = \sum_{i=1}^N (Y_i - a - bX_i)^2 = F(a, b) \rightarrow \min_{(a, b)}$
2. $\sum_{i=1}^N |e_i| = \sum_{i=1}^N |Y_i - a - bX_i| = F(a, b) \rightarrow \min_{(a, b)}$

МНК – минимизируем 1. Для нахождения минимума функции двух переменных, нам надо взять частные производные по каждой из них и приравнять их к нулю:

$$\frac{\partial F}{\partial a} = -2 \sum_{i=1}^N (Y_i - a - bX_i) = 0$$

$$\frac{\partial F}{\partial b} = -2 \sum_{i=1}^N (Y_i - a - bX_i)X_i = 0, \text{ или}$$

$$\begin{cases} \sum_{i=1}^N (Y_i - a - bX_i) = 0 \\ \sum_{i=1}^N (Y_i - a - bX_i)X_i = 0 \end{cases}$$

$$\text{или} \begin{cases} \sum_{i=1}^N e_i = 0 \\ \sum_{i=1}^N X_i e_i = 0 \end{cases}$$

Преобразуем систему:

$$\begin{cases} Na + b \sum_{i=1}^N X_i = \sum_{i=1}^N Y_i \\ a \sum_{i=1}^N X_i + b \sum_{i=1}^N X_i^2 = \sum_{i=1}^N X_i Y_i \end{cases}$$

Эта система называется система нормальных уравнений для нахождения коэффициентов парной линейной регрессионной модели по методу наименьших квадратов.

Из этой системы можно найти формулы для нахождения оценок коэффициентов по методу наименьших квадратов. Поделим обе части на N , раскроем скобки и перегруппируем слагаемые, получим

$$\hat{\beta} = \frac{\frac{\sum_{i=1}^N X_i Y_i}{N} - \frac{\sum_{i=1}^N X_i}{N} \frac{\sum_{i=1}^N Y_i}{N}}{\frac{\sum_{i=1}^N X_i^2}{N} - \left(\frac{\sum_{i=1}^N X_i}{N} \right)^2} = \frac{\sum_{i=1}^N X_i Y_i - N\bar{X}\bar{Y}}{\sigma_X^2} = \frac{Cov(X, Y)}{\sigma_X^2}$$

из (2.5) получим, что

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}.$$

Коэффициент наклона линии регрессии можно представить в другом виде:

$$\hat{\beta} = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^N (X_i - \bar{X})^2}$$

Замечание 1. Линия регрессии проходит через точку (\bar{X}, \bar{Y}) .

Замечание 2. Мы предполагаем, что среди X_i есть разные, тогда $\sigma_X \neq 0$. В противном случае, оценок по методу наименьших квадратов не существует.

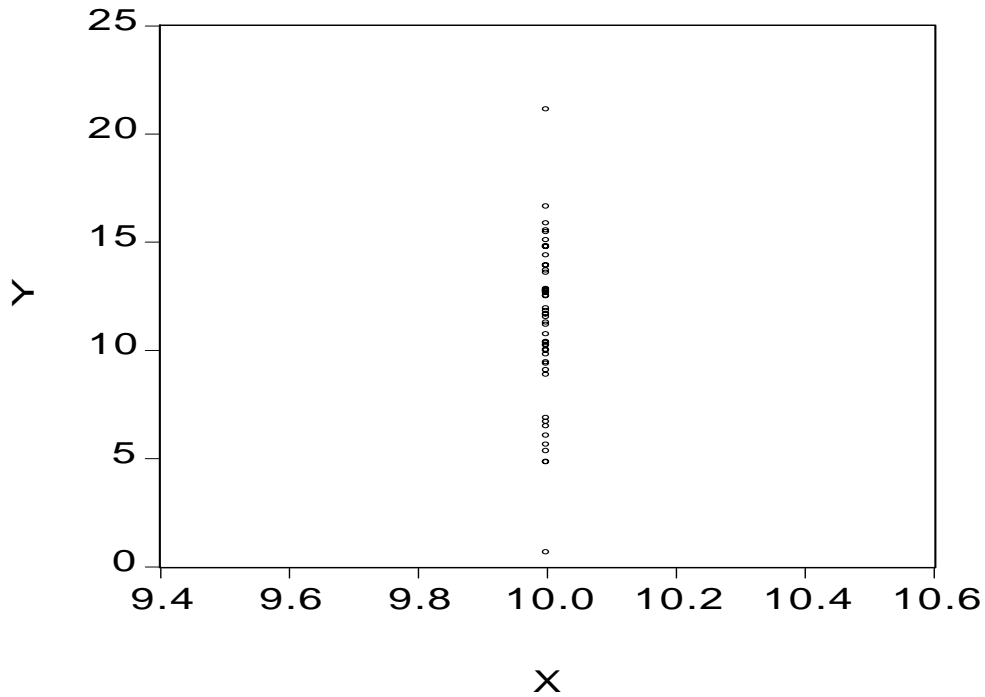


Рис. 11. Регрессионная прямая не существует.

Выборочный коэффициент корреляции и его свойства.

Введем меру близости данных наблюдений к линейной регрессии. В качестве меры такой близости будет служить выборочный коэффициент парной линейной корреляции, который вычисляется по формуле:

$$r_{xy} = \frac{\frac{\sum_{i=1}^N X_i Y_i}{N} - \bar{X}\bar{Y}}{\sqrt{\frac{\sum_{i=1}^N X_i^2}{N} - (\bar{X})^2} \sqrt{\frac{\sum_{i=1}^N Y_i^2}{N} - (\bar{Y})^2}}$$

В этом случае имеют место соотношения

$$r_{xy} = \hat{\beta} \frac{\sigma_x}{\sigma_y} \text{ и } \hat{\beta} = r_{XY} \frac{\sigma_y}{\sigma_x}.$$

Если из уравнения $\hat{Y} = \hat{\alpha} + \hat{\beta}X$ вычесть уравнение $\bar{Y} = \hat{\alpha} + \hat{\beta}\bar{X}$, то

получим $(\hat{Y} - \bar{Y}) = \hat{\beta}(X - \bar{X})$ или $(\hat{Y} - \bar{Y}) = \hat{\beta} \frac{\sigma_x}{\sigma_y} (X - \bar{X})$, тогда

$$\frac{(\hat{Y} - \bar{Y})}{\sigma_y} = r_{xy} \frac{(X - \bar{X})}{\sigma_x}$$

Свойства коэффициента корреляции.

1. $|r_{xy}| \leq 1$

Пусть y_j – наблюдаемое значение Y , $\hat{y}_j = ax_j + b$ – прогнозное или теоретическое значение Y при $X = X_j$. Рассмотрим выражение

$$S_y^2 = \frac{1}{N} \sum_{j=1}^N (y_j - \hat{y}_j)^2.$$

Произведя некоторые преобразования, получим $S_y^2 = \sigma_y^2 (1 - r_{xy}^2)$.

Поскольку как сумма квадратов $S_y^2 \geq 0$ и $\sigma_y^2 \geq 0$, то $(1 - r_{xy}^2) \geq 0$, откуда следует требуемое свойство.

2. Если $r_{xy} = \pm 1$, то это является необходимым и достаточным условием того, что все наблюдаемые значения (x_j, y_j) лежат на прямой регрессии, т. е. по данным наблюдений между переменными X и Y существует функциональная зависимость – без доказательства.

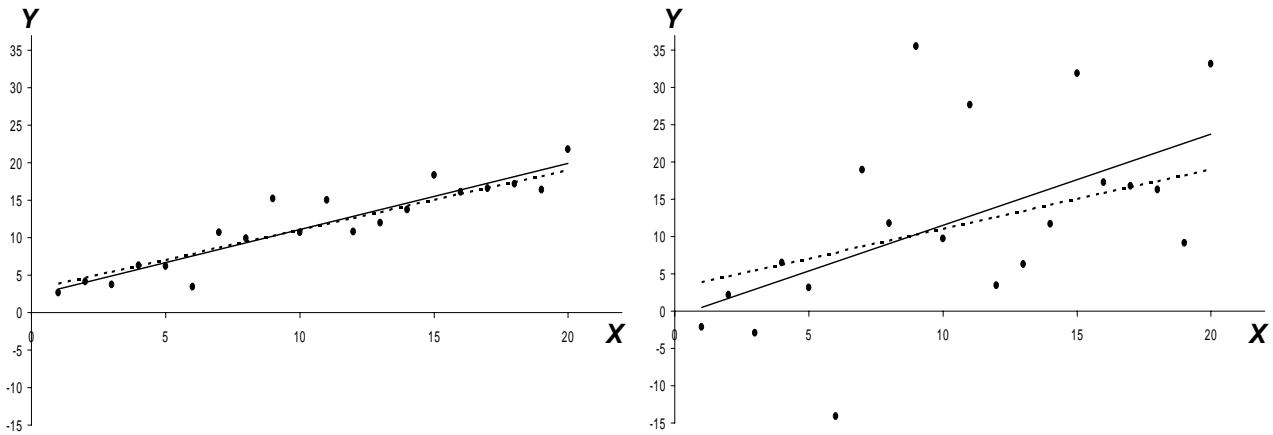
Но, на самом деле, мы можем добавить еще одно наблюдение и картина изменится. Такой вывод мы можем сделать именно на основании имеющихся у нас данных.

3. Пусть $r_{xy} = 0$ то из (1.7) следует при $\sigma_y > 0$, что $y = \bar{y}$

В этом случае говорят, что переменные не связаны линейной корреляционной зависимостью.

В том смысле, что условные средние сохраняют неизменные значения при изменении соответствующих аргументов. Однако, в этом случае признаки могут быть связаны нелинейной корреляционной зависимостью или даже быть функционально зависимыми.

4. Для случая $0 < |r_{xy}| < 1$ говорят, что между переменными существует линейная корреляционная зависимость, которая тем лучше (ближе к линейной функциональной), чем ближе $|r_{xy}|$ к 1, поскольку при $|r_{xy}| \rightarrow 1$ следует, что $S_y^2 \rightarrow 0$.



$$Y = 3.0 + 0.8X$$

Рис 12. Теснота линейной корреляционной связи

Уравнение одно, теснота линейной корреляционной связи разная.

Вопрос о наличии достаточно «хорошей» линейной корреляционной зависимости в каждом конкретном случае решается не только путем вычисления r_{xy} , но и с учетом опыта и интуиции исследователя.

5. Оценка тесноты связи не меняется при нормализации переменных.

Переменная Z называется нормализованной, если $\bar{Z} = 0$, $\sigma_z = 1$.

Пусть заданы переменные X и Y , проведена серия наблюдений и вычислены \bar{x} , \bar{y} , σ_x , σ_y . Сделаем замену переменных:

$$Z = \frac{x - \bar{x}}{\sigma_x}, \quad U = \frac{y - \bar{y}}{\sigma_y}. \quad \text{Нетрудно убедиться, что эти переменные}$$

нормализованы. Тогда

$$r_{xy} = \frac{\sum_{j=1}^N \frac{x_j y_j}{N} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}, \quad r_{UZ} = \frac{\sum_{j=1}^N \frac{Z_j U_j}{N} - \bar{Z} \cdot \bar{U}}{\sigma_Z \cdot \sigma_U} = \sum_{j=1}^N \frac{Z_j U_j}{N}, \quad \text{отсюда}$$

$$r_{UZ} = \sum_{j=1}^N \frac{(x_j - \bar{x})(y_j - \bar{y})}{N\sigma_x\sigma_y} = r_{xy}$$

Если мы получили значение коэффициента корреляции близкое к 1, мы делаем вывод о том, что переменные достаточно сильно связаны между собой.

Однако, если коэффициент корреляции между двумя исследуемыми переменными близок к 1, на самом деле они могут и не быть зависимыми. Пример с душевнобольными и радиоприемниками – пример так называемой «ложной корреляции». Высокое значение коэффициента корреляции может быть обусловлено и существованием третьей переменной, которая оказывает сильное влияние на первые две переменные, что и служит причиной их высокой коррелируемости. Поэтому возникает задача расчета «чистой» корреляции между переменными X и Y , т. е. корреляции, в которой исключено влияние (линейное) других переменных. Для этого и вводят понятие коэффициента частной корреляции.

Итак, мы хотим определить коэффициент частной корреляции между переменными X и Y , исключив линейное влияние переменной Z . Для его определения используется следующая процедура:

1. Оцениваем регрессию $Y = \alpha_1 + \alpha_2 Z + \varepsilon$,
2. Получаем остатки $e_i^Y = Y_i - \hat{Y}_i$,
3. Оцениваем регрессию $X = \alpha_1 + \alpha_2 Z + \varepsilon$,
4. Получаем остатки $e_i^X = X_i - \hat{X}_i$,
5. $r(XY | Z) = r_{e^X e^Y}$ - выборочный коэффициент частной корреляции,

измеряет степень связи между переменными X и Y , очищенную от влияния переменной Z .

Прямые вычисления:

$$r(XY | Z) = \frac{r_{YX} - r_{XZ}r_{YZ}}{\sqrt{1 - r_{XZ}^2} \sqrt{1 - r_{YZ}^2}}$$

Свойство: $0 \leq r(XY | Z) \leq 1$

Процедура построения коэффициента частной корреляции обобщается на случай, если мы хотим избавиться от влияния двух и более переменных.

3. МНОЖЕСТВЕННАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ

Множественный регрессионный анализ является расширением парного регрессионного анализа на случай, когда зависимая переменная гипотетически связана с более чем одной независимой переменной. В этом случае возникает новая проблема, которой не было в случае парной модели. При оценке влияния данной независимой переменной на зависимую переменную нам надо будет разграничить воздействие на зависимую переменную ее и другие переменные. Кроме того, мы должны будем решить проблему спецификации модели. Если в парном регрессионном анализе эта проблема заключалась только в выборе вида функции $f(X)$, то теперь нам, кроме этого, надо будет решить, какие мы будем включать в модель, а какие – нет. Иначе говоря, если предполагается, что несколько переменных могут оказывать влияние на зависимую переменную, то другие могут и не подходить для нашей модели.

Итак, у нас есть независимая переменная Y , которая характеризует состояние или поведение экономического объекта, и есть набор переменных X_1, \dots, X_k , характеризующие этот экономический объект качественно или количественно, которые, как мы предполагаем, оказывают влияние на переменную Y , т. е. мы предполагаем, что значения результирующей переменной Y выступают в виде функции, значения которой определяются, правда, с некоторой погрешностью, значениями объясняющих переменных, выступающих в роли аргументов этой функции, т. е.

$$Y = f(X_1, \dots, X_k) + \varepsilon,$$

где ε - случайный член, который входит в наше уравнение по тем же самым причинам, что и в случае парного регрессионного анализа.

Поначалу, среди всех возможных функций $f(X_1, \dots, X_k)$ мы выбираем линейные:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon \quad (*)$$

(*) – множественная линейная регрессионная модель (МЛРМ) со свободным членом.

$$Y = \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon - \text{МЛРМ без свободного члена.}$$

Например, если мы изучаем величину спроса на масло, то модель может выглядеть следующим образом:

$$Q^D = \beta_0 + \beta_1 P + \beta_2 X + \beta_3 PM + \varepsilon,$$

где Q^D – объем спроса на масло, X – доход, P – цена на масло, PM – цена на мягкое.

Здесь нам неизвестны коэффициенты β и параметры распределения ε , Зато мы имеем выборку из N наблюдений над переменными Y и X_1, \dots, X_k . Для каждого наблюдения должно выполняться следующее равенство:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \varepsilon_i$$

или в матричной форме:

$$Y = X\beta + \varepsilon, \text{ где}$$

$$Y = \begin{bmatrix} Y_1 \\ \dots \\ Y_N \end{bmatrix}, X = \begin{bmatrix} \mathbf{1} & X_{11} & \dots & X_{k1} \\ \dots & \dots & \dots & \dots \\ \mathbf{1} & X_{N1} & \dots & X_{kN} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \dots \\ \beta_k \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_N \end{bmatrix}.$$

Наша задача по результатам наблюдений, на основе этих наблюдений, получить надежные оценки неизвестных коэффициентов (оценить неизвестные параметры) и проверить, насколько хорошо выбранная модель соответствует исходным данным.

Каким образом получить эти оценки? Нам надо построить гиперплоскость. Из всех возможных гиперплоскостей мы хотим выбрать ту, чтобы она «наилучшим образом» подходила к нашим данным, *была бы в центре скопления наших данных*, т. е. чтобы все Y_i как можно ближе лежали к нашей гиперплоскости. В качестве меры близости точек к прямой мы введем разность

$$e_i = Y_i - \hat{Y}_i = Y_i - b_1 X_{1i} - \dots - b_k X_{ki}$$

Очевидно, значения b_1, \dots, b_k надо подбирать таким образом, чтобы минимизировать некоторую интегральную (т. е. по всем имеющимся наблюдениям) характеристику невязок или остатков:

$$S = \sum_{i=1}^N e_i^2 = \sum_{i=1}^N (Y_i - b_1 X_{1i} - \dots - b_k X_{ki})^2 = F(b_1, \dots, b_k) \rightarrow \min_{(b_1, \dots, b_k)}$$

Необходимое условие экстремума:

$$\left\{ \begin{array}{l} \frac{\partial S}{\partial b_0} = 0 \\ \frac{\partial S}{\partial b_1} = 0 \\ \frac{\partial S}{\partial b_2} = 0 \\ \dots \\ \frac{\partial S}{\partial b_k} = 0 \end{array} \right.$$

или

$$\left\{ \begin{array}{l} -2 \sum_{i=1}^N (Y_i - b_1 X_{1i} - \dots - b_k X_{ki}) = 0 \\ -2 \sum_{i=1}^N (Y_i - b_1 X_{1i} - \dots - b_k X_{ki}) X_{1i} = 0 \\ -2 \sum_{i=1}^N (Y_i - b_1 X_{1i} - \dots - b_k X_{ki}) X_{2i} = 0 \\ \dots \\ -2 \sum_{i=1}^N (Y_i - b_1 X_{1i} - \dots - b_k X_{ki}) X_{ki} = 0 \end{array} \right.$$

После преобразований получим:

$$\left\{ \begin{array}{l} Nb_0 + b_1 \sum_{i=1}^N X_{1i} + b_2 \sum_{i=1}^N X_{2i} + \dots + b_k \sum_{i=1}^N X_{ki} = \sum_{i=1}^N Y_{1i} \\ b_0 \sum_{i=1}^N X_{1i} + b_1 \sum_{i=1}^N X_{1i}^2 + b_2 \sum_{i=1}^N X_{1i} X_{2i} + \dots + b_k \sum_{i=1}^N X_{1i} X_{ki} = \sum_{i=1}^N Y_{1i} X_{1i} \\ b_0 \sum_{i=1}^N X_{2i} + b_1 \sum_{i=1}^N X_{1i} X_{2i} + b_2 \sum_{i=1}^N X_{2i}^2 + \dots + b_k \sum_{i=1}^N X_{2i} X_{ki} = \sum_{i=1}^N Y_{1i} X_{2i} \\ \dots \\ b_0 \sum_{i=1}^N X_{ki} + b_1 \sum_{i=1}^N X_{1i} X_{ki} + b_2 \sum_{i=1}^N X_{2i} X_{ki} + \dots + b_k \sum_{i=1}^N X_{ki}^2 = \sum_{i=1}^N Y_{1i} X_{ki} \end{array} \right.$$

Эта система называется система нормальных уравнений для нахождения коэффициентов множественной линейной регрессионной модели по методу наименьших квадратов.

Получим формулу для нахождения коэффициентов множественной линейной регрессионной модели в матричном виде.

$$\sum_{i=1}^N e_i^2 = e'e, \quad e = \begin{bmatrix} e_1 \\ \dots \\ e_N \end{bmatrix}, \quad \text{тогда } e = Y - X\hat{\beta}.$$

$$\begin{aligned} e'e &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) = Y'Y - Y'X\hat{\beta} - \hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta} = \\ &= Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta} \end{aligned}$$

Здесь мы воспользовались тем, что $\hat{\beta}'X'Y$ - скаляр, и поэтому он совпадает со своим транспонированным значением.

Необходимое условие минимума (в матричной форме):

$$\frac{\partial(e'e)}{\partial\hat{\beta}} = -2X'Y + 2X'X\hat{\beta}.$$

Здесь мы воспользовались свойствами векторного и матричного дифференцирования:

Что значит продифференцировать вектор-функцию по вектору переменных:

$$\frac{\partial \varphi(x)}{\partial x} = \begin{pmatrix} \frac{\partial \varphi_1(x)}{\partial x_1} & \dots & \frac{\partial \varphi_1(x)}{\partial x_n} \\ \frac{\partial \varphi_m(x)}{\partial x_1} & \dots & \frac{\partial \varphi_m(x)}{\partial x_n} \end{pmatrix}$$

Здесь $\varphi(x)$ – m -мерная вектор-функция, x – n -мерный вектор.

Случаи:

1) $\varphi(x) = a'x$, $a = (a_1, \dots, a_n)$, $x = (x_1, \dots, x_n)$

$$\frac{\partial \varphi(x)}{\partial x} = a'$$

2) $\varphi(x) = x'Ax$, $A: n \times n$ - матрица

$$\frac{\partial \varphi(x)}{\partial x} = x'(A + A')$$
, если матрица A симметричная, то $\frac{\partial \varphi(x)}{\partial x} = 2x'A$

3) $\varphi(x) = Ax$, $A: m \times n$ - матрица.

$$\frac{\partial \varphi(x)}{\partial x} = A$$

Итак, $\frac{\partial ESS}{\partial \hat{\beta}} = -2X'Y + 2X'X\hat{\beta}$

$$2X'Y = 2X'X\hat{\beta}$$

$X'Y = X'X\hat{\beta}$, если матрица $(X'X)$ невырождена, то

$$\boxed{\hat{\beta} = (X'X)^{-1} X'Y}$$
 - МНК оценки коэффициентов МЛРМ.

Коэффициенты по методу наименьших квадратов существуют не всегда, а только в том случае, когда определитель матрицы $(X'X)$ отличен от нуля. Определитель будет равен нулю в случае, если столбцы матрицы X линейно зависимы. Такое может произойти, если между независимыми переменными существует точное линейное соотношение. Если такое соотношение между переменными существует, мы говорим о том, что в модели присутствует полная мультиколлинеарность.

Пример. Рассмотрим модель со средней оценкой на экзамене, состоящую из трех объясняющих переменных: I – доход родителей, D – среднее число часов, затраченных на обучение в день, W – среднее число часов, затраченных на обучение в неделю. Очевидно, что $W=7D$. И это соотношение будет выполняться для каждого студента, который попадет в нашу выборку. Случай полной мультиколлинеарности отследить легко, поскольку в этом случае невозможно построить оценки по методу наименьших квадратов. Если в модели присутствует полная мультиколлинеарность, следует удалить из регрессионного уравнения одну из переменных, которые входят в линейное соотношение.

Наряду с коэффициентами исходного регрессионного уравнения рассматривают еще нормализованные коэффициенты.

Нормализуем исходные переменные, для чего вычислим

$$\bar{X}_j = \frac{\sum_{j=1}^N X_{ij}}{N} \quad (j = 1, \dots, k); \quad \bar{Y} = \frac{\sum_{j=1}^N Y_j}{N};$$

$$d_{x_i} = \sqrt{\frac{\sum_{j=1}^n X_{ij}^2}{N} - \bar{X}_i^2} \quad (i=1, \dots, k); \quad d_y = \sqrt{\frac{\sum_{j=1}^N Y_j^2}{N} - \bar{Y}^2}.$$

Введем новые переменные

$$Z_i = \frac{X_i - \bar{X}_i}{d_{x_i}} \quad (i=1, \dots, k); \quad U = \frac{Y - \bar{Y}}{d_y}$$

Очевидно, что $d_{z_1} = d_{z_2} = \dots = d_{z_n} = d_u = 1$, т. е.

новые переменные нормализованы. Вместо векторов наблюдений $(X_{1j}, \dots, X_{nj}, Y_j)$ будем рассматривать N векторов $(Z_{1j}, \dots, Z_{nj}, Y_j)$, которые получены путем использования предыдущих формул.

Будем оценивать коэффициенты следующей регрессионной функции $U = \alpha_1 Z_1 + \dots + \alpha_n Z_n + \alpha_0$, которые называются нормализованными коэффициентами регрессии. Если для их отыскания использовать метод

наименьших квадратов, то, очевидно, что $\alpha_0=0$. Поэтому линейная функция с нормализованными коэффициентами регрессии имеет вид

$$U = \alpha_1 Z_1 + \dots + \alpha_n Z_n$$

Построим функцию $\Phi(\hat{\alpha}_1, \dots, \hat{\alpha}_n) = \sum_{j=1}^n (\hat{\alpha}_1 Z_{1j} + \dots + \hat{\alpha}_n Z_{nj} - U_j)^2$ и будем

искать ее наименьшее значение по параметрам $\{\alpha_i\}$:

$$\min_{\{\alpha_1, \dots, \alpha_n\}} \left\{ \Phi(\hat{\alpha}_1, \dots, \hat{\alpha}_n) = \sum_{j=1}^n (\hat{\alpha}_1 Z_{1j} + \dots + \hat{\alpha}_n Z_{nj} - U_j)^2 \right\}$$

Продифференцировав эту функцию по всем параметрам $\hat{\alpha}_j$ ($j=1, \dots, k$) и приравняв эти производные к нулю, получим k уравнений. Например, для

частной производной по $\hat{\alpha}_j$ будет $\frac{\partial \Phi}{\partial \hat{\alpha}_j} = 2 \sum_{i=1}^N (\hat{\alpha}_1 Z_{1j} + \dots + \hat{\alpha}_n Z_{nj} - U_j) Z_{ji} = 0$.

Разделим обе части на N и раскроем скобки. Получим

$$\hat{\alpha}_1 \frac{\sum_{j=1}^N Z_{1j} Z_{ij}}{N} + \dots + \hat{\alpha}_n \frac{\sum_{j=1}^N Z_{nj} Z_{ij}}{N} = \frac{\sum_{j=1}^N U_j Z_{ij}}{N}.$$

Воспользовавшись предыдущими соотношениями, получим

$$\frac{\sum_{j=1}^N Z_{kj} Z_{ij}}{N} = r_{z_k z_i} = r_{x_k x_i} \quad (l = 1, \dots, k);$$

$$\frac{\sum_{j=1}^N U_j Z_{ij}}{N} = r_{z_u z_i} = r_{y x_i} \quad (j = 1, \dots, k);$$

$$\frac{\sum_{j=1}^N Z_{ij} U_j}{N} = r_{z_i z_i} = r_{x_i x_i} = 1 \quad (j = 1, \dots, k).$$

Введем следующие обозначения:

$r_{x_l x_k} = r_{lk}$ ($j=1, \dots, k, l=1, \dots, k$); $r_{y x_j} = r_{0j}$ ($j=1, \dots, k$). Подставив их в выражение

получим систему, состоящую из k линейных уравнений с k неизвестными $\hat{\alpha}_1, \dots, \hat{\alpha}_k$:

$$\begin{cases} r_{11}\beta_1 + r_{12}\beta_2 + \dots + r_{1n}\beta_n = r_{01} \\ r_{21}\beta_1 + r_{22}\beta_2 + \dots + r_{2n}\beta_n = r_{02} \\ \dots \\ r_{n1}\beta_1 + r_{n2}\beta_2 + \dots + r_{nn}\beta_n = r_{0n} \end{cases}$$

Матрица этой системы называется корреляционной матрицей R

$$R = \begin{pmatrix} 1 & r_{12} & r_{13} & \dots & r_{1n} \\ r_{21} & 1 & r_{23} & \dots & r_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ r_{n1} & r_{n2} & r_{n3} & \dots & r_{nn} \end{pmatrix}$$

По главной диагонали стоят 1, поскольку $r_{ii}=1$. Матрица является симметричной, поскольку $r_{ik} = r_{ki}$. Кроме того, при внимательном отношении к сбору данных она является невырожденной, т. е. решение системы (2.7) всегда существует.

Решив систему, получим значение нормализованных коэффициентов регрессии β_1, \dots, β_n . для получения оценок коэффициентов исходного уравнения подставим (2.4) в уравнение (2.5)

$$\frac{y - \bar{y}}{\sigma_y} = \beta_1 \frac{x_1 - \bar{x}_1}{\sigma_{x_1}} + \dots + \beta_n \frac{x_n - \bar{x}_n}{\sigma_{x_n}}$$

Преобразовав выражение, получаем:

$$y = \beta_1 \frac{\sigma_y}{\sigma_{x_1}} x_1 + \beta_2 \frac{\sigma_y}{\sigma_{x_2}} x_2 + \dots + \beta_n \frac{\sigma_y}{\sigma_{x_n}} x_n + (\bar{y} - \beta_1 \frac{\sigma_y}{\sigma_{x_1}} \bar{x}_1 - \beta_2 \frac{\sigma_y}{\sigma_{x_2}} \bar{x}_2 - \dots - \beta_n \frac{\sigma_y}{\sigma_{x_n}} \bar{x}_n)$$

Сравнивая выражение с исходным, получаем формулы для вычисления коэффициентов регрессии a_0, a_1, \dots, a_n :

$$\left\{ \begin{array}{l} a_1 = \beta_1 \frac{\sigma_y}{\sigma_{x_1}}, \\ a_2 = \beta_2 \frac{\sigma_y}{\sigma_{x_2}}, \\ \dots \\ a_n = \beta_n \frac{\sigma_y}{\sigma_{x_n}}, \\ a_0 = \bar{y} - a_1 \bar{x}_1 - \dots - a_n \bar{x}_n \end{array} \right.$$

Таким образом, зная нормализованные коэффициенты, можно по этим формулам, найти исходные коэффициенты регрессии.

4. ОЦЕНКА КАЧЕСТВА ПОДГОНКИ ЛИНИИ РЕГРЕССИИ К ИМЕЮЩИМСЯ ДАННЫМ

Итак, гиперплоскость мы построили. Насколько хорошо нам удалось объяснить изменение переменной Y нашей моделью. Разложим вариацию Y на две части. Насколько наше уравнение объясняет вариацию Y и какова часть Y , которую мы не можем объяснить нашим уравнением.

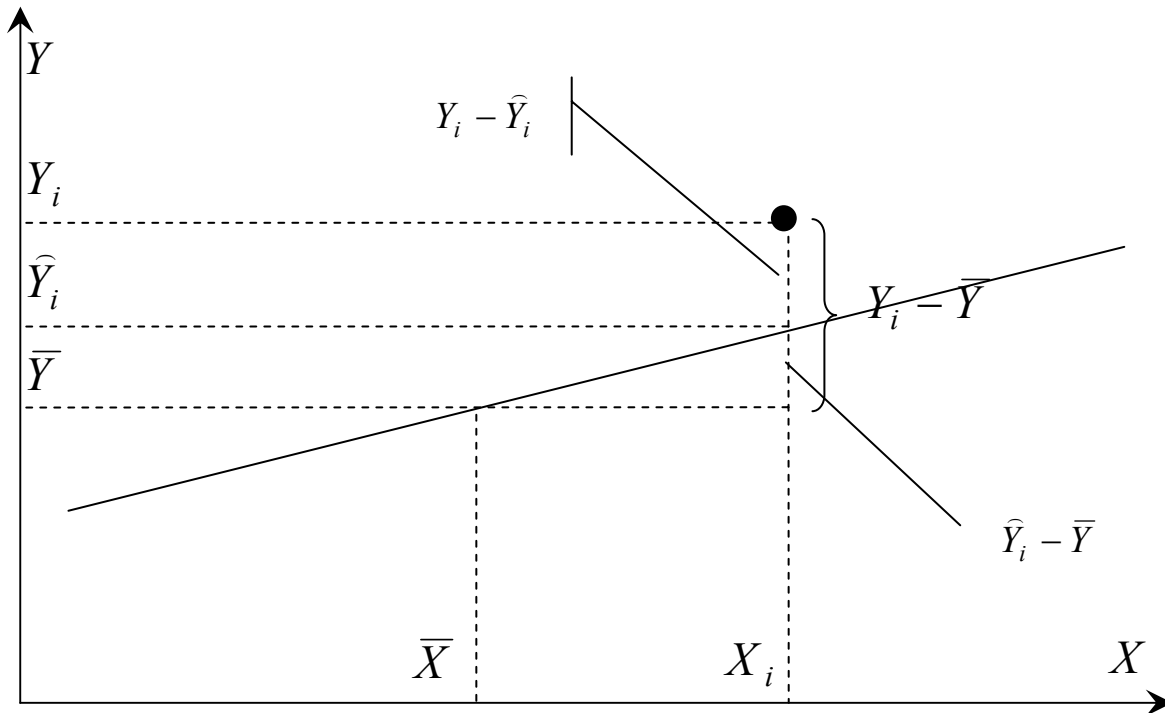


Рис 1. Разложение отклонения переменной Y от среднего на две составляющие.

Рассмотрим $\sum_{i=1}^N (Y_i - \bar{Y})^2$ - величина, являющаяся мерой вариации

переменной Y вокруг ее среднего значения. Распишем эту величину:

$$\sum_{i=1}^N (Y_i - \bar{Y})^2 = \sum_{i=1}^N (Y_i - \hat{Y}_i + \hat{Y}_i - \bar{Y})^2 = \sum_{i=1}^N (Y_i - \hat{Y}_i)^2 - 2 \sum_{i=1}^N (Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) + \sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2$$

I
II
III

В этой сумме $II = 0$, если в уравнении есть свободный член.

$$\sum_{i=1}^N (Y_i - \bar{Y})^2 = \sum_{i=1}^N (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2$$

TSS ESS RSS

где

TSS – total sum of squares – вся дисперсия или вариация Y , характеризует степень случайного разброса значений функции регрессии около среднего значения Y ;

ESS – error sum of squares – есть сумма квадратов остатков регрессии, та величина, которую мы минимизируем при построении прямой, часть дисперсии, которая нашим уравнением не объясняется;

RSS – regression sum of squares – объясненная часть дисперсии.

Определение. Коэффициентом детерминации или долей объясненной нашим уравнением дисперсии называется величина

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS}$$

Свойства коэффициента детерминации:

1. $0 \leq R^2 \leq 1$ в силу определения;

2. $R^2 = 0$ - в это м случае $RSS = 0$, т. е. наша регрессия ничего не объясняет, ничего не дает по сравнению с тривиальным прогнозом $\hat{Y}_i = \bar{Y}$. Наши данные позволяют сделать вывод о независимости Y и X , изменение в переменной X никак не влияет на изменение среднего значения переменной Y (примеры, когда зависимость между переменными есть, а коэффициент детерминации равен нулю);

3. $R^2 = 1$ - в этом случае все точки (X_i, Y_i) лежат на одной прямой ($ESS = 0$). Тогда на основании наших данных можно сделать вывод о наличии функциональной, а именно, линейной, зависимости между переменными Y и X . Изменение переменной Y полностью объясняется изменением переменной X ;

4. $0 < R^2 < 1$ - в этом случае чем ближе R^2 к 1, тем лучше качество подгонки кривой к нашим данным, тем точнее \hat{Y} аппроксимирует Y .

5. R^2 , вообще говоря, возрастает при добавлении еще одного регрессора, поэтому для выбора между несколькими регрессионными уравнениями не следует полагаться только на R^2

Попыткой устранить эффект, связанный с ростом R^2 при увеличении числа регрессоров, является коррекция R^2 на число регрессоров - наложение "штрафа" за увеличение числа независимых переменных. Скорректированный

$$R^2 - R_{adj}^2 : R_{adj}^2 = 1 - \frac{ESS / (N - k)}{TSS / (N - 1)} \quad (3.9)$$

здесь в числителе - несмещенная оценка дисперсии ошибок (как увидим позднее), в знаменателе - несмещенная оценка дисперсии Y . (*Совпадают ли они?*).

Свойства:

1. $R_{adj}^2 = 1 - (1 - R^2) \frac{N - 1}{N - k}$ - доказать самим;

2. $R^2 > R_{adj}^2$, $k > 1$:

$$R^2 - R_{adj}^2 = R^2 - 1 + (1 - R^2) \frac{N - 1}{N - k} = (1 - R^2) + \left(\frac{N - 1}{N - k} - 1 \right) > 0, k > 1;$$

3. $R_{adj}^2 \leq 1$, но может быть и < 0 .

В определенном смысле использование R_{adj}^2 для сравнения регрессий при изменении числа регрессоров более корректно.

Упражнение. Показать, что статистика R_{adj}^2 увеличится при добавлении новой переменной тогда и только тогда, когда t -статистика коэффициента при этой переменной по модулю больше 1.

Следовательно, если в результате регрессии с новой переменной R_{adj}^2 увеличилась, это еще не означает, что коэффициент при этой переменной значимо отличается от нуля, поэтому мы не можем сказать, что спецификация модели улучшилась. Это первая причина, почему R_{adj}^2 не стал широко использоваться в качестве диагностической величины. Вторая причина -

уменьшение внимания к самому R^2 . На практике даже плохо определенная модель регрессии может давать высокий коэффициент R^2 . Поэтому теперь он рассматривается в качестве одного из целого ряда диагностических показателей, которые должны быть проверены при построении модели регрессии. Следовательно, и корректировка его мало что дает.

Итак, при помощи регрессионного анализа мы с вами получили оценки интересующей нас зависимости (*): $\hat{\beta} = (X'X)^{-1}X'Y$

Однако, это всего лишь оценки. Возникает вопрос, насколько они хороши. Оказывается, что при выполнении некоторых условий наши оценки получаются достаточно надежными.

5. НЕЛИНЕЙНЫЕ РЕГРЕССИОННЫЕ МОДЕЛИ.

Выделяют два типа нелинейных моделей:

- Нелинейные зависимости, приводящиеся преобразованием переменных к линейным
- Нелинейные зависимости, не приводящиеся преобразованием переменных к линейным

Наиболее универсальным инструментом анализа является метод наименьших квадратов, который предназначен для оценки линейной модели. Поэтому при анализе нелинейной регрессионной модели основным приемом является сведение задачи линейной модели и последующая оценка ее методом наименьших квадратов. Сведение это является искусственным и зависит от опыта и интуиции исследователя.

Парная нелинейная регрессионная модель. Пусть заданы зависимая переменная Y – случайная величина, X – независимая переменная. За парой переменных (X, Y) проведена серия из N наблюдений. Данные сгруппированы и представлены в виде следующей таблицы:

X_i	зависимая переменная	n_i	\bar{Y}_{X_i}	D_i
X_1	$Y_{11} \dots Y_{1j} \dots Y_{1n_1}$	n_1	\bar{y}_{x_1}	D_1
		...		
X_i	$Y_{i1} \dots Y_{ij} \dots Y_{in_i}$	n_i	\bar{y}_{x_i}	D_i
		...		
X_m	$Y_{m1} \dots Y_{mj} \dots Y_{mn_m}$	n_m	\bar{y}_{x_m}	D_m

Здесь среди X_i нет одинаковых, n_i – число наблюдений при $X = X_i$,

$\sum_{i=1}^m n_i = N$, \bar{y}_{x_i} – условные средние и D_i – групповые дисперсии, которые

характеризует рассеяние наблюдений внутри i – й группы:

$$\bar{y}_{x_i} = \frac{\sum_{j=1}^N y_{ij}}{n_i}, D_i = \frac{\sum_{j=1}^N (y_{ij} - \bar{y}_i)^2}{n}.$$

В этом случае можно вычислить:

$$D_{\text{общ}} = \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2}{N}, \text{ где } \bar{y} = \frac{\sum_{i=1}^m n_i \bar{y}_i}{N};$$

$$D_{\text{МГ}} = \frac{\sum_{i=1}^m n_i (\bar{y}_i - \bar{y})^2}{N}, D_{\text{ВГ}} = \frac{\sum_{i=1}^m n_i \bar{y}_i}{N};$$

$$D_{\text{общ}} = D_{\text{МГ}} + D_{\text{ВГ}}.$$

$D_{\text{общ}}$ – общая дисперсия, $D_{\text{ВГ}}$ – внутригрупповая дисперсия, $D_{\text{МГ}}$ – межгрупповая дисперсия.

Введем в рассмотрение показатель, который обычно называется теоретическим корреляционным отношением:

$$R_0 = \sqrt{\frac{D_{\text{МГ}}}{D_{\text{общ}}}} = \sqrt{\frac{D_{\text{общ}} - D_{\text{ВГ}}}{D_{\text{общ}}}}$$

Необходимо подчеркнуть, что R_0 вычисляется только по данным наблюдений и не зависит от вида корреляционной зависимости.

Это теоретическое корреляционное отношение обладает следующими основными свойствами:

1. $0 \leq R_0 \leq 1$. Это очевидное свойство следует из соотношения $D_{\text{общ}} = D_{\text{МГ}} + D_{\text{ВГ}}$, так как $D_{\text{общ}} \geq D_{\text{МГ}}$.
2. Если $R_0 = 0$, то $D_{\text{МГ}} = 0$ и отсюда следует, что $\bar{y}_{x_i} = \bar{y}$ для любого $i=1, \dots, m$. В этом случае говорят, что между Y и X не существует никакой корреляционной зависимости.
3. $R_0 \geq |r_{XY}|$, причем равенство достигается только в случае, когда между переменными Y и X существует линейная функциональная зависимость.

4. Если $R_0 = 0$, то $D_{\text{ВГ}} = 0$, т. е. все $D_i = 0$. Это возможно только тогда, когда каждому X_i соответствует одно Y_i . В этом случае говорят, что можно построить такую корреляционную функцию $Y = f(X)$, что она пройдет через все точки (X_i, Y_i) , $i = 1, \dots, m$.
5. Если $0 < R_0 < 1$, то говорят, что между переменными Y и X существует корреляционная зависимость и можно подобрать такую корреляционную функцию $Y = f(X)$, что кривая пройдет через все точки (X_i, \bar{y}_{x_i}) , $i=1, \dots, m$, и оценка степени приближения данных совпадет с R_0 .

Вид такой функции определяется достаточно просто с учетом возможности сведения расчетов к множественной линейной модели.

$$\bar{Y} = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_k X^k$$

т. е. нам надо подобрать такой полином степени k , который пройдет через все точки (X_i, \bar{Y}_{x_i}) . Сведение осуществляется путем замены переменных $Z_k = X^k$. В этом случае получается линейное уравнение

$$\bar{Y} = \beta_0 + \beta_1 Z_1 + \dots + \beta_k Z_k$$

Методом наименьших квадратов получаем оценки $\hat{\beta}$ коэффициентов уравнения.

Однако такой подход к определению вида зависимости $Y = f(X)$ применять нецелесообразно по следующим причинам:

1. При больших k возникают вычислительные трудности.
2. Переменные Z_i и Z_j достаточно хорошо коррелируют, что приводит к неустойчивости получаемых результатов.
3. Самое важное, что при построении зависимости случайным возмущениям придается характер закономерности.

Поэтому на практике поступают следующим образом. Рассматриваются несколько возможных видов функций $f(X)$. Некоторые из них приведены в таблице 3.2., где так же указаны способы их сведения к линейному случаю.

Проводится серия наблюдений проверяются гипотезы о виде функции $Y = f(X)$. Выбирается тот вид функции, где R^2 принимает наибольшее значение. При равной объясняющей способности из двух моделей мы всегда выбираем более простую.

Название функции	Аналитическое выражение	Преобразование функции
степенная	$y = ax^b$	$\ln y = \ln a + b \ln x$
показательная	$y = ab^x$	$\ln y = \ln a + x \ln b$
показательно-степенная	$y = ax^b c^x$	$\ln y = \ln a + b \ln x + x \ln c$
экологическая	$y = ae^{-b^2(x-c)^2}$	$\ln y = \ln a - b^2 c^2 + 2b^2 cx - b^2 x^2$
функция Гомперца	$\ln y = \ln a + bc^x$	$\ln \ln y = \ln \ln a + \ln b + x \ln c$
гиперболическая	$y = \frac{1}{a + bx}$	$\frac{1}{y} = a + bx$
дробно-рациональная	$y = \frac{x}{a + bx + cx^2}$	$\frac{x}{y} = a + bx + cx^2$
модифицированная экспоненциальная	$y = ae^{bx}$	$\ln y = \ln a + bx$
функция Торн-Квиста	$y = \frac{ax}{b + x}$	$\frac{1}{y} = \frac{b}{ax} + \frac{1}{a}$

6. СТАТИСТИЧЕСКИЕ СВОЙСТВА ОЦЕНОК КОЭФФИЦИЕНТОВ МЛРМ.

Полученные оценки неизвестных коэффициентов регрессионного уравнения $\hat{\beta}$ мы с вами можем рассматривать как случайные величины. Действительно, при повторении наблюдений над экономическим объектом – получении выборок того же самого объема N при тех же самых значениях объясняющей переменной X значение результирующего параметра Y будет варьироваться за счет случайного члена ε , а, следовательно, будут варьироваться зависящие от y_1, \dots, y_N значения оценок. Если же X – случайная величина, то тогда вариация оценок будет зависеть и от вариации X . Таким образом, свойства коэффициентов регрессии будут существенным образом зависеть от свойств случайного члена ε и от свойств X , если X – случайная величина.

Для того чтобы оценки, полученные по МНК, давали «наилучшие» результаты, мы потребуем от остаточного члена или ошибки ε и от X выполнения следующих условий (предположения относительно того, как генерируются наблюдения):

1. $Y = \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$ - спецификация модели;

2. X_1, \dots, X_k – детерминированные вектора, линейно независимые в R^n , т. е. матрица X имеет максимальный ранг k (в повторяющихся наблюдениях единственным источником случайных возмущений вектора Y являются случайные возмущения вектора ε);

3. $M\varepsilon_i = 0$;

4. $M\varepsilon_i^2 = D\varepsilon_i = \sigma_\varepsilon^2$, дисперсия ошибки не зависит от номера наблюдения;

5. $M(\varepsilon_i \varepsilon_j) = 0$ при $i \neq j$, т. е. некоррелированность ошибок разных наблюдений;

6. $\varepsilon_i \in N(0, \sigma_\varepsilon^2)$, т. е. ε_i – нормально распределенная случайная величина со средним 0 и дисперсией σ_ε^2 .

1-5 - КЛРМ, 1-6 - НЛРМ, условия 1-5 - условия Гаусса-Маркова В матричной форме:

$$M\varepsilon = 0,$$

$$\Omega = \begin{pmatrix} M(\varepsilon_1^2) & M(\varepsilon_1\varepsilon_2) & \dots & M(\varepsilon_1\varepsilon_N) \\ M(\varepsilon_2\varepsilon_1) & M(\varepsilon_2^2) & \dots & M(\varepsilon_2\varepsilon_N) \\ \dots & \dots & \dots & \dots \\ M(\varepsilon_N\varepsilon_1) & M(\varepsilon_N\varepsilon_2) & \dots & M(\varepsilon_N^2) \end{pmatrix} = M(\varepsilon\varepsilon') . \text{ – матрица ковариаций}$$

вектора ошибок. Матрица Ω предполагается положительно определенной, т. е. $\forall x \quad x'\Omega x > 0$.

В классической регрессионной модели матрица ковариаций имеет следующий вид:

$$\Omega = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix}$$

$$D\varepsilon = M(\varepsilon\varepsilon^T) = \sigma^2 I_N \text{ - матрица ковариаций вектора } \varepsilon;$$

$\varepsilon \sim N(0, \sigma^2 I_N)$, т. е. ε_i имеют совместное нормальное распределение со средним 0 и матрицей ковариаций $\sigma^2 I_N$ (про матрицу ковариаций)

.В случае НЛРМ условие 5. эквивалентно условию статистической независимости ошибок для разных наблюдений. Действительно, если две нормально распределенные величины не коррелированы, то они независимы.

Обсудим эти условия.

1. Спецификация модели отражает наше представление о механизме зависимости Y и X и выбор объясняющей переменной X .

2. Мы будем предполагать, что X_i – детерминированные константы, т. е. значения X_i (значение объясняющей переменной в каждом наблюдении) считается экзогенным, полностью определяемым внешними причинами. Такое

предположение подразумевает то, что переменная X полностью контролируется исследователем, который может изменять ее значение в целях эксперимента. Это предположение нереалистично во многих экономических и бизнес моделях. Позже мы посмотрим, сохраняются ли свойства оценок в случае, если X – случайная величина.

3. В матричной форме это условие выглядит так: $M\varepsilon = 0$.

Это условие состоит в том, что математическое ожидание случайного члена равно нулю в любом наблюдении. Иногда случайный член бывает положительным, иногда отрицательным, но он не должен иметь смещения ни в одном возможном направлении.

Надо сказать, что если в уравнение включается постоянный член, то бывает разумным предположить, что первое условие выполняется автоматически, т. к. роль константы и состоит в определении любой систематической составляющей в Y , которую не учитывают объясняющие переменные (если спецификация модели выбрана правильно).

Иллюстрация: предположим, что $M\varepsilon_i = \mu$, тогда

$$\begin{aligned} Y_i &= \alpha + \beta X_i + \varepsilon_i = \alpha + \beta X_i + \varepsilon_i + \mu - \mu = (\alpha + \mu) + \beta X_i + (\varepsilon_i - \mu) = \\ &= \alpha' + \beta X_i + \varepsilon'_i \end{aligned}$$

$$M(\varepsilon_i - \mu) = \mu - \mu = 0$$

Таким образом, исходная модель эквивалентна новой модели с ошибкой, имеющей нулевое математическое ожидание и другим свободным членом.

4. Второе условие говорит нам о том, что дисперсии ошибок постоянны для всех наблюдений. Иногда случайный член будет больше, иногда меньше, иногда больше, но не должно быть априорной причины для того, чтобы он порождал большую ошибку в одних наблюдениях, чем в других. Условие независимости ошибок от номера наблюдения называют *гомоскедастичностью*. Случай, когда условие гомоскедастичности нарушается, называется *гетероскедастичностью*. Этот случай можно иногда наблюдать графически:

Рисунок 1.

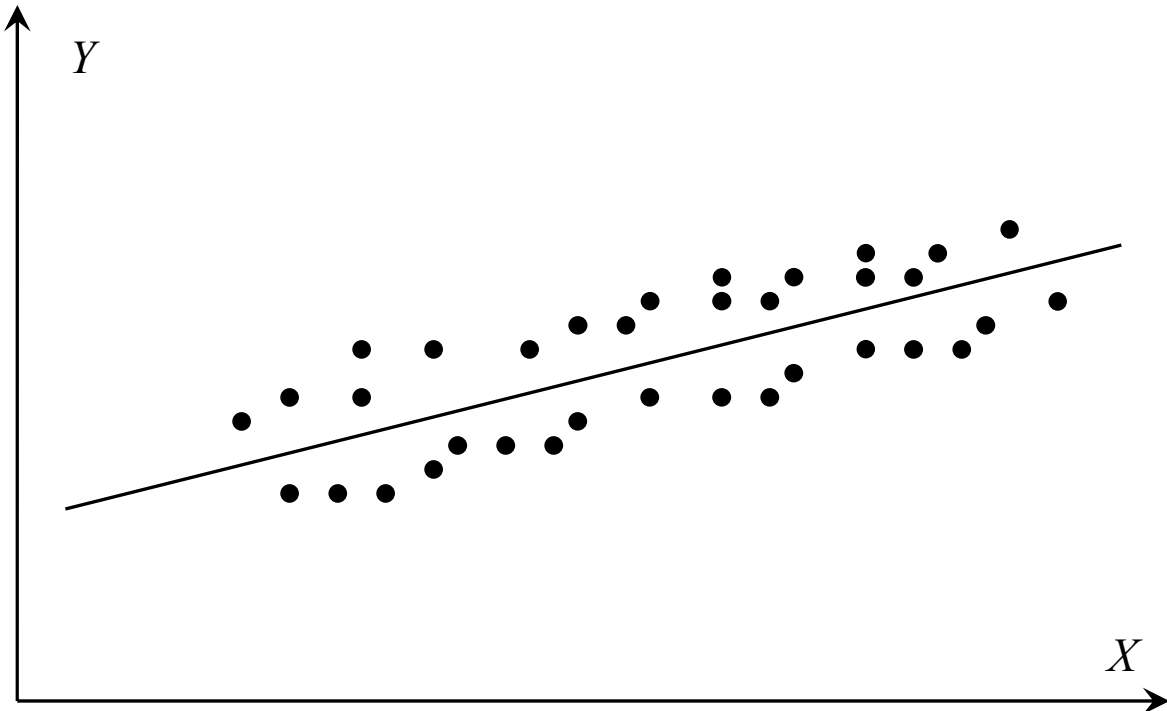


Рис. 1. Гомоскедастичность

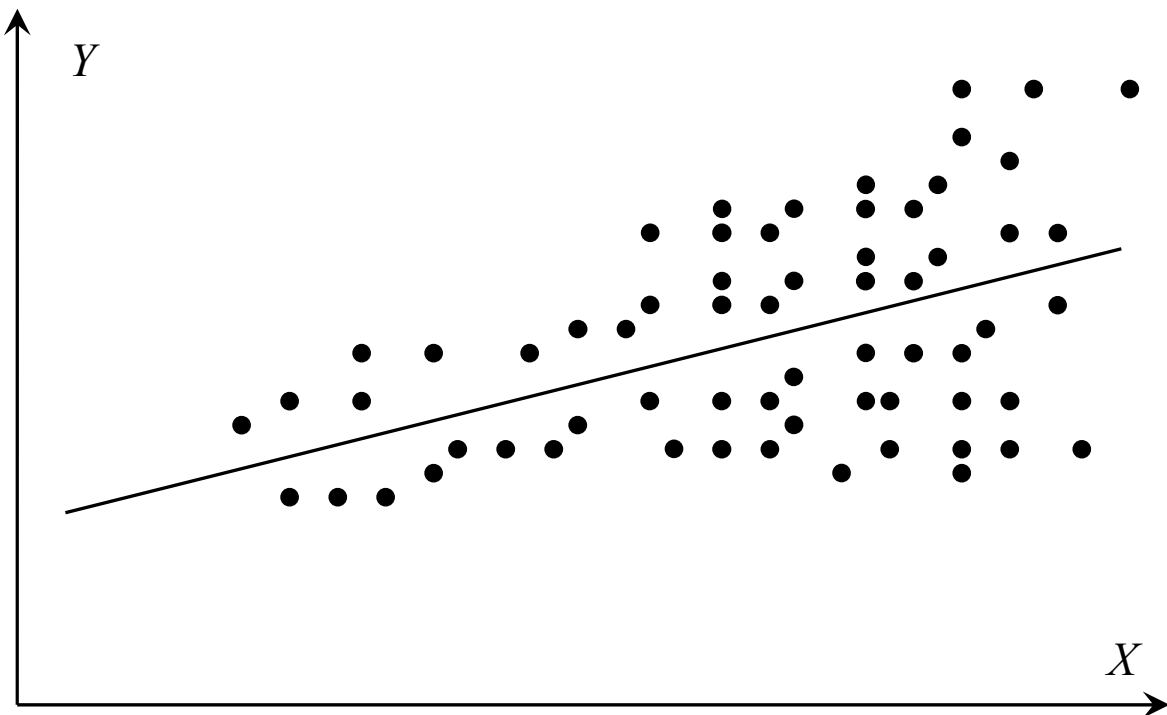


Рис. 2. Гетероскедастичность

5. Условие указывает на некоррелированность ошибок для разных наблюдений. Условие предполагает отсутствие систематической связи между

значениями случайного члена в любых двух наблюдениях. Это условие почти всегда нарушается, если наши данные представляют собой временные ряды. В случае если это условие не выполняется, говорят об автокорреляции остатков. Для простейшего случая $M(\varepsilon_i \varepsilon_{i+1}) = \rho$ - автокорреляционный процесс первого порядка – типичный вид данных представлен на рисунке 2.

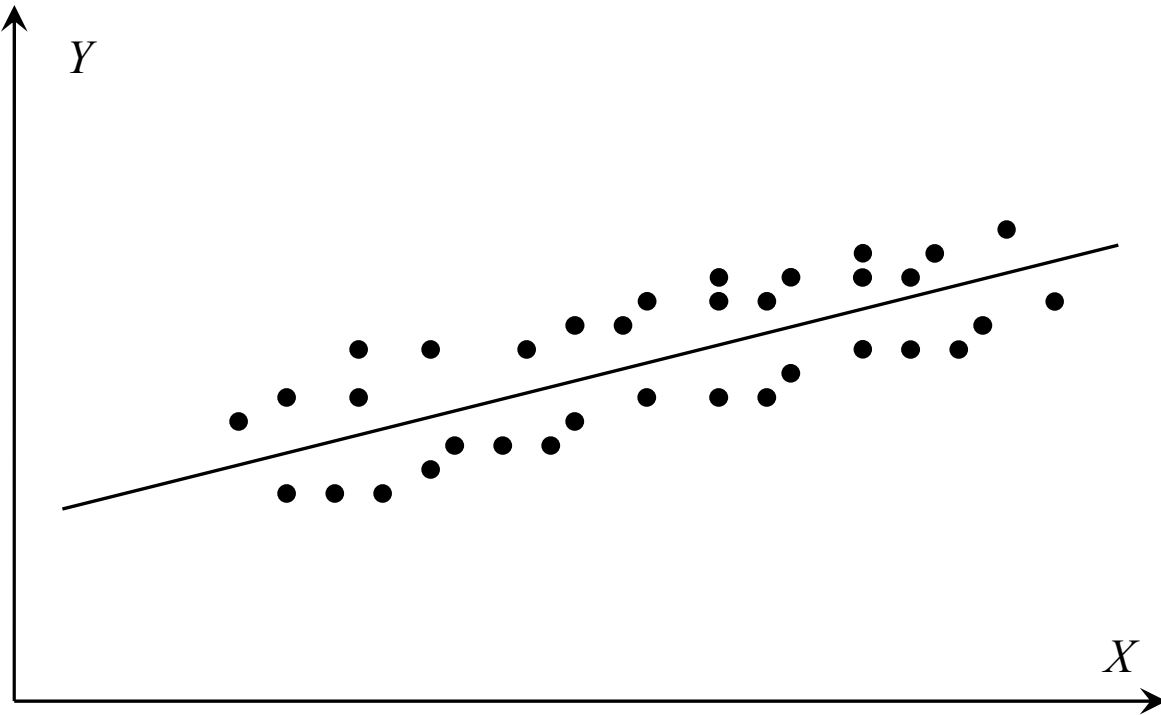


Рис. 3. Автокорреляция отсутствует

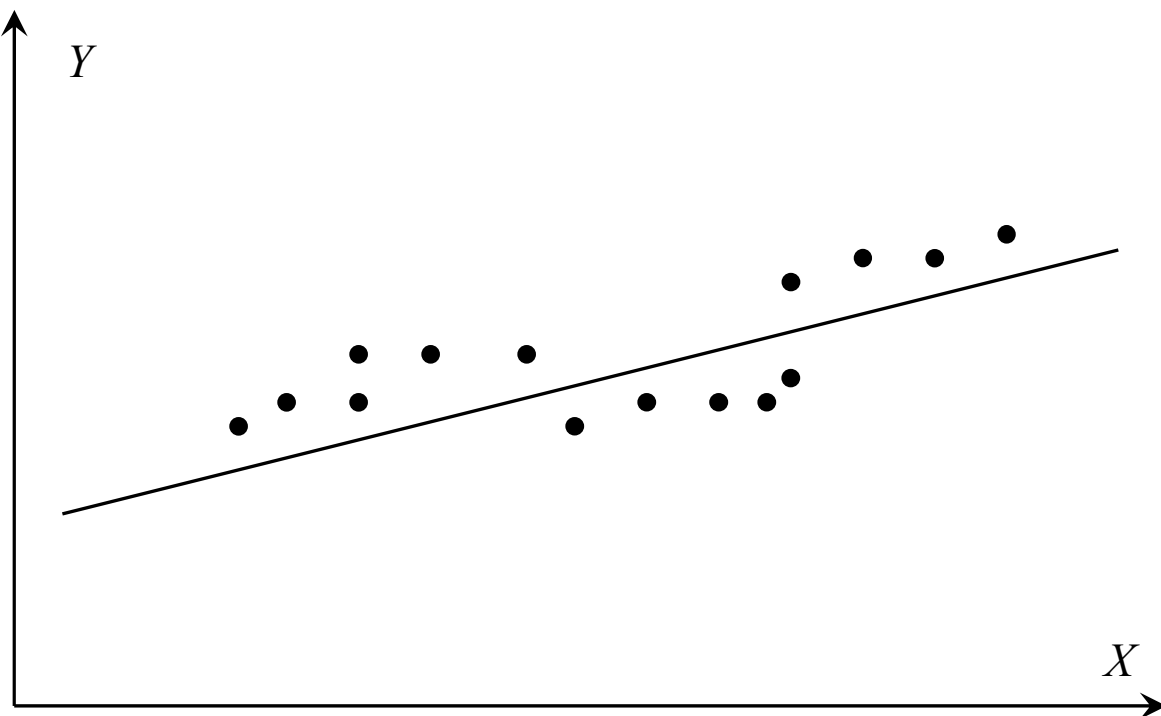


Рис. 4. Положительная автокорреляция первого порядка $\rho > 0$

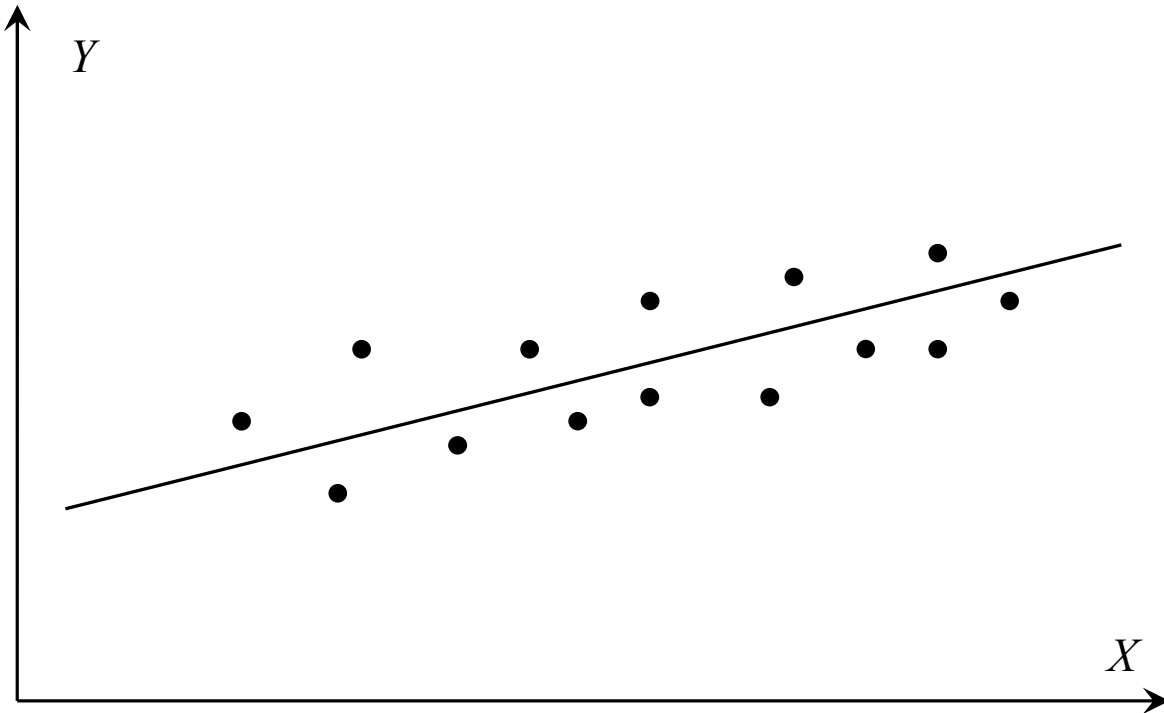


Рис. 6. Отрицательная автокорреляция первого порядка $\rho < 0$

Автокорреляция иногда является следствием неправильного выбора формы зависимости:

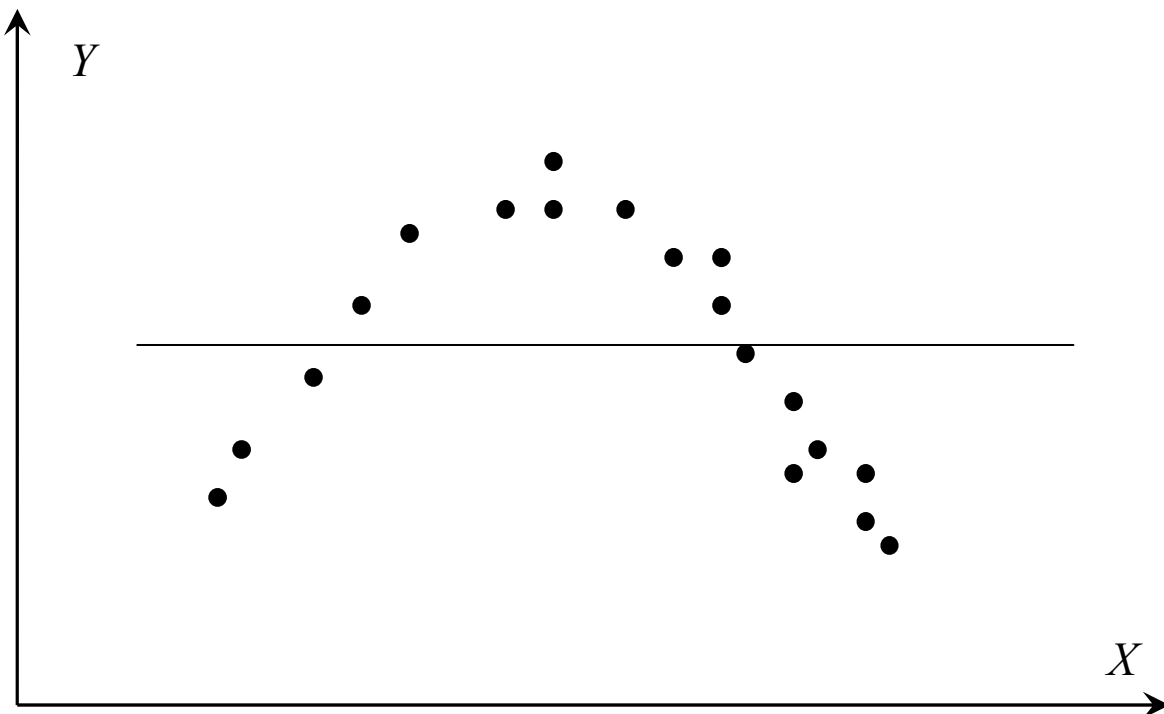


Рис. 7. Автокорреляция, вызванная неправильным выбором регрессионной функции

6. Это предположение не является чем-то сверхъестественным и высосанным из теоретического пальца. Действительно, как мы отмечали на прошлом занятии, ε_i включает в себя много факторов, которые, в принципе, можно считать независимыми. Отсюда, как следует из центральной предельной теоремы Ляпунова, ε_i будут иметь почти нормальное распределение.

Отметим, что в случае КЛРМ условие 6 эквивалентно условию статистической независимости ошибок для разных наблюдений. Действительно, если две нормально распределенные величины не коррелированы, то они независимы. В общем случае это не выполняется. А поскольку они независимы, то вектор ошибок ε имеет множественное нормальное распределение или величины ε_i будут иметь совместное нормальное распределение с вектором средних 0 и ковариационной матрицы $\sigma_\varepsilon^2 I_N$.

Итак, мы с вами находимся в условиях КЛРМ. Посмотрим, какими свойствами обладают в этом случае наши оценки

Коэффициенты, рассчитанные при помощи метода наименьших квадратов, являются статистическими оценками неизвестных коэффициентов регрессионного уравнения. По имеющейся выборке мы можем построить несколько оценок одного и того же параметра. Нас будут интересовать не все возможные оценки, а лишь оценки, обладающие определенными свойствами. Вот эти свойства:

Свойства статистических оценок:

- асимптотические – проявляется при больших объемах выборки, показывает, что происходит со статистической оценкой при увеличении объема выборки (состоятельность, асимптотическая несмещенность, асимптотическая нормальность, асимптотическая эффективность)

- свойства при фиксированном объеме выборки (несмещенность, эффективность)

Желаемые свойства оценок следующие – несмещенность, эффективность, состоятельность.

Состоятельность. Оценка $\hat{\theta}$ называется *состоятельной*, если при увеличении объема выборки значения оценки стремятся по вероятности к истинному значению оцениваемого параметра:

$$p \lim_{N \rightarrow \infty} \hat{\theta} = \theta.$$

Для доказательства состоятельности статистических оценок используются теоремы, относящиеся к законам больших чисел (теорема Чебышева) и теорема Слуцкого.

Теорема Слуцкого. Пусть $f(x, y)$ непрерывна в точке (a, b) и случайные последовательности X_n и Y_n сходятся по вероятности к a и b соответственно:

$$p \lim_{n \rightarrow \infty} X_n = a, \quad p \lim_{n \rightarrow \infty} Y_n = b, \quad \text{Тогда } f(X_n, Y_n) \text{ сходится по вероятности к } f(a, b).$$

Несмещенность. *Несмещенной* называют статистическую оценку $\hat{\theta}$, математическое ожидание которой равно истинному значению оцениваемого параметра, т. е. $E\hat{\theta} = \theta$.

Оценку, которая не удовлетворяет этому свойству, называют смещенной: $E\hat{\theta} \neq \theta$. Смещенность оценки означает присутствие в оценке систематических ошибок (ошибок одного знака), т. е. смещенная оценка завышает или занижает истинное значение параметра.

Величину смещения обозначают следующим образом:

$$bias\hat{\theta} = \theta - E\hat{\theta}$$

Для несмещенных оценок $bias\hat{\theta} = 0$

Несмещенность статистической оценки доказывается непосредственно.

Теорема. Если $\hat{\theta}$ - несмещенная статистическая оценка и $Var\hat{\theta} \xrightarrow{N \rightarrow \infty} 0$, то она состоятельна.

Что лучше: смещенная или несмещенная оценка? Однозначного ответа нет.

Для одного и того же параметра существует бесконечно много несмещенных, смещенных, состоятельных оценок. Как выбрать самую точную оценку?

Если рассматриваются две несмещенные оценки, то критерием точности является дисперсия – выбираем ту статистическую оценку, дисперсия которой меньше.

Говорят, что несмещенная оценка $\hat{\theta}_1$ **более эффективна**, чем несмещенная оценка $\hat{\theta}_2$, если ее дисперсия меньше

$$Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$$

Эффективной в классе несмещенных оценок называют несмещенную оценку, которая при заданном объеме выборки N имеет наименьшую возможную дисперсию

Для доказательства эффективности несмещенной статистической оценки используется неравенство Рао-Фреше-Крамера.

Информацией Фишера о неизвестном параметре θ называется величина

$$I(\theta) = E\left(\frac{\partial \ln f(x, \theta)}{\partial \theta}\right)^2$$

Теорема Рао-Фреше-Крамера (Рао-Крамера). Пусть плотность распределения случайной величины X $f_X(x, \theta)$ удовлетворяет условиям регулярности:

- область возможных значений случайной величины не зависит от θ ;
- Информация Фишера конечна и положительна $I(\theta)$.

Тогда для произвольной несмещенной оценки $\hat{\theta}$, построенной по выборке объема N , выполняется неравенство (Рао-Фреше-Крамера):

$$Var(\hat{\theta}) \geq \frac{1}{N * I(\theta)}.$$

Так что если для какой-то несмещенной оценки ее дисперсия достигает нижней границы, определяемой неравенством Рао-Фреше-Крамера, то она является эффективной.

Если рассматривать все оценки, смещенные и несмещенные, то статистическая оценка называется эффективной, если она доставляет минимум ее среднеквадратической ошибки:

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + bias^2(\hat{\theta}) = Var(\hat{\theta}) + (\theta - E\hat{\theta})^2$$

Как правило, эконометристов более интересует состоятельность оценки, чем ее Несмещенность. Смещенная, но состоятельная оценка может не равняться истинному значению в среднем, но с ростом выборки будет приближаться к истинному значению параметра.

Свойства (с доказательствами для парного случая):

Свойство 1. Линейная зависимость оценок от наблюдаемых значений Y .

$$\hat{\beta} = \frac{\sum z_i u_i}{\sum z_i^2} = \frac{\sum z_i (y_i - \bar{y})}{\sum z_i^2} = \frac{\sum z_i y_i}{\sum z_i^2} - \frac{\bar{y} \sum z_i}{\sum z_i^2} = \sum w_i y_i$$

поскольку $\sum z_i = 0$ в силу того, что

$$\sum (x_i - \bar{x}) = \sum x_i - N\bar{x} = \sum x_i - \sum x_i = 0$$

$w_i = \frac{z_i}{\sum z_i^2}$, если X - детерминированный вектор, то w -

детерминированный вектор (при повторении выборок значения не меняются).

Легко убедиться, что

$$\sum w_i = 0, \sum w_i^2 = \frac{1}{\sum z_i^2}, \sum w_i z_i = \sum w_i x_i = 1$$

Аналогично преобразовывая выражение для $\hat{\alpha}$, мы получим

$$\hat{\alpha} = \sum \left(\frac{1}{N} - \bar{x} w_i \right) y_i$$

Свойство 2. $\hat{\beta} = (X'X)^{-1} X'Y = (X'X)^{-1} X'(X\beta + \varepsilon) = \beta + (X'X)^{-1} X'\varepsilon$

$M\hat{\beta} = \beta + M((X'X)^{-1} X'\varepsilon) = \beta$, т. е. $\hat{\beta}$ - несмещенная оценка β .

$$\hat{\beta} = \sum w_i y_i = \sum w_i (\alpha + \beta x_i + \varepsilon_i) = (2.9) = \beta + \sum w_i \varepsilon_i$$

$$M\hat{\beta} = \beta + \sum w_i M\varepsilon_i = \beta,$$

Для доказательства мы использовали 2 и 3.

Свойство 3. Матрица ковариаций оценок:

$$\hat{\beta} - \beta = \sum_{i=1}^N w_i Y_i$$

$$D\hat{\beta} = M(\hat{\beta} - \beta)^2 = M\left(\sum_{i=1}^N w_i \varepsilon_i\right)^2 = M(w_1^2 \varepsilon_1^2 + \dots + w_N^2 \varepsilon_N^2 + 2w_1 w_2 \varepsilon_1 \varepsilon_2 + \dots + 2w_{N-1} w_N \varepsilon_{N-1} \varepsilon_N) = \sigma_\varepsilon^2 \sum_{i=1}^N w_i^2 = \frac{\sigma_\varepsilon^2}{\sum_{i=1}^N (X_i - \bar{X})^2}$$

$$\sigma_{\hat{\beta}} = \frac{\sigma_\varepsilon}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2}}.$$

Аналогично выводится формула для $D\hat{\alpha}$

$$D\hat{\alpha} = \sigma_\varepsilon^2 \frac{\sum_{i=1}^N X_i^2}{N \sum_{i=1}^N (X_i - \bar{X})^2}$$

Подобным образом можно отыскать ковариацию:

$$\text{cov}(\hat{\alpha}, \hat{\beta}) = \frac{-\bar{X}}{\sum_{i=1}^N (X_i - \bar{X})^2} \sigma_{\varepsilon}^2.$$

$\hat{\beta} - \beta = (X'X)^{-1} X' \varepsilon$ - из предыдущего пункта.

$$\begin{aligned} D\hat{\beta} &= M((\hat{\beta} - \beta)(\hat{\beta} - \beta)') = M((X'X)^{-1} X' \varepsilon \varepsilon' X (X'X)^{-1}) = \\ &= (X'X)^{-1} X' M(\varepsilon \varepsilon') X (X'X)^{-1} = (X'X)^{-1} X' \sigma_{\varepsilon}^2 I_N X (X'X)^{-1} = \\ &= \sigma_{\varepsilon}^2 (X'X)^{-1} X' X (X'X)^{-1} = \sigma_{\varepsilon}^2 (X'X)^{-1} \end{aligned}$$

$$D\hat{\beta} = \sigma_{\varepsilon}^2 (X'X)^{-1}$$

(пользовались тем, что матрица, обратная к симметричной, так же симметричная).

пользовались 3, 4 и 5.

$$D\hat{\beta}_i = \sigma_{\hat{\beta}_i}^2 = \sigma_{\varepsilon}^2 a^{ii}, \text{ где } a^{ii} - i\text{-й диагональный элемент матрицы } (X'X)^{-1}$$

Свойство 4. Теорема Гаусса-Маркова.

В условиях 1-5 МНК-оценки МЛРМ представляют собой наилучшие линейные несмещенные оценки, т. е. в классе линейных несмещенных оценок МНК-оценки обладают наименьшей дисперсией.

Best Linear Unbiased Estimation (BLUE)

Важность теоремы Гаусса-Маркова. Мы можем придумать много оценок возможных для коэффициентов β , в частности, можем придумать много линейных оценок, т. е. таких оценок, которые выражаются в виде взвешенного среднего наблюдений объясняемой переменной. Некоторые из этих оценок могут быть несмещенными как, например, «наивная» оценка. Так вот, оценки коэффициентов уравнения по методу наименьших квадратов в случае классической парной модели – это наилучшие оценки в том смысле, что среди всех возможных линейных несмещенных оценок эти оценки имеют наименьшую дисперсию. *Best Linear Unbiased Estimator – BLUE* Вопрос нахождения такой оценки будет возникать в нашем курсе снова и снова, т. к. мы увидим, что при нарушении условий Гаусса-Маркова МНК-оценки уже не

будут «BLUE». В этом случае наша цель будет заключаться в построении других оценок, не МНК, которые уже будут «BLUR».

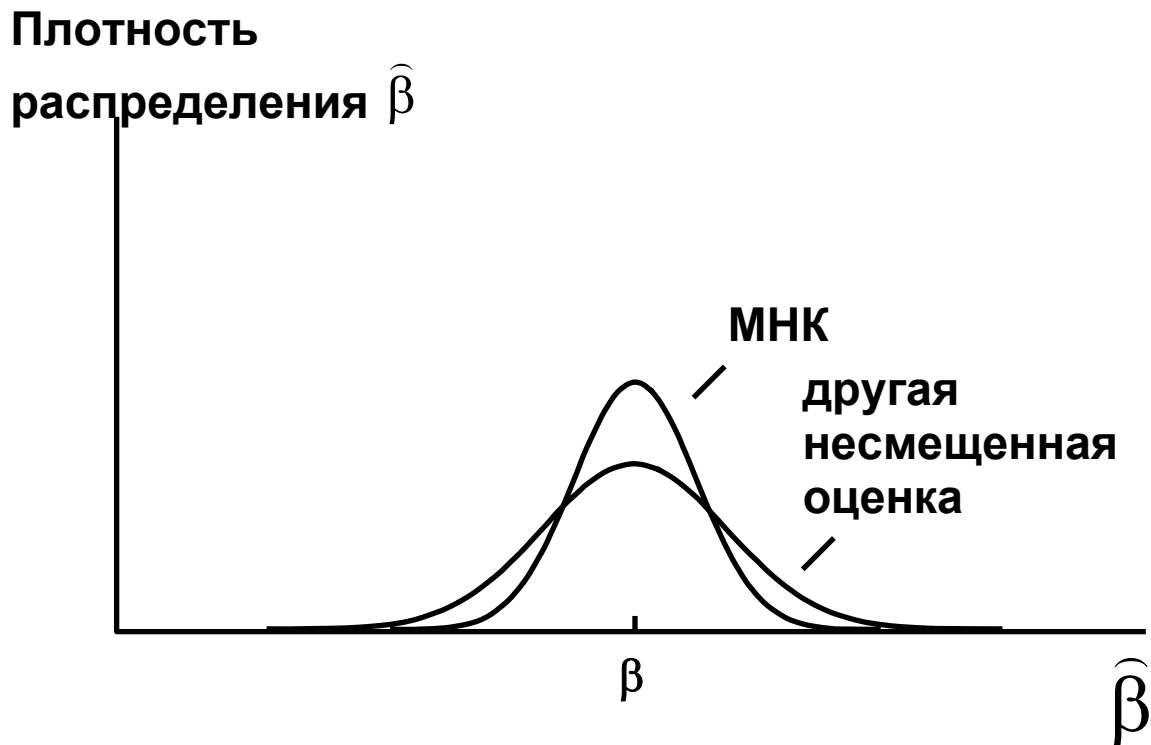


Рис 8. «BLUR» - оценка.

Обратите внимание, что в выражении матрицы ковариаций $\hat{\beta}$ фигурирует дисперсия остаточного члена. Однако на практике мы эту дисперсию не знаем, поскольку не знаем ε_i , поэтому не можем вычислить теоретическую матрицу ковариаций $\hat{\beta}$. Мы сможем построить оценку этой матрицы, если сможем оценить σ^2 по результатам наблюдений. Никакой информацией об остаточном члене ε_i мы не располагаем. Единственно, на что мы можем опираться - на остатки или невязки e_i . Разброс остатков относительно линии регрессии будет отражать разброс ε относительно истинной неизвестной прямой. В общем случае остаток и ошибка в любом данном наблюдении неравны друг другу. Для оценки σ_ε^2 используем $\sum_{i=1}^N e_i^2$:

Свойство 5. $s_{\varepsilon}^2 = \frac{\sum_{i=1}^N e_i^2}{N-k-1} = \frac{e'e}{N-k-1}$ - несмещенная оценка σ_{ε}^2

Итак, оценка s_{ε}^2 является несмещенной оценкой дисперсии σ_{ε}^2 . Тогда оценки матрицы ковариаций оценок будут следующими:

$$\widehat{D}\widehat{\beta} = s_{\varepsilon}^2 (X'X)^{-1}$$

$$s_{\widehat{\beta}_i}^2 = s_{\varepsilon}^2 a^{ii}$$

Для парной модели

$$s_{\widehat{\beta}}^2 = \frac{s_{\varepsilon}^2}{\sum_{i=1}^N (x_i - \bar{x})^2}, \quad s_{\widehat{\alpha}}^2 = s_{\varepsilon}^2 \frac{\sum_{i=1}^N x_i^2}{\sum_{i=1}^N (x_i - \bar{x})^2}$$

Стандартные отклонения коэффициентов регрессии, вычисленные на основе предыдущей формулы, приводятся в результатах регрессии практически во всех статистических пакетах.

До сих пор мы нигде не использовали свойство 6, т. е. не делали никаких предположений о распределении вероятностей ошибок ε_i . Что будет, если мы запостулируем нормальную форму этого распределения.

Свойство 6.

В предположениях НЛРМ $\widehat{\beta} \sim N(\beta, \sigma_{\varepsilon}^2 (X'X)^{-1})$

Свойство 7. В случае НРЛМ

$$\frac{e'e}{\sigma_{\varepsilon}^2} = \frac{(N-k)s_{\varepsilon}^2}{\sigma_{\varepsilon}^2} \sim \chi^2(N-k) \text{ - без доказательства.}$$

Свойство 8. В условиях НЛРМ оценки s_{ε}^2 $\widehat{\beta}$ независимы. - без доказательства.

7. ПРОВЕРКА ГИПОТЕЗ ОТНОСИТЕЛЬНО КОЭФФИЦИЕНТОВ РЕГРЕССИИ.

Предположим, что мы находимся в условиях НМЛРМ.

1. Проверка гипотезы о равенстве коэффициента регрессионного уравнения некоторому числу.

$$H_0: \beta = \beta_0$$

$$H_a: \beta \neq \beta_0$$

или учитывая, что $\hat{\beta}$ - несмещенная оценка β , можем переписать гипотезу:

$$H_0: M\hat{\beta} = \beta_0.$$

$$H_a: M\hat{\beta} \neq \beta_0$$

Поскольку $\hat{\beta} \sim N(\beta, \sigma_\varepsilon^2 (X'X)^{-1})$, то $\hat{\beta} - \beta \sim N(0, \sigma_\varepsilon^2 (X'X)^{-1})$ или $\hat{\beta}_i - \beta_i \sim N(0, \sigma_{\hat{\beta}_i}^2)$, где $\sigma_{\hat{\beta}_i}^2 = \sigma_\varepsilon^2 a^{ii}$.

$$\text{Поэтому } \frac{\hat{\beta}_i - \beta_i}{\sigma_\varepsilon \sqrt{a^{ii}}} \sim N(0,1).$$

Далее, $(N-k) \frac{s_\varepsilon^2}{\sigma_\varepsilon^2} \sim \chi^2(N-k-1)$ и оценки s_ε^2 и $\hat{\beta}$ независимы,

следовательно,

$$\frac{\frac{\hat{\beta}_i - \beta_i}{\sigma_\varepsilon \sqrt{a^{ii}}}}{\sqrt{\frac{(N-k-1)s_\varepsilon^2}{(N-k-1)\sigma_\varepsilon^2}}} = \frac{\hat{\beta}_i - \beta_i}{s_\varepsilon \sqrt{a^{ii}}} = \frac{\hat{\beta}_i - \beta_i}{s_{\hat{\beta}_i}} \sim t(N-k-1).$$

Вычисляем наблюдаемое значение критерия $t_{\text{набл.}}$.

Для проверки нулевой гипотезы при различных альтернативных гипотезах:

$$H_a: \beta_i \neq \beta_{i0}.$$

$t_{кр}$ находим из таблиц критических точек распределения Стьюдента с $N-k-1$ степенями свободы для выбранного уровня значимости α и учитывая, что критическая область двусторонняя - $t_{кр}^{\partial\epsilon}(\nu, N-k-1)$. Далее, если

$|t_{набл}| < t_{кр}^{\partial\epsilon}(\nu, N-k-1)$, то мы говорим, что у нас нет оснований отвергнуть нулевую гипотезу, если же

$|t_{набл}| > t_{кр}^{\partial\epsilon}(\nu, N-k-1)$, то мы нулевую гипотезу отвергаем.

Если же у нас критерий односторонний, то все сохраняется, за исключением критического значения статистики. Его мы ищем по таблицам критических точек распределения Стьюдента с $N-k-1$ степенями свободы для выбранного уровня значимости α и учитывая, что критическая область односторонняя - $t_{кр}^{одн}(\alpha, N-k-1)$. Выполняется следующее соотношение между односторонними и двусторонними критическими точками:

$$t_{кр}^{одн}(\alpha/2, N-k-1) = t_{кр}^{\partial\epsilon}(\alpha, N-k-1)$$

Особенно просто критерий выглядит в случае, когда $\beta_{i0} = 0$, т. е. в случае, когда мы хотим убедиться в значимости этого коэффициента и таким образом убедиться в наличии связи между Y и X_i : $t = \frac{\hat{\beta}_i}{s_{\hat{\beta}_i}}$ - t -статистика i -го коэффициента МЛРМ. Значение этой статистики приводятся почти всеми статистическими пакетами.

Если мы теперь рассмотрим неравенство

$$P(|t| > t_{кр}^{\partial\epsilon}) = \mu$$

$$P(|t| < t_{кр}^{\partial\epsilon}) = 1 - \mu = \gamma$$

$$P\left(\left|\frac{\hat{\beta} - \beta}{s_{\hat{\beta}}}\right| < t_{кр}^{\partial\epsilon}\right) = \gamma$$

Разрешим это неравенство относительно β :

$$P(\hat{\beta} - s_{\hat{\beta}} t_{кр}^{\partial\epsilon} < \beta < \hat{\beta} + s_{\hat{\beta}} t_{кр}^{\partial\epsilon}) = \gamma$$

$(\hat{\beta}_i - s_{\hat{\beta}_i} t_{кр}^{доуст}; \hat{\beta}_i + s_{\hat{\beta}_i} t_{кр}^{доуст})$ - доверительный интервал для параметра β_i с

уровнем надежности γ . В этом случае говорят, что доверительный интервал с вероятностью γ покрывает истинное значение параметра β_i .

Не говорят, что доверительный интервал содержит с вероятностью γ истинное значение параметра β . Поскольку истинное значение параметра существует независимо от нас, а доверительный интервал мы строим, т. о. не β попадает в доверительный интервал, а доверительный интервал с той или иной вероятностью попадает на β .

2. Тестирование регрессионного уравнения.

Пусть константа включена в число регрессоров.

Процедура разделения вариации переменной Y на две составляющие позволяет провести нам тест на существование линейной зависимости между переменной Y и переменными X_1, \dots, X_k .

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

Таким образом, справедливость нулевой гипотезы означает, что ни одна из переменных X_1, \dots, X_k не помогает нам объяснить вариацию Y . Эта гипотеза позволяет нам судить о значимости регрессии в целом. Эта гипотеза об отсутствии линейной связи между Y и X_1, \dots, X_k .

Проверка нулевой гипотезы осуществляется при помощи следующего критерия:

$$F_{k, N-k-1} = \frac{R^2}{1-R^2} \frac{N-k-1}{k} = \frac{RSS/k}{ESS/(N-k-1)}$$

При справедливости нулевой гипотезы данная статистика имеет распределение Фишера с числом степеней свободы числителя k и знаменателя $N-k-1$.

Если нулевая гипотеза верна, то следует ожидать, что RSS , R^2 и, следовательно, F , близки к нулю. Таким образом, если значение F -статистики велико, мы нулевую гипотезу отвергаем. Граничное значение, начиная с

которого мы отвергаем гипотезу, находится из таблиц распределения Фишера для выбранного уровня значимости ν и числу степеней свободы числителя k и знаменателя $N-k-1$ - $F_{кр}(\nu, k, N - k - 1)$. Таким образом, если $F > F_{кр}(\nu, k, N - k - 1)$, мы нулевую гипотезу отвергаем, делаем вывод о том, что хотя бы одна из объясняющих переменных, участвующих в модели, действительно линейно влияет на переменную Y .

Итак, при помощи F -статистики мы проверяем значимость коэффициента детерминации. Если F -статистика незначимо отличается от нуля, это означает, что объясняющие переменные, участвующие в модели на самом деле не очень-то нам помогают объяснить вариацию переменной Y .

Для парного случая F – статистика выглядит следующим образом:

$$F = (N - 2) \frac{R^2}{1 - R^2} = \frac{\hat{\beta}^2 \sum_{i=1}^N (x_i - \bar{x})^2}{\sum_{i=1}^N e_i^2 / (N - 2)} \text{ - Упражнение}$$

Сравнивая предыдущее выражение и выражение для t -статистики коэффициента наклона, получим, что $F = t^2$:

$$t^2 = \left(\frac{\hat{\beta} - \beta}{s_{\hat{\beta}}} \right)^2 = \frac{(\hat{\beta} - \beta)^2}{\frac{s_{\varepsilon}^2 \sum_{i=1}^N (x_i - \bar{x})^2}{N - 2}} = \frac{(\hat{\beta} - \beta)^2 \sum_{i=1}^N (x_i - \bar{x})^2}{\sum_{i=1}^N e_i^2 / (N - 2)} = F .$$

Таким образом, проверка гипотезы $H_0: \beta = 0$, используя F и t -статистики, дает для одномерной регрессионной модели идентичные результаты.

3. Объединенный тест на несколько коэффициентов регрессии.

При помощи F -статистики мы теперь умеем проверять гипотезу о том, что все коэффициенты при объясняющих переменных равны нулю. Иногда возникают ситуации, когда нам необходимо проверить гипотезу о том, что

нулю равны не все коэффициенты при объясняющих переменных, а некоторые из них. В этом случае осуществляется следующая процедура.

Рассмотрим модель множественной регрессии:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon - \text{«длинная регрессия»}.$$

Назовем эту модель моделью без ограничений (UR), поскольку здесь мы не делаем никаких ограничений на возможные значения коэффициентов регрессии. Предположим, что мы хотим протестировать гипотезу о том, что q последних коэффициентов регрессии одновременно равны нулю. Т. е. мы хотим проверить гипотезу о том, что $\beta_{k-q} = \dots = \beta_k = 0$. Перепишем предыдущее уравнение следующим образом:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_{k-q-1} X_{k-q-1} + \beta_{k-q} X_{k-q} + \dots + \beta_k X_k + \varepsilon$$

нулевая гипотеза выглядит следующим образом:

H_0 : $\beta_{k-q} = \dots = \beta_k = 0$, т. е. последние q коэффициентов одновременно равны нулю.

В случае, если эта гипотеза справедлива, то истинная модель выглядит следующим образом:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_{k-q-1} X_{k-q-1} + \varepsilon - \text{«короткая регрессия»}$$

Назовем эту модель моделью с ограничениями (R – restricted model).

Оценим обе эти модели и посчитаем сумму квадратов остатков в модели с ограничениями и в модели без ограничений – ESS_R и ESS_{UR} соответственно. ESS_R всегда больше, чем ESS_{UR} . Этот результат эквивалентен тому, что R^2 всегда увеличивается при добавлении в модель новых объясняющих переменных. Если нулевая гипотеза справедлива, выбрасывание из уравнения q последних объясняющих переменных несильно скажется на объясняющих качествах уравнения, и ESS_R будет ненамного отличаться от ESS_{UR} . Таким образом, если нулевая гипотеза справедлива, разница $ESS_R - ESS_{UR}$ будет ненамного отличаться от нуля. Статистический критерий для проверки нулевой гипотезы следующий:

$$F_{q, N-k-1} = \frac{(ESS_R - ESS_{UR})/q}{ESS_{UR}/(N-k-1)}$$

При справедливости нулевой гипотезы данная статистика имеет распределение Фишера с числом степеней свободы числителя q и знаменателя $N-k-1$.

Если нулевая гипотеза справедлива, выбрасывание из уравнения q последних объясняющих переменных несильно скажется на объясняющих качествах уравнения, и ESS_R будет ненамного отличаться от ESS_{UR} . Таким образом, если нулевая гипотеза справедлива, разница $ESS_R - ESS_{UR}$ будет ненамного отличаться от нуля. Следовательно, F -статистика будет достаточно мала. Граничное значение, при котором нулевую гипотезу отвергают, зависит от выбранного уровня значимости ν . Оно находится из таблиц распределения Фишера для выбранного уровня значимости ν и числу степеней свободы числителя q и знаменателя $N-k-1$. Таким образом, если мы нулевую гипотезу отвергаем, то делаем вывод о том, что наши переменные действительно оказывают влияние на переменную Y и включение их в модель существенно повышает объясняющую силу уравнения.

Похожий подход – рассмотрение регрессии с ограничением регрессии без ограничений – можно применить и для проверки гипотезы о наличии линейных связей между коэффициентами. Например, нам может понадобиться в ходе нашего исследования проверить гипотезу о равенстве между собой нескольких коэффициентов регрессии.

4. Проверка гипотезы о наличии линейных ограничений на коэффициенты.

Предположим, мы рассматриваем и оцениваем функцию потребления:

$C = \beta_0 + \beta_1 X_L + \beta_2 X_{NL} + \varepsilon$, где X_L – трудовые доходы, а X_{NL} – нетрудовые доходы. В этом случае нам может понадобиться проверить гипотезу о том, что предельные склонности к потреблению равны между собой ($\beta_1 = \beta_2$) или

гипотезу о том, что общая предельная склонность к потреблению равна 1 ($\beta_1 + \beta_2 = 1$).

Рассмотрим сначала первый случай.

Суть подхода к проверке таких гипотез такая же, как и в предыдущем пункте. Мы оцениваем две регрессии – регрессию без ограничений и регрессию с ограничениями, составляем F – статистику и проверяем ее значимость при помощи таблиц распределения Фишера.

Рассмотрим сначала первый случай.

Нулевая гипотеза: $H_0: \beta_1 = \beta_2$

Модель без ограничений: $C = \beta_0 + \beta_1 X_L + \beta_2 X_{NL} + \varepsilon$;

модель с ограничениями: $C = \beta_0 + \beta_1 (X_L + X_{NL}) + \varepsilon$.

Во втором случае моделью с ограничениями будет следующая модель:

$C - X_{NL} = \beta_0 + \beta_1 (X_L - X_{NL}) + \varepsilon$.

Здесь мы просто подставили в исходную модель выражение для β_2 : $\beta_2 = 1 - \beta_1$.

Статистический критерий для проверки нулевой гипотезы следующий:

$$F_{q, N-k-1} = \frac{(ESS_R - ESS_{UR}) / q}{ESS_{UR} / (N - k - 1)}$$

При справедливости нулевой гипотезы данная статистика имеет распределение Фишера с числом степеней свободы числителя q и знаменателя $N-k-1$, где q – число ограничений, накладываемых на коэффициенты. В нашем случае оно равно 1.

В статистических пакетах проверка гипотезы о наличии линейных ограничений на коэффициенты называется тестом Вальда (Wald test).

Рассмотрим эту гипотезу в общем виде:

$H_0: H\beta = r$.

Например:

$$\begin{pmatrix} 2 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ означает, что } \begin{cases} 2\beta_1 + \beta_3 = 0 \\ \beta_2 + \beta_3 = 1 \end{cases} .$$

H – матрица размера $q \times k$, где q – число ограничений, r – вектор из q компонент.

Для проверки такой гипотезы используется статистика Вальда:

$$W = \frac{1}{s^2} [H\hat{\beta} - r]' [h(X'X)^{-1}H']^{-1} [H\hat{\beta} - r]$$

При справедливости нулевой гипотезы эта статистика распределена асимптотически как $\chi^2(q)$. Для проверки нулевой гипотезы находим критическую точку распределения $\chi^2(q)$ для выбранного уровня значимости ν – $W_{кр}$. Если $W_{набл} > W_{кр}$, то мы нулевую гипотезу отвергаем, если $W_{набл} < W_{кр}$, то говорим, что нет оснований отвергнуть нулевую гипотезу.

Ту же самую гипотезу можно проверить при помощи статистики Фишера, вычислив суммы квадратов остатков для моделей с ограничением и модели без ограничений. Как связаны между собой эти статистики? Оказывается, что

$$F = \frac{W}{q} .$$

В пакете Eviews приводятся наблюдаемые значения обеих статистик и значения Probability для каждой из них.

значения Probability для каждой из них.

5. Проверка гипотезы о равенстве коэффициентов различных регрессионных уравнений (тест Чоу).

Предположим, что мы рассматриваем регрессионное уравнение $Y = \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$ и данные для его оценки содержат наблюдения для разных по качеству объектов: для мужчин и женщин, для белых и черных. вопрос, который нас может здесь заинтересовать, следующий – верно ли, что рассматриваемая модель совпадает для двух выборок, относящихся к объектам разного качества? Ответить на этот вопрос можно при помощи теста Чоу.

Рассмотрим модели:

$$Y_i = \beta'_1 X_{1i} + \dots + \beta'_k X_{ki} + \varepsilon''_i, \quad i=1, \dots, N \quad (1);$$

$$Y_i = \beta''_1 X_{1i} + \dots + \beta''_k X_{ki} + \varepsilon''_i, \quad i=N+1, \dots, N+M \quad (2).$$

В первой выборке N наблюдений, во второй – M наблюдений. Пример: Y – заработная плата, объясняющие переменные – возраст, стаж, уровень образования. Следует ли из имеющихся данных, что модель зависимости заработной платы от объясняющих переменных, стоящих в правой части одинакова для мужчин и женщин?

$$H_0: \beta'_1 = \beta''_1, \beta'_2 = \beta''_2, \dots, \beta'_k = \beta''_k$$

Для проверки этой гипотезы можно воспользоваться общей схемой проверки гипотез при помощи сравнения регрессии с ограничениями и регрессии без ограничений. Регрессией без ограничений здесь является объединение регрессий (1) и (2), т. е. $ESS_{UR} = ESS_1 + ESS_2$, число степеней свободы – $N + M - 2k$. Регрессией с ограничениями (т. е. регрессией в предположении, что выполнена нулевая гипотеза) будет являться регрессия $Y = \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$ для всего имеющегося набора наблюдений:

$$Y = \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon, \quad i = 1, \dots, N+M \quad (3).$$

Оценивая (3), получаем ESS_R . Для проверки нулевой гипотезы используем следующую статистику:

$$F_{k, N+M-2k} = \frac{(ESS_R - ESS_{UR})/k}{ESS_{UR}/(N+M-2k)}, \quad \text{которая в случае справедливости}$$

нулевой гипотезы имеет распределение Фишера с числом степеней свободы числителя k и знаменателя $N + M - 2k$.

Если нулевая гипотеза справедлива, мы можем объединить имеющиеся выборки в одну и оценивать модель для $N + M$ наблюдений. Если же нулевую гипотезу отвергаем, то мы не можем слить две выборки в одну, и нам придется оценивать эти две модели по отдельности.

Изучение общей линейной модели, рассмотренной нами ранее, весьма существенно, как мы видели, опирается на статистический аппарат. Однако, как и во всех приложениях мат. статистики, сила метода зависит от предположений, лежащих в его основе и необходимых для его применения. Некоторое время мы будем рассматривать ситуации, когда одна или более гипотез, лежащих в основе линейной модели, нарушается. Мы рассмотрим альтернативные методы оценивания в этих случаях. Мы увидим, что роль одних гипотез более существенна по сравнению с ролью других. Нам надо посмотреть, к каким последствиям может привести нарушения тех или иных условий (предположений), уметь проверить, удовлетворяются они или нет и знать, какие статистические методы можно и целесообразно применять, когда не подходит классический метод наименьших квадратов.

2. Связь между переменными линейная и выражается уравнением $Y = \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$ - ошибки спецификации модели (невключение в уравнение существенных объясняющих переменных, включение в уравнение лишних переменных, неправильный выбор формы зависимости между переменными);

3. X_1, \dots, X_k – детерминированные переменные – стохастические регрессоры, линейно независимые – полная мультиколлинеарность;

4. $M\varepsilon_i = 0$;

5. $M\varepsilon_i^2 = D\varepsilon_i = \sigma_\varepsilon^2$ - гетероскедастичность;

5. $M(\varepsilon_i, \varepsilon_j) = 0$ при $i \neq k$ – автокорреляция ошибок

Прежде чем приступить к разговору, рассмотрим следующие понятия: парный коэффициент корреляции и частный коэффициент корреляции.

8. МУЛЬТИКОЛЛИНЕАРНОСТЬ.

1. Совершенная мультиколлинеарность.

Одно из требований Гаусса-Маркова говорит нам о том, чтобы объясняющие переменные не были связаны никаким точным соотношением. Если такое соотношение между переменными существует, мы говорим о том, что в модели присутствует совершенная мультиколлинеарность. Пример. Рассмотрим модель со средней оценкой на экзамене, состоящую из трех объясняющих переменных: I – доход родителей, D – среднее число часов, затраченных на обучение в день, W – среднее число часов, затраченных на обучение в неделю. Очевидно, что $W=7D$. И это соотношение будет выполняться для каждого студента, который попадет в нашу выборку. Случай полной мультиколлинеарности отследить легко, поскольку в этом случае невозможно построить оценки по методу наименьших квадратов.

2. Частичная мультиколлинеарность или просто мультиколлинеарность.

Гораздо чаще встречается ситуация, когда между объясняющими переменными точной линейной зависимости не существует, но между ними существует тесная корреляционная зависимость – этот случай носит название реальной или частичной мультиколлинеарности (просто мультиколлинеарность) – существование тесных статистических связей между переменными. Надо сказать, что вопрос мультиколлинеарности – это вопрос скорее степени выраженности явления, а не его вида. Оценка любой регрессии будет страдать от нее в том или ином виде, если только все независимые переменные не окажутся абсолютно некоррелированными. Рассмотрение данной проблемы начинается только тогда, когда это начинает серьезно влиять на результаты оценки регрессии (наличие статистических связей между регрессорами вовсе не обязательно дает неудовлетворительные оценки). Итак, мультиколлинеарность – это проблема, когда тесная корреляционная

зависимость между регрессорами ведет к получению ненадежных оценок регрессии.

Последствия мультиколлинеарности:

Формально, поскольку $(X'X)$ – невырожденная, то мы можем построить МНК-оценки коэффициентов регрессии. Однако вспомним, как выражаются теоретические дисперсии оценок коэффициентов регрессии: $D\hat{\beta}_i = \sigma_{\hat{\beta}_i}^2 = \sigma_{\varepsilon}^2 a^{ii}$, где a^{ii} - i -й диагональный элемент матрицы $(X'X)^{-1}$. Поскольку матрица $(X'X)$ близка к вырожденной и $\det(X'X) \approx 0$, то

1) на главной диагонали обратной матрицы стоят очень большие числа, поскольку элементы обратной матрицы обратно пропорциональны $\det(X'X)$. Следовательно, теоретическая дисперсия i -го коэффициента достаточно большая и оценка дисперсии $s_{\hat{\beta}_i}$ так же большая, следовательно, t - статистики небольшие, что может привести к статистической незначимости i -го коэффициента. Т. е. переменная оказывает значимое влияние на объясняемую переменную, а мы делаем вывод о ее незначимости.

2) Поскольку оценки $\hat{\beta}_i$ и $s_{\hat{\beta}_i}$ зависят от $(X'X)^{-1}$, элементы которой обратно пропорциональны $\det(X'X)$, то если мы добавим или уберем одно-два наблюдения, добавив или убрав, таким образом, одну-две строки к матрице $X'X$, то значения $\hat{\beta}_i$ и $s_{\hat{\beta}_i}$ могут измениться существенным образом, вплоть до смены знака – неустойчивость результатов оценивания.

3) Трудность интерпретации уравнения регрессии. Допустим, у нас в уравнении есть две переменные, которые связаны между собой между собой: X_1 и X_2 . Коэффициент регрессии при X_1 интерпретируется как мера изменения Y за счет изменения X_1 при прочих равных условиях, т.е. значения всех других переменных остаются прежними. Однако, поскольку переменные X_1 и X_2 связаны, то изменения в переменной X_1 повлекут за собой предсказуемые изменения в переменной X_2 и значение X_2 не останется прежним.

Пример: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots$, где X_1 – общая площадь, X_2 – жилая площадь. Мы говорим: "Если жилая площадь увеличиться на 1 кв. м., то при прочих равных условиях цена квартиры увеличиться на β_2 долл". Однако в этом случае и жилая площадь увеличится на 1 кв. м. и прирост цены будет $\beta_1 + \beta_2$. Разграничить влияние на переменную Y каждой переменной в отдельности уже не представляется возможным. Выход в данной ситуации с ценой на квартиру – включить в модель не общую площадь, а так называемую "добавочную" или "дополнительную" площадь.

Признаки мультиколлинеарности.

Точных критериев для определения наличия (отсутствия) мультиколлинеарности не существует. Однако есть эвристические рекомендации по ее выявлению:

1) Анализируют матрицу парных коэффициентов корреляции между регрессорами и если значение коэффициента корреляции близко к 1, то это считается признаком мультиколлинеарности.

2) Анализ матрицы корреляции – лишь поверхностное суждение о наличии (отсутствии) мультиколлинеарности. Более внимательное изучение этого вопроса достигается при помощи расчета коэффициентов частной корреляции или расчетов коэффициентов детерминации каждой из объясняющих переменных по всем другим объясняющим переменным в регрессии $X_i = \beta_0 + \beta_1 X_1 + \dots + \beta_{i-1} X_{i-1} + \beta_{i+1} X_{i+1} + \dots + \beta_k X_k + \varepsilon$.

3) Можно посчитать определитель матрицы $(X'X)$ и если он близок к нулю, то это тоже свидетельствует о наличии мультиколлинеарности.

4) $(X'X)$ – симметричная положительно определенная матрица, следовательно, все ее собственные числа неотрицательны. Если определитель матрицы $(X'X)$ равен нулю, то минимальное собственное число так же ноль и непрерывность сохраняется. Следовательно, по значению минимального собственного числа λ_{\min} можно судить и о близости к нулю определителя

матрицы $(X'X)$. Кроме этого свойства минимальное собственное число важно еще и потому, что стандартная ошибка коэффициента обратно пропорциональна λ_{\min} .

5) О наличии мультиколлинеарности можно судить по внешним признакам, являющимся следствиями мультиколлинеарности:

а) некоторые из оценок имеют неправильные с точки зрения экономической теории знаки или неоправданно большие значения;

б) небольшое изменение исходных экономических данных приводит к существенному изменению оценок коэффициентов модели;

с) большинство t -статистик коэффициентов незначимо отличаются от нуля, в то же время модель в целом является значимой, о чем говорит высокое значение F -статистики.

Как избавиться от мультиколлинеарности, как ее устранить:

1) Использование факторного анализа. Переход от исходного набора регрессоров, среди которых есть статистически зависимые, к новым регрессорам Z_1, \dots, Z_m при помощи метода главных компонент – вместо исходных переменных вместо исходных переменных рассматриваем некоторые их линейные комбинации, корреляция между которыми мала или отсутствует вообще. Задача здесь – дать содержательную интерпретацию новым переменным Z . Если не удалось – возвращаемся к исходным переменным, используя обратные преобразования. Полученные оценки будут, правда, смещенными, но будут иметь меньшую дисперсию.

2) Среди всех имеющихся переменных отобрать наиболее существенно влияющих на объясняемую переменную факторов. Процедуры отбора будут рассмотрены ниже.

3) Переход к смещенным методам оценивания.

Когда мы сталкиваемся с проблемой мультиколлинеарности, то у неопытного исследователя поначалу возникает желание просто исключить лишние регрессоры, которые, возможно, служат ее причиной. Однако не всегда ясно, какие именно переменные являются лишними в указанном смысле. Кроме

того, как будет показано ниже, отбрасывание так называемых существенно влияющих переменных приводит к смещенности МНК-оценок.

9. ОШИБКИ СПЕЦИФИКАЦИИ

Построение экономической модели включает в себя спецификацию ее соотношений, выбор переменных, входящих в соотношение, определение математической функции, входящей в каждое соотношение. В данном пункте мы рассмотрим второй элемент.

Если точно известно, какая переменная должна быть включена в уравнение, то наша задача состоит в определении коэффициентов, построении доверительных интервалов, проверке различных гипотез. На практике мы никогда не можем быть уверены, что уравнение специфицировано правильно. Что случится, если мы включим в уравнение переменные, которых там быть не должно, и что случится, если мы не включим в уравнение переменные, которые там должны присутствовать. Свойства оценок коэффициентов в значительной степени зависят от правильности спецификации модели.

Ошибки спецификации бывают двух видов:

- 1) невключение в уравнение существенной объясняющей переменной;
- 2) включение в уравнение переменной, которая не должна там присутствовать.
- 3) неправильный выбор формы зависимости между переменными, мы предположили, что модель линейная, а она может быть более сложной.

1. Влияние отсутствия в уравнении переменной, которая должна быть включена.

Рассмотрим ситуацию для случая двух переменных.

Истинная модель выглядит следующим образом: $Y = \alpha + \beta X + \gamma Z + \varepsilon$. Но мы не уверены в значимости Z , поэтому оцениваем «короткую» модель: $Y = \alpha + \beta X + \varepsilon$. По методу наименьших квадратов вычисляем $\hat{\beta}$:

$$\widehat{\beta} = \frac{\frac{\sum_{i=1}^N X_i Y_i}{N} - \bar{X}\bar{Y}}{\frac{\sum_{i=1}^N X_i^2}{N} - \bar{X}^2}$$

$\widehat{\beta}$ - несмещенная оценка β , если $M\widehat{\beta} = \beta$. Посчитаем, чему равно $M\widehat{\beta}$:

$$M\widehat{\beta} = M \frac{\frac{\sum_{i=1}^N X_i Y_i}{N} - \bar{X}\bar{Y}}{\frac{\sum_{i=1}^N X_i^2}{N} - \bar{X}^2} = \frac{\frac{\sum_{i=1}^N X_i M Y_i}{N} - \bar{X} M \bar{Y}}{\frac{\sum_{i=1}^N X_i^2}{N} - \bar{X}^2}$$

$$M Y_i = M(\alpha + \beta X_i + \gamma Z_i + \varepsilon_i) = \alpha + \beta X_i + \gamma Z_i$$

$$M \bar{Y} = M \left(\frac{\sum_{i=1}^N Y_i}{N} \right) = \frac{\sum_{i=1}^N M Y_i}{N} = \frac{\sum_{i=1}^N (\alpha + \beta X_i + \gamma Z_i)}{N} = \alpha + \beta \bar{X} + \gamma \bar{Z}$$

Таким образом, получаем в числителе:

$$\frac{\sum_{i=1}^N X_i M Y_i}{N} - \bar{X} M \bar{Y} = \frac{\sum_{i=1}^N X_i (\alpha + \beta X_i + \gamma Z_i)}{N} - \bar{X} (\alpha + \beta \bar{X} + \gamma \bar{Z}) =$$

$$= \alpha \bar{X} + \beta \frac{\sum_{i=1}^N X_i^2}{N} + \gamma \frac{\sum_{i=1}^N X_i Z_i}{N} - \alpha \bar{X} - \beta \bar{X}^2 - \gamma \bar{X} \bar{Z} =$$

$$= \beta \sigma_X^2 + \gamma \left(\frac{\sum_{i=1}^N X_i Z_i}{N} - \bar{X} \bar{Z} \right)$$

$$\text{Итак, } M\widehat{\beta} = \beta + \gamma \frac{\frac{\sum_{i=1}^N X_i Z_i}{N} - \bar{X} \bar{Z}}{\sigma_X^2}.$$

Таким образом, мы получили смещенную оценку. Оценка будет несмещенной в двух случаях:

- 1) $\gamma = 0$;
- 2) X и Z статистически независимы.

Наша оценка будет завышать или занижать истинное значение коэффициента в зависимости от знака смещения.

Интуитивное объяснение.

Предположим, что β и γ положительны, а X и Z положительно коррелированы, тогда с увеличением X

- 1) Y будет иметь тенденцию к росту, поскольку β положителен;
- 2) Z будет иметь тенденцию к увеличению, поскольку X и Z положительно коррелированы;
- 3) Y получит дополнительное ускорение из-за увеличения Z , поскольку γ положительно.

Другими словами, изменение Y будет преувеличивать влияние текущих значений X , т. к. отчасти они будут связаны с изменениями Z . Т.е. часть изменения Y за счет изменения Z будет приписано X .

Однако смещение оценок коэффициентов здесь – не единственная неприятность. Что будет с оценками дисперсий?

$$D\hat{\beta}_k = \frac{\sigma^2}{\sum_{i=1}^N (X_i - \bar{X})^2} - \text{в короткой регрессии (без доказательства).}$$

$$D\hat{\beta}_d = \frac{\sigma^2}{\sum_{i=1}^N (X_i - \bar{X})^2} \frac{1}{1 - r_{xz}^2} - \text{в длинной регрессии (без доказательства).}$$

Таким образом, $D\hat{\beta}_k \leq D\hat{\beta}_d$, т. е. $\hat{\beta}$ - смещенная оценка, но обладает меньшей дисперсией.

Что будет с оценкой $\sigma^2 - s^2$? Оказывается, что в случае, если мы не включаем в регрессию существенную переменную, эта оценка будет смещенной. Поскольку s^2 участвует во многих статистических тестах, то используя их для проверки гипотез, мы можем получить ложные выводы.

Итак, в случае невключения объясняющих переменных, МНК-оценка короткой регрессии смещена, и обладает меньшей дисперсией, чем у оценки в

длинной регрессии. Оценка дисперсии ошибки имеет неотрицательное смещение.

2. Включение несущественных переменных.

Теперь у нас ситуация противоположная предыдущей. Истинная модель выглядит следующим образом: $Y = \alpha + \beta X + \varepsilon$, а мы оцениваем «длинную» регрессию $Y = \alpha + \beta X + \gamma Z + \varepsilon$. Таким образом, включая в уравнение несущественную переменную, мы не учитываем информацию о том, что коэффициент при Z равен нулю. Следует всегда ожидать, что неучитывание всей информации о модели потеря эффективности оценок. Т. е. в нашем случае дисперсия оценки в «длинной» регрессии будет больше, чем дисперсия оценки коэффициента при X в истинной модели, поскольку мы вынуждены по тем же самым наблюдениям оценивать два параметра вместо одного. Тем не менее, оценки «длинной» регрессии останутся несмещенными.

Потеря эффективности не случится, если переменные X и Z некоррелированы. Потеря эффективности приводит к тому, что мы с большей трудностью отвергаем гипотезу о незначимости коэффициента, тем не менее оценка дисперсии β останется несмещенной.

Выводы здесь мы приводить не будем. $\hat{\beta}$ и s^2 - несмещенные оценки, но ее дисперсия больше, чем в правильной модели, т. е. точность оценки ухудшается.

Рисунок с графиками плотностей распределения.

3. Неправильный выбор функциональной зависимости.

Еще одна ошибка спецификации происходит, когда исследователь решает оценить линейную модель, в то время как истинная регрессионная модель нелинейная. Пример: $Y = \alpha + \beta X + \gamma X^2 + \varepsilon$, а оцениваем мы модель

$Y = \alpha + \beta X + \varepsilon$. Приведенная выше ситуация является частным случаем ситуации с пропущенными переменными. Выбор линейной модели, в то время как истинная модель нелинейная может привести к смещенности и несостоятельности оценок регрессии. Поэтому исследователи часто используют полиномиальную регрессию как тест на нелинейность в объясняющих переменных.

Итак, мы с вами рассмотрели теоретические аспекты включения лишних или невключения нужных переменных в уравнение. Что же делать на практике, когда мы никогда точно не знаем, какие переменные входят в модель, а какие нет. В таких ситуациях используют различные эвристические процедуры отбора регрессоров.

10. ПРОЦЕДУРЫ ОТБОРА РЕГРЕССОРОВ

(отбор наиболее существенных объясняющих переменных).

В самом начале нашего курса мы разбирали вопрос, откуда возникает ошибка ε в i -м наблюдении. Мы тогда говорим про невключение в уравнение переменных в силу различных обстоятельств – про возможность перехода от исходного числа p анализируемых объясняющих переменных к существенно меньшему числу объясняющих переменных, наиболее информативных в некотором смысле. Некоторые объясняющие переменные оказывают несущественное влияние на объясняющую переменную и им можно пренебречь. Если же у нас есть сильно зависимые признаки, то информация, поставляемая ими, дублирует друг друга, так, что дополнительным влиянием одной из переменных можно пренебречь. Поэтому стремление исследователя отобрать из имеющегося у него набора объясняющих переменных лишь самые существенные (с точки зрения влияния на Y), представляется вполне естественным. В предположении, что объясняющие переменные неслучайны, возможны две точки зрения на оценку уравнения регрессии, получаемого после отбора наиболее существенных предсказывающих переменных:

1. Модель регрессии является истинной, тогда при помощи метода наименьших квадратов получается несмещенная и эффективная оценка коэффициентов регрессии (в условиях мультиколлинеарности эта оценка может быть неудовлетворительной, но, тем не менее, останется несмещенной). Тогда принудительное приравнивание части коэффициентов к нулю, что и происходит при отборе регрессоров, приводит, как мы убедились, к смещенным оценкам коэффициентов при оставшихся переменных, т. е. мы переходим к классу смещенных оценок, о чем говорилось выше.

2. Процесс отбора существенных переменных можно рассматривать как процесс выбора истинной модели из множества возможных линейных моделей, которые могут быть построены с помощью набора объясняющих переменных,

и тогда полученные после отбора оценки коэффициентов можно рассматривать как несмещенные. этой точки зрения мы и будем придерживаться в дальнейшем.

3. Для случая, когда объясняющие переменные – случайные величины, вопрос о правильности (истинности) модели не стоит. Все, что мы ищем в этом случае – модель, сохраняющую ошибку предсказания на разумном уровне при ограниченном количестве переменных.

Существует несколько подходов к решению задачи отбора наиболее существенных объясняющих переменных. Мы остановимся на двух процедурах, реализующих идею «от простого к сложному» – последовательного наращивания числа объясняющих переменных.

Пусть у нас всего p переменных, претендующих на участие в правой части.

1. «Все возможные регрессии».

1) Проведем p парных регрессий Y на X_1, \dots, X_p и выберем ту переменную, для которой коэффициент детерминации наибольший - R_1^2 . на этом шаге мы найдем одну объясняющую переменную, которую можно назвать наиболее информативной объясняющей переменной при условии, что в регрессионную модель мы можем включить только одну из имеющегося набора объясняющих переменных.

2) проведем $p*(p-1)$ регрессий, каждый раз включая две из p переменных и выберем ту, которая дает наибольшее значение R_2^2 – пара $(X^{(1)}, X^{(2)})$ – наиболее информативная пара переменных: эта пара будет иметь наиболее тесную статистическую связь с результирующим показателем Y . В состав этой пары переменная из первого шага может и не войти.

3) находим три наиболее информативных объясняющих переменных, проведя $p*(p-1)*(p-2) - R_3^2$

...

Вопрос – когда остановиться. Строгих правил нет, только рекомендации. Изобразим на графике зависимость скорректированного коэффициента детерминации наиболее информативной совокупности переменных от числа этих переменных. Одновременно будем откладывать следующую величину:

$$R_{\min}^2 = R_{adj}^2(k) - 2 \sqrt{\frac{2k(N-k-1)}{(N-1)(N^2-1)}} (1 - R^2(k)).$$

Предлагается выбрать в качестве оптимального числа объясняющих переменных то число, для которого R_{\min}^2 достигает своего максимума. Теоретическое обоснование этому мы здесь не приводим.

Однако реализация метода всех возможных регрессий требует значительных вычислительных трудностей, поскольку число регрессий, которые необходимо оценить, большое (равное $2^p - 2$, для $p = 20$ число возможных переборов будет больше миллиона (вспомнить байку про шахматы)). Есть несколько выходов из этой ситуации. Мы рассмотрим

II. Пошаговая процедура отбора переменных (в двух реализациях).

Здесь мы на каждом шаге учитываем результаты предыдущего шага, и в этом состоит отличие этого метода от предыдущего.

Первый шаг такой же, как и в предыдущем случае:

1) Среди имеющихся p переменных выбираем ту, для которой коэффициент корреляции с объясняемой переменной наибольший.

2) а) Теперь мы перебираем не все возможные пары переменных, а лишь те, в которых участвует переменная, полученная на первом шаге. Число переборов в этом случае существенно уменьшится

б) среди оставшихся переменных выбираем ту, которая имеет с объясняемой переменной наибольший коэффициент частной корреляции, очищенный от влияния переменной, полученной на первом шаге.

3)...

Число переборов для а) - $\frac{(p+2)(p+1)}{2}$, т. е. для $p = 20$ число переборов будет 209.

Опять остается вопрос – когда же остановится. Ответ может быть такой, например, когда новый коэффициент частной корреляции будет уже незначимо отличаться от нуля и др. Здесь так же можно сконструировать величину R_{\min}^2 и остановится тогда, когда она достигнет максимума.

Вообще говоря, пошаговые процедуры не гарантируют получения оптимального с точки зрения «всех пошаговых регрессий» набора, но в большинстве ситуаций, наборы переменных, получаемых методами пошагового отбора, будут близки к ним.

Кроме описанных, существуют различные методы пошаговые: другой метод пошагового присоединения, метод присоединения-удаления, метод удаления и др.

11. ГЕТЕРОСКЕДАСТИЧНОСТЬ

Как было сказано выше, гетероскедастичность – ситуация, когда нарушено пятое условие Гаусса-Маркова: ошибки для разных наблюдений имеют разную дисперсию ($D\varepsilon_i = \sigma_i^2$). Пример с фирмами, работающими в одной сфере. Естественно ожидать, что ошибки для больших фирм будут иметь большую дисперсию, чем ошибки маленьких фирм.

1. Последствия гетероскедастичности.

а. МНК-оценки, хотя останутся несмещенными, уже не будут являться эффективными, т. е. не будут обладать наименьшей дисперсией. Мы сможем построить оценки с меньшей дисперсией при помощи другого метода.

б. МНК-оценки стандартных ошибок будут неверны ($s_{\hat{\beta}_i}^2 = s_\varepsilon^2 a^{ii}$). Дело в том, что $D\hat{\beta}_i = \sigma_\varepsilon^2 a^{ii}$ только в случае выполнения условия Гаусса-Маркова. т. е. дисперсия в условиях гетероскедастичности будет другой, что же на самом деле оценивают $s_{\hat{\beta}_i}$ - ? В этом случае говорят, что МНК-оценки стандартных ошибок смещены. Скорее всего, они занижают истинное значение дисперсии. Теперь мы с вами уже не сможем для оценки гипотезы о значимости коэффициентов пользоваться t -статистиками., поскольку в них фигурируют неверные (смещенные) оценки стандартного отклонения оценки коэффициента, заниженное. Следовательно, t -статистики будут завышены. Это значит, что мы можем принять неверное решение о значимости коэффициента, хотя он будет незначим.

$$\text{Для случая парной модели } D\hat{\beta} = \frac{\sum_{i=1}^N (X_i - \bar{X})^2 \sigma_i^2}{\left(\sum_{i=1}^N (X_i - \bar{X})^2 \right)^2}.$$

Интуиция неэффективности. Наблюдение, дисперсия ошибки которого будет меньше, обычно будет находиться ближе к линии регрессии, поэтому будет служить хорошим ориентиром, указывающим место этой линии. Наблюдение же, которое имеет большую дисперсию, будет обычно находиться дальше от линии и не сможет существенно помочь в определении местоположения этих линий.

2. Тесты на гетероскедастичность.

Мы по-прежнему рассматриваем модель $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$

Нулевая гипотеза – отсутствие в модели гетероскедастичности, т. е. гомоскедастичность:

Вид альтернативной гипотезы специфичен для каждого теста, т. е. зависит от теста, при помощи которого мы проверяем наличие гетероскедастичности. Не смотря на то, что таких тестов существует несколько, все они базируются на одном: анализе квадратов остатков исходной регрессии. Поскольку остатки регрессии снабжают нас информацией об ошибках регрессии, мы можем проанализировать остатки для того, чтобы посмотреть, отличается ли разброс остатков (вокруг нуля) или разброс наблюдений вдоль линии регрессии от наблюдения к наблюдению разброс остатков вокруг нуля будет отражать разброс ошибок вокруг нуля. Эти рассуждения должны навести нас на мысль, что ситуацию гетероскедастичности можно отследить графически. Если наши данные представляют собой временной ряд, то отсортировав остатки или квадраты остатков по времени и изобразив их на графике мы можем заметить, что остатки растут во времени. Если же мы анализируем пространственные данные, изобразив остатки на графике в зависимости от одной из объясняющих переменных, можно заметить разницу в разбросе остатков.

Это эвристический, опытный подход. Теперь приведем несколько формальных тестов. Все тесты предполагают, что дисперсии ошибок наблюдений зависят от некоторой переменной, которая может входить в модель, а может и не входить.

1. Тест Голфилда–Квандта.
2. Тест ранговой корреляции Спирмена.
3. Тест Бреуш–Пагана.
4. Тест Уайта.

Исходная модель: $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$

$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_N^2$. Предположим, что нулевая гипотеза неверна и в нашей модели присутствует гетероскедастичность и $D\varepsilon_i = \sigma_i^2$. Предположим также, что $\sigma_i^2 = f(Z_i)$, где Z_i – может быть одной из объясняющих переменных, группой объясняющих переменных, или вообще переменной, не участвующей в модели. Форма $f(Z)$ может быть различной – линейной, логарифмической, квадратичной.

Тест Уайта заключается в следующем:

- 1) Оцениваем имеющуюся модель и получаем величины остатков e_i
- 2) Осуществляем регрессию

$$e_i^2 = \gamma_0 + \gamma_1 X_{1i} + \dots + \gamma_k X_{ki} + \gamma_{k+1} X_{1i}^2 + \dots + \gamma_{2k} X_{ki}^2 + \gamma_{2k+1} X_{1i} X_{2i} + \dots + u_i$$

В этой регрессии мы учитываем больше форм зависимостей σ_i^2 от независимых переменных. Если нулевая гипотеза справедлива и σ_i^2 не зависит никак ни от одной из независимых переменных, то наша регрессия практически ничего не объясняет, следовательно, ее R^2 мал. Если же есть гетероскедастичность, то R^2 «большой». Границы «малости»: при справедливости нулевой гипотезы статистика NR^2 имеет распределение «хи-квадрат» с числом степеней свободы q , где q – число переменных в регрессии пункта 2 вместе со свободным членом.

Пример

2. Коррекция на гетероскедастичность.

Задача – уточнить оценки коэффициентов и исправить стандартные ошибки, чтобы модно было пользоваться тестами для проверки гипотез.

Предположим ненадолго, что мы знаем величины ошибок σ_i^2 . Тогда поделим обе части уравнения нашей модели на σ_i :

$$\frac{Y_i}{\sigma_i} = \beta_0 \frac{1}{\sigma_i} + \beta_1 \frac{X_{1i}}{\sigma_i} + \dots + \beta_k \frac{X_{ki}}{\sigma_i} + v_i, \text{ где } v_i = \frac{\varepsilon_i}{\sigma_i}.$$

Но σ_i^2 мы никогда не знаем.

Гетероскедастичность	
ничего не знаем о σ_i^2	есть априорная информация о σ_i^2
Можем исправить стандартные ошибки, чтобы можно было использовать статистические тесты для проверки гипотез относительно коэффициентов – стандартные ошибки в форме Уайта или Невье-Веста	можем уточнить оценки коэффициентов уравнения – двухшаговая процедура коррекции на гетероскедастичность.

1. Стандартные ошибки в форме Уайта (White Standart Errors) – состоятельные оценки стандартных отклонений оценок коэффициентов регрессионного уравнения.

Для случая парной модели:

$$D\hat{\beta} = \frac{\sum_{i=1}^N (X_i - \bar{X})^2 \sigma_i^2}{\left(\sum_{i=1}^N (X_i - \bar{X})^2\right)^2}, \quad D\hat{\beta} = \frac{\sum_{i=1}^N (X_i - \bar{X})^2 e_i^2}{\left(\sum_{i=1}^N (X_i - \bar{X})^2\right)^2}.$$

Стандартные ошибки в форме Уайта можно получить практически во всех статистических пакетах, в том числе и в Eviews-е.

2. Процедура коррекции на гетероскедастичность.

Пусть у нас есть основания предполагать, что значения дисперсий ошибок в i -м наблюдении пропорционально значениям некоторой объясняющей переменной (пусть, для определенности, X_1), т. е. $\sigma_i^2 \sim X_{1i}^2$ или $\sigma_i^2 = CX_{1i}^2$

Тогда мы можем сделать следующее: поделим обе части уравнения нашей модели на X_{1i} :

$$\frac{Y_i}{X_{1i}} = \beta_0 \frac{1}{X_{1i}} + \beta_1 \frac{X_{1i}}{X_{1i}} + \dots + \beta_k \frac{X_{ki}}{X_{1i}} + v_i, \text{ где } v_i = \frac{\varepsilon_i}{X_{1i}}.$$

Упражнение. Показать, что дисперсия v_i не зависит от номера наблюдения.

Если же дисперсия ошибок зависит от значений нескольких переменных и форма этой зависимости не обязательно линейная (логарифмическая, например), то проводим двухшаговую процедуру коррекции на гетероскедастичность:

- 1) Оцениваем исходную модель (*) МНК, получаем остатки e_i
- 2) Оцениваем следующую регрессию:

$$e_i^2 = \gamma_0 + \gamma_1 X_{1i} + \gamma_2 X_{2i} + \dots + \gamma_k X_{ki} + u_i, \text{ получаем } \hat{e}_i$$

- 3) Оцениваем взвешенную регрессию:

$$\frac{Y_i}{\hat{e}_i} = \beta_0 \frac{1}{\hat{e}_i} + \beta_1 \frac{X_{1i}}{\hat{e}_i} + \dots + \beta_k \frac{X_{ki}}{\hat{e}_i} + v_i$$

- 4) проверяем на гетероскедастичность, если нет, то ОК, если не удалось – возвращаемся к шагу 2 и придумываем другие формы зависимости (добавляем квадраты, перекрестные члены и др.).

12. АВТОКОРРЕЛЯЦИЯ ВРЕМЕННЫХ РЯДОВ.

Во временных рядах практически всегда наблюдается явление, называемое автокорреляцией. Автокорреляция представляет собой корреляционную зависимость между последующими и предшествующими членами временного ряда, т. е. корреляцию между рядами Y_1, Y_2, \dots, Y_T и $Y_L, Y_{L+1}, \dots, Y_{T+L}$, где L – длина временного смещения. L зависит от наибольшего числа периодов во временном ряду.

Автокоррелированными могут оказаться остатки регрессионных моделей, построенных на базе временных рядов.

Рассмотрим множественную линейную регрессионную модель

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon, \quad (2.1)$$

в которой присутствует автокорреляция ошибок:

$$M(\varepsilon_i \varepsilon_j) \neq 0 \text{ при } i \neq k.$$

Например, ошибки могут подчиняться автокорреляционному процессу первого порядка:

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t \quad (2.2),$$

тогда параметр ρ называют параметром автокорреляции.

Наличие автокорреляции может быть обусловлено следующими причинами:

1. Если в модели не учтен некоторый существенный фактор, то его влияние может быть отражено в остатках, вследствие чего последние могут оказаться автокоррелированными.

2. Выбран неправильный тип модели.

3. Специфическая структура случайных остатков.

Последствия автокорреляции.

1. МНК-оценки коэффициентов модели остаются смещенными и состоятельными, но перестают быть эффективными, т. е. мы можем построить оценки с меньшими дисперсиями.

2. МНК-оценка дисперсии оценок коэффициентов смещены и несостоятельны, они занижают истинное значение дисперсии.

3. В модели с лаговой зависимой переменной ...

Обнаружение автокорреляции.

Итак, игнорирование автокорреляции регрессионных остатков создает серьезные трудности для применения обыкновенного МНК. Поэтому важно владеть методами, позволяющими устанавливать ее присутствие. Большинство тестов на автокорреляцию используют следующую идею: если корреляция есть у ошибок, то она присутствует и в остатках, получаемых после применения к (2.1) обычного метода наименьших квадратов. Мы рассмотрим только одну реализацию этой процедуры, а именно, тест на наличие в модели автокорреляции первого порядка (тест Дарбина-Уотсона).

$$H_0 : \rho = 0.$$

В качестве альтернативной могут выступать различные гипотезы.

Критическая статистика Дёрбина-Уотсона имеет вид

$$DW = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2}$$

Если постоянный член включен в число регрессоров, то нетрудно показать (Магнус, Катышев, Пересецкий), что статистика Дарбина-Уотсона тесно связана с выборочным коэффициентом корреляции между e_t и e_{t-1} (r):

$$DW \approx 2(1 - r)$$

Содержательный смысл статистики Дарбина-Уотсона следующий: если между e_t и e_{t-1} имеется достаточно высокая положительная автокорреляция, то в определенном смысле e_t и e_{t-1} близки друг к другу и величина статистики DW мала. Это согласуется с формулой (2.4).

Поскольку, как оказалось, распределение статистики DW (в предположении справедливости гипотезы H_0) зависит от наблюдаемых значений объясняющих переменных X , Дарбину и Уотсону удалось установить (для двух заданных величин уровня значимости критерия $\alpha = 0,05$ и $\alpha = 0,01$) лишь такие пороговые значения DW_u и DW_l , которые позволяют построить следующие два варианта процедуры проверки гипотезы (в зависимости от альтернативы о наличии в остатках положительной или отрицательной автокорреляции 1-го порядка):

а) При $DW < 2$ (альтернатива: существование в остатках положительной автокорреляции первого порядка):

- по заданному α находим из таблиц пороговые значения DW_u и DW_l ;
- по формуле подсчитываем значение критической статистики DW ;
- если $DW < DW_l$, то гипотеза H_0 отвергается (с вероятностью ошибиться, равной α) в пользу гипотезы о положительной автокорреляции;
- если $DW > DW_u$, то гипотеза H_0 не отвергается;
- если $DW_l < DW < DW_u$, то сделать определенный вывод по имеющимся исходным данным нельзя (зона неопределенности).

б) При $DW > 2$ (альтернатива: существование в остатках отрицательной автокорреляции первого порядка).

- первые два действия — те же, что и в п. а);
- если $4 - DW_l < DW < 4$, то гипотеза H_0 отвергается (с вероятностью ошибиться, равной α) в пользу гипотезы об отрицательной автокорреляции,
- если $DW_u < DW < 4 - DW_u$, то гипотеза H_0 не отвергается;
- если $4 - DW_u < DW < 4 - DW_l$, то сделать определенный вывод по имеющимся исходным данным нельзя.

Замечание. Тест Дарбина-Уотсона построен в предположении о том, что регрессоры и ошибки некоррелированы. Поэтому, этот тест нельзя применять, например, в случае, когда среди регрессоров содержатся лагированные значения зависимой переменной.

Оценивание в модели с авторегрессией.

В ситуациях, когда значение коэффициента корреляции ρ между соседними по времени регрессионными остатками известно, исследователь не должен испытывать затруднений в практической реализации основных формул ОМНК. В этом случае матрица Ω выглядит следующим образом:

$$\Omega = \sigma^2 \begin{pmatrix} \mathbf{1} & \rho & \dots & \rho^{T-1} \\ \rho & \mathbf{1} & \dots & \rho^{T-2} \\ & & \dots & \\ \rho^{T-1} & \rho^{T-2} & \dots & \mathbf{1} \end{pmatrix}.$$

Взвешенный МНК выглядит следующим образом:

$$HY = HX\beta + H\varepsilon$$

где

$$H = \frac{1}{\sqrt{1-\rho^2}} \begin{pmatrix} 1-\rho^2 & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 \\ & & & \dots & & \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{pmatrix}$$

Поэтому остановимся на ситуации (гораздо более реалистичной), когда значение параметра ρ априори неизвестно исследователю. Практически все процедуры, предложенные для реализации доступного ОМНК в модели регрессии с автокоррелированными остатками при неизвестном значении ρ , имеют итерационный характер. Приведем здесь описание одной из наиболее распространенных процедур подобного типа, известной в литературе под названием процедуры Кохрейна-Оркатта:

1) вычисляются обычные МНК-оценки модели (2.1);

2) подсчитываются невязки 1-й итерации $e_i = Y_i - \hat{Y}_i$

3) первое приближение $\hat{\rho}$ - оценки неизвестного параметра ρ определяется в качестве МНК-оценки коэффициента регрессии ρ в модели $e_t = \rho e_{t-1} + u_t$, где остатки удовлетворяют условиям классической модели;

4) вычисляются ДОМНК-оценки по формуле (2.5) с матрицей H определенной соотношением (2.6), в котором вместо ρ подставлены значения $\hat{\rho}$ из первой итерации;

5) подсчитываются невязки 2-й итерации $e_i = Y_i - \hat{Y}_i$

6) процедура повторяется, начиная с пункта 3.

и т. д.

Процедуру заканчивают при стабилизации получаемых значений, т.е. на стадии, когда очередное приближение ρ мало отличается от предыдущего. «Тонкое место» метода определяется типовым недостатком подобных процедур, заключающимся в возможности «скатиться» в ходе итераций в локальный, а не глобальный минимум критерия наименьших квадратов. В этом случае значение параметра ρ может быть определено с большой ошибкой. Чтобы этого избежать, используются «решетчатые» процедуры, например, процедура Хилдрета-Лу.

Оценка матрицы ковариаций оценок коэффициентов МЛРМ.

Как уже было сказано выше, МНК-оценки матрицы ковариаций оценок коэффициентов модели смещены и несостоятельны, поэтому, как и в случае гетероскедастичности, необходима коррекция стандартных ошибок в присутствии автокорреляции. Такая оценка есть – Матрица ковариаций в форме Навье-Веста (Newey-West). Предположим, что в матрице ковариаций ошибок Ω ненулевые элементы стоят не только на главной диагонали, но и на соседних

диагоналях, отстоящих от главной не более чем на L (т. е. $w_{ij} = \mathbf{0}$, если $|i - j| > L$).

Невье и Вест показали в 1987 г., что оценка

$$\widehat{\Omega} = T(X'X)^{-1} \left[\frac{\mathbf{1}}{T} \sum_{s=1}^T e_s^2 x_s x_s' + \frac{\mathbf{1}}{T} \sum_{L=1}^L \sum_{t=j+1}^T w_j e_t e_{t-j} (x_t x_{t-j}' + x_{t-j} x_t') \right] (X'X)^{-1}$$

Существует несколько способов выбора весовых коэффициентов, например, весовые коэффициенты Бартлетта и Парзена.

Весовые коэффициенты Бартлетта:

$$w_j = \frac{\mathbf{1}}{L - \mathbf{1}}$$

13. ОБОБЩЕННЫЙ МЕТОД НАИМЕНЬШИХ КВАДРАТОВ.

Одно из предположений классической регрессионной модели гласит о том, что ошибки имеют одинаковую дисперсию и некоррелированы друг с другом. Это предположение во многих ситуациях нереалистично. При анализе временных рядов в очень редких ситуациях можно предполагать, что ошибки некоррелированы для разных периодов времени. Кроме того, в некоторых ситуациях возникает пространственная автокорреляция, если мы рассматриваем данные по регионам России (региональные данные) или анализируем поведение индивидуумов, связанных между собой, например, родственными или дружескими отношениями. Гетероскедастичность может возникнуть, например, если наши данные в некотором смысле неоднородны (исследуем зависимость расходов на питание у семей с различным уровнем доходов). Поэтому, естественно изучать модели без предположения выполнения таких условий Гаусса-Маркова.

Рассмотрим модель

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$$

или в матричной форме

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

где $M(\varepsilon_i) = 0$, $V(\varepsilon) = \Omega$ - матрица ковариаций ε :

$$\Omega = \begin{pmatrix} M(\varepsilon_1^2) & M(\varepsilon_1\varepsilon_2) & \dots & M(\varepsilon_1\varepsilon_N) \\ M(\varepsilon_2\varepsilon_1) & M(\varepsilon_2^2) & \dots & M(\varepsilon_2\varepsilon_N) \\ \dots & \dots & \dots & \dots \\ M(\varepsilon_N\varepsilon_1) & M(\varepsilon_N\varepsilon_2) & \dots & M(\varepsilon_N^2) \end{pmatrix} = M(\varepsilon\varepsilon').$$

Кроме того, матрица Ω предполагается положительно определенной, т. е. $\forall x \quad x'\Omega x > 0$.

В классической регрессионной модели матрица ковариаций имеет следующий вид:

$$\Omega = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix}$$

В модели с гетероскедастичностью:

$$\Omega = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & \sigma_N^2 \end{pmatrix}$$

В модели с автокорреляцией первого порядка ($\varepsilon_t = \rho\varepsilon_{t-1} + u_t$):

$$\Omega = \sigma^2 \begin{pmatrix} 1 & \rho & \dots & \rho^{N-1} \\ \rho & 1 & \dots & \rho^{N-2} \\ & & \dots & \\ \rho^{N-1} & \rho^{N-2} & \dots & 1 \end{pmatrix}$$

Мы можем оценить обобщенную модель обычным методом наименьших квадратов. В этом случае, как мы помним, МНК-оценки останутся несмещенными, однако, МНК-оценки дисперсий оценок коэффициентов будут являться смещенными. Напомним, что МНК-оценка матрицы ковариаций вектора $\hat{\beta}$ выражается следующим образом:

$$\hat{V}(\hat{\beta}) = \frac{\sum_{i=1}^N e_i^2}{N - k} (X'X)^{-1}.$$

Оценка $\hat{\beta}$ Хотя останется состоятельной (без доказательства), она уже не будет оптимальной в смысле теоремы Гаусса-Маркова. МНК-оценки коэффициентов уже не будут эффективными, т. е. обладать наименьшей дисперсией из всех возможных линейных несмещенных оценок. Иными словами, мы можем придумать другую линейную несмещенную оценку, дисперсия которой будет меньше, чем дисперсия МНК-оценок.

Такую оценку и строят при помощи обобщенного метода наименьших квадратов (ОМНК). Мы повышаем эффективность оценок за счет

дополнительной информации о матрице Ω (обычный метод наименьших квадратов эту информацию не учитывает).

Попытаемся такую оценку построить.

Задача оценивания может быть решена различными эквивалентными способами, из которых мы выбрали простейший. Поскольку матрица Ω положительно определена по условию, мы можем воспользоваться базовой теоремой алгебры матриц, которая гласит, что для любой положительно определенной матрицы Ω существует невырожденная матрица H такая, что

$$H\Omega H' = I$$

Перепишем равенство (3.2) следующим образом:

$$\Omega = H^{-1}(H')^{-1}$$

Откуда

$$\Omega^{-1} = H' H .$$

Эта матрица H нам понадобится в дальнейшем, для того, чтобы преобразовать нашу исходную модель:

$$HY = HX\beta + H\varepsilon$$

или

$$\tilde{Y} = \tilde{X}\beta + \tilde{\varepsilon}$$

Теперь найдем матрицу ковариаций нового случайного члена:

$$M(\tilde{\varepsilon}\tilde{\varepsilon}') = M(H\varepsilon\varepsilon'H') = HM(\varepsilon\varepsilon')H' = H\Omega H' = I .$$

Итак, матрица ковариаций преобразованной модели (3.3) удовлетворяет условиям Гаусса-Маркова. Поэтому, для оценки этой модели можно применять обычный метод наименьших квадратов:

$$\hat{\beta}_{ОМНК} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y}$$

эта оценка является несмещенной и эффективной, согласно теореме Гаусса-Маркова. В терминах начальной модели наша оценка по обобщенному методу наименьших квадратов будет выглядеть следующим образом:

$$\begin{aligned} \hat{\beta}_{ОМНК} &= (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y} = ((HX)'HX)^{-1}(HX)'HY = (X'H'HX)^{-1}X'H'HY = \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y \end{aligned}$$

Матрица ковариаций вектора $\hat{\beta}_{ОМНК}$ выражается следующим образом:

$$V(\hat{\beta}_{ОМНК}) = (X' \Omega^{-1} X)^{-1}$$

Теорема Айткена. В классе линейных несмещенных оценок коэффициентов уравнения (3.1) оценка $\hat{\beta}_{ОМНК} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} Y$ является эффективной, т. е. обладает наименьшей матрицей ковариаций (не совсем корректно говорить дисперсией).

Как можно видеть из (3.4), результаты ОМНК прекрасно согласуются с результатами МНК: если $\Omega = \sigma^2 I$, то МНК-оценка совпадает с ОМНК-оценкой, таким образом, МНК – частный случай ОМНК для особого вида матрицы Ω .

Однако, для построения ОМНК-оценки нам необходимо знать матрицу Ω . А ее мы никогда не знаем, поскольку не знаем ε . Поэтому матрицу Ω тоже надо оценить. Поскольку в этой матрице всего $N(N+1)/2$ элементов, то нет никакой надежды получить приемлемые (состоятельные) оценки, имея всего N наблюдений. Поэтому, для получения состоятельной оценки матрицы Ω приходится накладывать некоторые ограничения на ее структуру.

Пусть V – состоятельная оценка матрицы Ω , тогда, подставляя ее в (3.5) и (3.6), осуществляя таким образом доступный обобщенный метод наименьших квадратов (feasible GLS), получаем оценки по доступному методу наименьших квадратов:

$$\hat{\beta}_{ДОМНК} = (X' V^{-1} X)^{-1} X' V^{-1} Y$$

$$V(\hat{\beta}_{ДОМНК}) = (X' V^{-1} X)^{-1}.$$

Что со свойствами этих оценок – теряют свойства несмещенности, но остается состоятельность. Так что предпочесть – смещенность, но меньшую дисперсию или несмещенность, но неэффективность. Ответ, по-видимому, такой: на малых выборках ДОМНК оценки ведут себя непредсказуемым образом, поэтому в некоторых ситуациях лучше использовать МНК-оценки (в этом случае МНК-оценки матрицы ковариаций $\hat{\beta}$ по-прежнему плохая), на

больших же – ДОМНК (состоятельность, следовательно, смещение элиминируется с ростом выборки).

Проверять гипотезы можно, непосредственно используя $V(\hat{\beta}_{\text{ДОМНК}}) = (X'V^{-1}X)^{-1}$, либо при помощи вспомогательной регрессии

Для обобщенной регрессионной модели, в отличие от классической, уже нельзя использовать R^2 в качестве удовлетворительной мерой качества подгонки. Он не обязательно лежит в интервале $[0;1]$, а добавление или удаление регрессоров не обязательно влечет за собой его увеличение или уменьшение. Так же нет смысла рассматривать коэффициент детерминации и для вспомогательной регрессии, поскольку

- 1) среди преобразованных регрессоров уже может и не быть константы;
- 2) в общем случае трудно установить связь между качеством подгонки вспомогательной регрессии и исходной модели.

Прежде чем закончить разговор об ОМНК, рассмотрим преобразование матрицы Ω в (3.2), т. е. приведем матрицу H для случая гетероскедастичности и автокорреляции первого порядка остатков.

Для случая гетероскедастичности:

$$H = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{\sigma_N} \end{pmatrix}$$

Преобразование данных в соответствии с (3.3) эквивалентно взвешенному методу наименьших квадратов.

Для случая автокорреляции первого порядка случайного члена:

$$H = \frac{1}{\sqrt{1-\rho^2}} \begin{pmatrix} 1-\rho^2 & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 \\ & & & \dots & & \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{pmatrix}$$

14. Прогнозирование при помощи регрессионных моделей.

Пусть имеется выборка $(Y_i, X_{1i}, \dots, X_{ki})$, $i = \overline{1, N}$, на основании которой мы при помощи МНК построили выборочное регрессионное уравнение $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \dots + \hat{\beta}_k X_{ki}$.

Теперь по полученному уравнению мы можем построить прогнозные значения переменной Y для наблюдений, не входящих в выборку.

Задача. Построить прогноз $\hat{Y}(X_{1N+1}, \dots, X_{kN+1})$.

$\hat{Y}(X_{1N+1}, \dots, X_{kN+1}) = \hat{\beta}_0 + \hat{\beta}_1 X_{1N+1} + \dots + \hat{\beta}_k X_{kN+1}$ - точечный прогноз.

???

Прогноз чего

Напомним, что линия регрессии описывает изменение математического ожидания Y при изменении X , т. е. Поведение условного математического ожидания $M_Y(X) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$, поэтому $\hat{Y}(X_1, \dots, X_k)$ будет являться прогнозным значением (оценкой) условного математического ожидания (среднего значения) Y для данных X : $\hat{Y}(X_{1N+1}, \dots, X_{kN+1}) = \hat{M}_Y(X_{1N+1}, \dots, X_{kN+1})$.

Помимо точечного прогноза $\hat{Y}(X_{1N+1}, \dots, X_{kN+1})$ условного математического ожидания Y , строят так же доверительные интервалы для:

- а) функции регрессии, т. е. УМО Y ,
- б) индивидуальных значений переменной Y для данных X , т.е. прогноз самого Y .

Доверительный интервал – интервал, который с некоторой вероятностью (уровнем надежности) покрывает истинное значение исследуемого параметра:

- а) $M_Y(X)$;
- б) $Y(X)$.

Проиллюстрируем построение доверительного интервала на парной модели, после чего приведем результат для общего случая.

а) для построения доверительного интервала нужна дисперсия $\hat{Y}(X)$. Найдем ее:

$$D\hat{Y}(X) = D(\hat{\alpha} + \hat{\beta}X) = D(\bar{Y} - \hat{\beta}\bar{X} + \hat{\beta}X) = D(\bar{Y} + \hat{\beta}(X - \bar{X})) =$$

:

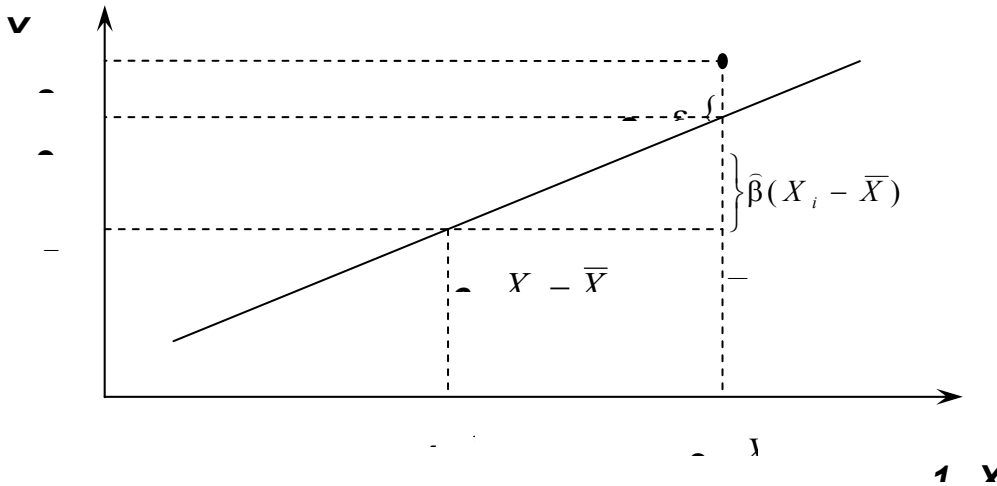


Рис 1. Графическая иллюстрация

без доказательства \bar{Y} и $\hat{\beta}(X - \bar{X})$ независимы, тогда

$$= D\bar{Y} + (X - \bar{X})^2 D\hat{\beta} =$$

X – неслучайная переменная, поэтому выносится из под знака дисперсии в квадрате

$$\left. \begin{aligned} D\bar{Y} &= D\frac{\sum Y_i}{N} = D\left(\alpha + \beta\bar{X} + \frac{\sum \varepsilon_i}{N}\right) = \frac{\sigma^2}{N} \\ D\hat{\beta} &= \frac{\sigma^2}{\sum (X_i - \bar{X})^2} \end{aligned} \right\} \text{- для случая гомоскедастичности,}$$

$$\left. \begin{aligned} D\bar{Y} &= \frac{\sum \sigma_i^2}{N^2} \\ D\hat{\beta} &= \frac{\sum \sigma_i^2 (X_i - \bar{X})^2}{\left(\sum (X_i - \bar{X})^2\right)^2} \end{aligned} \right\} \text{- для случая гетероскедастичности.}$$

Можно показать, что в условиях КМЛРМ можно показать, что статистика $t = \frac{\hat{Y} - M_Y(X)}{s_{\hat{Y}}}$ имеет распределение Стьюдента (t – распределение) с числом степеней свободы $N - 2$. Тогда доверительный интервал для УМО $M_Y(X)$ с уровнем надежности γ будет рассчитываться по следующей формуле:

$$(\hat{Y} - s_{\hat{Y}} t_{кр}^{\delta\gamma}(1 - \gamma, N - 2); \hat{Y} + s_{\hat{Y}} t_{кр}^{\delta\gamma}(1 - \gamma, N - 2))$$

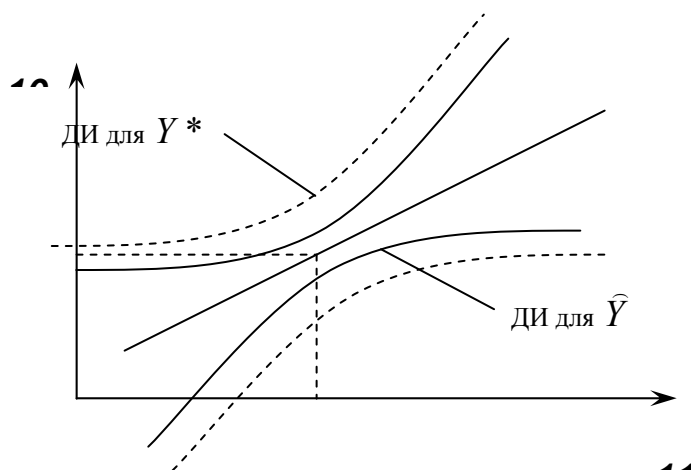


Рис. 2. Доверительные интервалы.

Из приведенных формул видно, что ширина доверительного интервала зависит от значений объясняющей переменной X : чем ближе X к \bar{X} , тем она меньше и минимальна для $X = \bar{X}$. Таким образом, построение прогноза значений зависимой переменной оправдано, если значение объясняющей переменной X не выходит за диапазон значений по выборке. Причем прогноз тем более точный, чем ближе X к \bar{X} .

Для МЛРМ доверительный интервал для функции регрессии или для условного математического ожидания зависимой переменной $M_Y(X)$ в предположении, что объясняющие переменные приняли значения, задаваемые вектором $X_{N+1}(1, X_{1N+1}, \dots, X_{kN+1})$ выглядит следующим образом:

$$(\hat{Y} - s_{\hat{Y}} t_{кр}^{\delta\gamma}(1 - \gamma, N - k - 1); \hat{Y} + s_{\hat{Y}} t_{кр}^{\delta\gamma}(1 - \gamma, N - k - 1)),$$

где $s_{\hat{Y}}^2 = s^2 X_{N+1} (X' X)^{-1} X'_{N+1}$ - квадрат стандартной ошибки групповой средней \hat{Y} .

б) при построении доверительного интервала для индивидуальных значений Y - Y_{N+1}^* необходимо учитывать еще один источник вариации переменной Y - рассеяние вокруг линии регрессии. В оценку суммарной дисперсии следует включить величину s^2 . Таким образом, оценка дисперсии индивидуальных значений Y будет рассчитываться по следующей формуле:

$$s_{Y^*}^2 = s^2 \left(1 + \frac{1}{N} + \frac{(X_{N+1} - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right)$$

а соответствующий доверительный интервал будет строиться по следующей формуле:

$$(\hat{Y}_{N+1} - s_{Y^*} t_{kp}^{\alpha} (1 - \gamma, N - 2); \hat{Y}_{N+1} + s_{Y^*} t_{kp}^{\alpha} (1 - \gamma, N - 2))$$

Для МЛРМ:

15. Временные ряды

Ряд называется строго стационарным, если совместное распределение вероятностей Y_{t_1}, \dots, Y_{t_m} не зависит от сдвига по времени, т. е. совпадает с распределением вероятностей $Y_{t_1+L}, \dots, Y_{t_m+L}$ для любых L, t_1, \dots, t_m . Обычно нас интересует не все распределение, а только дисперсии, средние значения и ковариации. Ряд называется слабо стационарным или просто стационарным, если средние, дисперсии и ковариации не зависят от времени t . Таким образом, для стационарного ряда $MY(t) = \mu$, $DY(t) = v_0$, $Cov(Y_t, Y_{t+L}) = v_L$, $ACF(t, L) = \varphi(L)$. Заметим, что из строгой стационарности следует слабая стационарность.

Введем понятие автокорреляционной функции:

$$ACF(L) = \frac{Cov(Y_t, Y_{t+L})}{DY_t} = \frac{v_L}{v_0} = \varphi_L, L = 1, 2, \dots$$

Выборочная автокорреляционная функция временного ряда называется коррелограммой и определяется следующим образом:

$$ACF(L) = \frac{\sum_{t=L+1}^T (Y_t - \bar{Y})(Y_{t-L} - \bar{Y})}{\sum_{t=1}^T (Y_t - \bar{Y})^2} \quad L = 1, 2, \dots$$

Автоковариационная функция – математическое ожидание произведений отклонений уровней ряда, сдвинутых на период L . Автокорреляционная функция тесноту связи между уровнями временного ряда.

Кроме автокорреляционной функции рассматривают еще частную автокорреляционную функцию $PACF_Y(L)$. Содержательно частная автокорреляционная функция представляет собой «чистую корреляцию между Y_t и Y_{t+L} при исключении влияний промежуточных значений $Y_{t+1}, \dots, Y_{t+L-1}$. Формула для ее записи достаточно сложна, поэтому здесь мы ее не приводим.

Примеры временных рядов.

1) «белый шум» – все уровни временного ряда распределены одинаково.

$$Y_t = \varepsilon_t, \text{ где } \varepsilon_t \sim iid(0, \sigma^2) \quad MY_t = 0, \quad DY_t = \sigma^2, \quad ACF(L) = \begin{cases} 1, L = 0; \\ 0, L > 0 \end{cases}.$$

2) авторегрессионный ряд первого порядка (AR(1))

$$Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma^2) \text{ и } |\rho| < 1.$$

$$MY_t = \alpha + \rho MY_{t-1} = \alpha + \rho\alpha + \rho^2 MY_{t-2} = \alpha(1 + \rho + \rho^2 + \dots) = \frac{\alpha}{1 - \rho}$$

Таким образом, если $|\rho| < 1$, средние не зависят от времени.

$$DY_t = \frac{\sigma^2}{1 - \rho^2} - \text{показать.}$$

$$v_L = \frac{\rho^L \sigma^2}{1 - \rho^2} - \text{можно показать. Т. о. при } |\rho| < 1 \text{ ряд стационарен.}$$

4) «случайное блуждание»

$$Y_t = Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma^2).$$

Предыдущий случай, когда $\rho = 1$.

Пример случайного блуждания (пленка).

$$MY_t = MY_{t-1}, \quad DY_t = DY_{t-1} + \sigma^2, \quad DY_1 = \sigma^2, \text{ таким образом, } DY_t = t\sigma^2, \text{ т. е.}$$

дисперсия ряда неограниченно возрастает со временем. Ряд не является стационарным.

Мнимая регрессия.

Предположим, что мы с вами рассматриваем два временных ряда, являющихся «случайными блужданиями»:

$$X_t = X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma^2) \text{ и}$$

$$Y_t = Y_{t-1} + e_t, \quad e_t \sim iid(0, \sigma^2).$$

ε и e независимы. Следовательно, независимы X и Y . Однако, осуществляя регрессию Y на X : $Y = \alpha + \beta X + u$. Проверяя значимость регрессии и коэффициентов стандартными методами, при помощи коэффициента детерминации и t-статистик мы можем сделать ложный вывод о

наличии зависимости между переменными. Это происходит потому, что ряд, полученный из остатков регрессии, в общем случае будет являться нестационарным, следовательно, его дисперсия будет зависеть от номера наблюдения и неограниченно возрастать со временем, не будет удовлетворять классическим условиям регрессионной модели. Как было показано, в этом случае t -статистика расходится при $T \rightarrow \infty$, следовательно, чем больше выборка, тем больше шансов прийти к ложному заключению. На практике признаками мнимой регрессии являются высокое значение коэффициента детерминации и низкое значение статистики Дарбина-Уотсона. Однако, в случае если мы делаем регрессию одного нестационарного ряда на другой, все может быть не так уж и плохо. Случай коинтеграции временных рядов рассмотрим ниже.

Прогнозирование случайных блужданий

Отличие рядов 2) и ряда 3) «случайное блуждание» в том, что в первом ряду влияние возмущений затухает со временем, во втором же ряду – нет.

Иллюстрация

5) ряд с трендом, например, линейным.

$$Y_t = \alpha + \beta t + \varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma^2).$$

$MY_t = \alpha + \beta t$ - ряд не является стационарным.

6) ряд с сезонной компонентой не является стационарным.

$$Y_t = S(t) + \varepsilon_t, \quad \text{где } \varepsilon_t \text{ - стационарный ряд с нулевым средним.}$$

В случае 4) и 5) методы моделирования стационарных временных рядов применяются к остаткам регрессии или к сглаженным уровням временного ряда, т. е. к уровням, очищенным от тренда, циклической и сезонной составляющей.

Обнаружение нестационарности.

1. Визуальный анализ временного ряда. Возможно, временной ряд содержит видный на глаз временной тренд и сезонность (периодичную компоненту). Возможно, что разброс значений возрастает или убывает со временем. (признак «случайного блуждания»). Это может служить указанием

на зависимость среднего и , соответственно, дисперсии от времени. Во всех трех случаях ряд, скорее всего, не будет стационарным.

2. Построить график выборочной автокорреляционной функции или коррелограмму. Коррелограмма стационарного временного ряда быстро убывает со временем, быстро уходит почти в ноль после нескольких первых значений – «влияние предыдущих уровней затухает». Если график показывает, что АСФ убывает медленно, с колебаниями, то ряд, скорее всего, будет нестационарным.

Примеры графиков АСФ для ряда с трендом, сезонностью, «случайного блуждания», стационарного временного ряда. (пленка)

3. Формальные тесты на стационарность.

Формальные тесты на стационарность могут определить не только есть ли нестационарность, но и какой природы. Ряд является нестационарным потому что он содержит временной тренд или потому что ряд является случайным блужданием. сновной вопрос: экономические переменные, такие как ВВП, уровень занятости, ставка процента, возвращаются к тренду в случае шока или экономический шок перманентен, т. е. его эффект не исчезнет после нескольких лет, т. е. переменные являются случайными блужданиями.

Этот вопрос важен по двум причинам:

- 1) если эти переменные ведут себя как случайные блуждания, регрессия между ними может быть ложной, и исключение тренда проблемы не решит- ряд снова останется нестационарным.
- 2) Если такие переменные, как ВВП, являются случайными блужданиями, то эффект любого мгновенного шока, такого, как рост цен на нефть, обвал национальной валюты, никуда не денется после нескольких лет и будет перманентным.

Итак, является ли ряд случайным блужданием или нет – вопрос экономической теории. Во многих работах было показано, что почти все макроэкономические переменные представляют из себя случайные блуждания или имеют в качестве

своей составляющей компоненту случайного блуждания. Львиная доля этих исследований использует тест Дикки-Фуллера на единичные корни.

$$Y_t = Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma^2) \quad - \text{ как по имеющимся наблюдениям}$$

определить, является ли ряд «случайным блужданием» или нет, еще говорят, есть ли у нас в уравнении единичный корень. Т. е. в модели

$$Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t \text{ или } Y_t = \rho Y_{t-1} + \varepsilon_t, \text{ или } Y_t = \alpha + \beta t + \rho Y_{t-1} + \varepsilon_t \quad \rho = 1.$$

Очень просто – ответите мне вы, осуществить регрессию Y_t на Y_{t-1} и проверить при помощи t-статистики Стьюдента нулевую гипотезу о том, что $\rho = 1$, т. е. посчитать $\frac{\hat{\rho} - 1}{s_{\hat{\rho}}}$ и сравнить полученное значение с критическим

значением, найденным по таблицам распределения Стьюдента для выбранного уровня значимости с числом степеней свободы $T-2$ ($T-1$) и учитывая односторонность критерия. Однако, в случае, если $\rho = 1$, t-статистика уже не имеет распределение Стьюдента. Ее распределение описано Дики и Фуллером в 1976 году (Dickey & Fuller).

Рассмотрим три регрессии:

$$Y_t = \rho Y_{t-1} + \varepsilon_t \text{ - AR - модель} \tag{1}$$

$$Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t \text{ - AR – модель с константой} \tag{2}$$

$$Y_t = \alpha + \beta t + \rho Y_{t-1} + \varepsilon_t \text{ - AR – модель с константой и трендом} \tag{3}$$

$$H_0: \rho = 1;$$

$$H_a: \rho < 1.$$

Составляем t-статистику $\frac{\hat{\rho} - 1}{s_{\hat{\rho}}}$ и смотрим в таблицу критических точек

распределения Дики-Фуллера. Например, для $T=100$. Таким образом, если мы будем применять стандартный t-тест, мы будем часто отвергать верную гипотезу о наличии единичного корня.

Критические значения, указанные в таблице 1, остаются справедливыми, если в правые части регрессий (1)-(3) добавляются лагированные значения переменной Y , с величиной сдвига больше 1. Это позволяет проверять наличие

единичного корня в авторегрессионных моделях порядка больше первого. Такой тест называется аугментированным тестом Дики-Фуллера. Если порядок авторегрессионного процесса неизвестен, то рекомендуется включать в уравнение возможно большее число лагов, чтобы устранить возможную автокорреляцию ошибок. Дело в том, что в ADF тесте предполагается, что ошибки являются «белым шумом» и критические значения, указанные в Таблице 1. справедливы только при выполнении этого условия. Однако включение большого числа лагов снижает мощность критерия, так что тут главное не переборщить. Определить подходящее число лагов можно при помощи критериев выбора порядка ARMA моделей (см. ниже). Кроме этого, можно проверить статистическую значимость дополнительной лаговой переменной.

Избавление от нестационарности:

1. Выделить тренд и сезонность, т. е. неслучайную составляющую временного ряда. Как это сделать – см. выше.
2. Если ряд представляет «случайное блуждание», то взятие последовательных разностей делает ряд стационарным. На практике порядок разностей, как правило, не больше двух.

МОДЕЛИ СТАЦИОНАРНЫХ ВРЕМЕННЫХ РЯДОВ.

(Модели Бокса-Дженкинса. Модели авторегрессии и скользящего среднего.)

$$Y_t = \alpha + \rho_1 Y_{t-1} + \dots + \rho_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \quad \text{где } \varepsilon_t \sim iid(0, \sigma^2) -$$

процесс авторегрессии порядка p и скользящего среднего порядка q – ARMA(p,q).

$$\sum_{\tau=1}^p \rho_{\tau} Y_{t-\tau} - \text{авторегрессионный член порядка } p,$$

$$\sum_{\tau=1}^q \theta_{\tau} \varepsilon_{t-\tau} - \text{член скользящего среднего порядка } q.$$

Рассмотрим примеры (везде предполагаем, что ряды стационарны).

I. Авторегрессионный процесс порядка p – AR(p).

$$Y_t = \alpha + \rho_1 Y_{t-1} + \dots + \rho_p Y_{t-p} + \varepsilon_t, \text{ где } \varepsilon_t \sim iid(0, \sigma^2).$$

Поскольку ряд стационарен, то $MY_t = MY_{t-1} = \dots = MY_{t-p} = \mu$,

следовательно, $\mu = \frac{\alpha}{1 - \rho_1 - \dots - \rho_p}$. Эта формула дает нам условие

стационарности авторегрессионного ряда. Если ряд стационарный, то μ - конечно, следовательно, $\rho_1 + \dots + \rho_p < 1$. Это условие необходимое, но не достаточное (не является достаточным для стационарности), поскольку есть и другие условия, которые мы должны наложить для того, чтобы ряд AR(p) был стационарным.

Дисперсия и автокорреляционная функция рассчитываются уже не так просто. Формула для расчета ACF рекуррентная, позволяющая по первым p элементам вычислить остальные. Частичная автокорреляционная функция процесса AR(p) имеет ненулевые значения лишь для первых p элементов. Для всех других элементов PACF равна нулю. Это свойство частной автокорреляционной функции часто используется при подборе порядка p в модели авторегрессии для конкретных анализируемых временных рядов. Однако, следует иметь в виду, что этот результат верен для теоретической частной автокорреляционной функции и не обязательно выполняется для выборочной автокорреляционной функции. На практике поступают следующим образом: если, к примеру, рассчитанные на основании исходных данных все коэффициенты частной автокорреляции статистически незначимо отличаются от нуля, начиная с k , то порядок авторегрессии, подбираемой для анализируемого ряда, естественно определить числом $p = k-1$.

а) Авторегрессионный процесс первого порядка – AR(1).

С таким процессом мы уже сталкивались ранее неоднократно:

$$Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t, \varepsilon_t \sim iid(0, \sigma^2), \text{ условие стационарности: } |\rho| < 1.$$

$$MY_t = \alpha + \rho MY_{t-1} = \alpha + \rho\alpha + \rho^2 MY_{t-2} = \alpha(1 + \rho + \rho^2 + \dots) = \frac{\alpha}{1 - \rho}$$

$$DY_t = \frac{\sigma^2}{1-\rho} - \text{(было упражнение) показать.}$$

$$v_L = \frac{\rho^L \sigma^2}{1-\rho^2} - \text{можно показать. Т. о. при } |\rho| < 1 \text{ ряд стационарен.}$$

$$ACF(L) = \frac{Cov(Y_t, Y_{t+L})}{DY_t} = \frac{v_L}{v_0} = \rho^L, L = 0, 1, 2, \dots$$

Коррелограмма (график автокорреляционной функции) простая – она начинается в 1 и убывает геометрически.

Пример. Рассмотрим следующий процесс: $Y_t = 3 + 0.75Y_{t-1} + \varepsilon_t$. На рисунке изображена коррелограмма этого ряда:

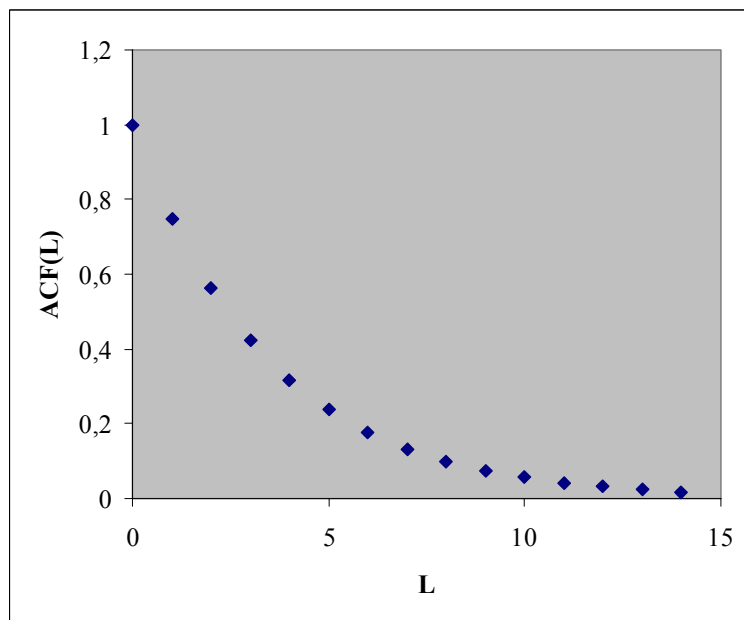


Рис. 1. Коррелограмма ряда $Y_t = 3 + 0.75Y_{t-1} + \varepsilon_t$.

Обратите внимание, что рассматриваемый процесс имеет бесконечную память. Текущее значение процесса зависит от всех его прошлых значений, хотя это влияние с течением времени ослабевает. Реализация рассматриваемого процесса приведена на рис. 2.

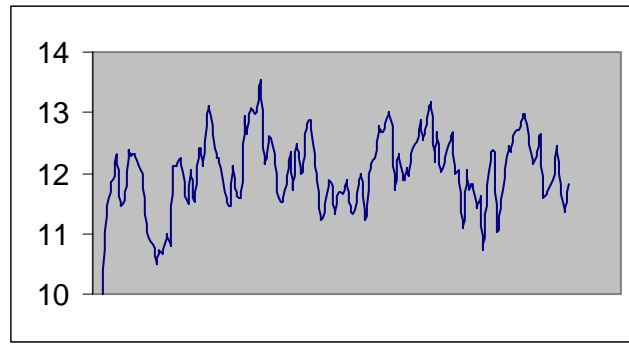


Рис. 2. Типичная реализация процесса $Y_t = 3 + 0.75Y_{t-1} + \varepsilon_t$

$\mu = 12$ для этого процесса.

б) Авторегрессионный процесс второго порядка – AR(2).

$$Y_t = \alpha + \rho_1 Y_{t-1} + \rho_2 Y_{t-2} + \varepsilon_t, \varepsilon_t \sim iid(0, \sigma^2).$$

Математическое ожидание (среднее процесса) есть $\mu = \frac{\alpha}{1 - \rho_1 - \rho_2}$.

Необходимое условие стационарности: $\rho_1 + \rho_2 < 1$. Расчитаем дисперсию процесса и его ACF (в дальнейшем для простоты вычислений положим $\alpha = 0$, в этом случае $\mu = 0$). Дисперсия процесса:

$$DY_t = v_0 = Cov(Y_t, Y_t) = Cov(Y_t, \rho_1 Y_{t-1} + \rho_2 Y_{t-2} + \varepsilon_t) = \rho_1 v_1 + \rho_2 v_2 + \sigma^2$$

$$v_L = Cov(Y_t, Y_{t+L}) = Cov(Y_t, \rho_1 Y_{t+L-1} + \rho_2 Y_{t+L-2} + \varepsilon_t) = \rho_1 v_{L-1} + \rho_2 v_{L-2}.$$

$$ACF(L) = \frac{Cov(Y_t, Y_{t+L})}{DY_t} = \frac{v_L}{v_0} = \varphi_L = \rho_1 \varphi_{L-1} + \rho_2 \varphi_{L-2}.$$

Рассмотрим φ_1 и φ_2 . Получим выражение для них в явной форме:

$$\begin{cases} \varphi_1 = \rho_1 + \rho_2 \varphi_2 \\ \varphi_2 = \rho_1 \varphi_1 + \rho_2 \end{cases} \text{ - система уравнений Юла-Уолкера для AR(2) процесса}$$

(воспользовались тем, что $\varphi_1 = 1$). Решая эту систему, найдем первые два

значения автокорреляционной функции: $\varphi_1 = \frac{\rho_1}{1 - \rho_2}$ и $\varphi_2 = \frac{\rho_1^2}{1 - \rho_2} + \rho_2$.

Тогда, подставляя полученные формулы в выражение для дисперсии процесса,

получим, что $v_0 = \rho_1 v_1 + \rho_2 v_2 + \sigma^2 = \frac{(1 - \rho_2)\sigma^2}{(1 + \rho_2)((1 - \rho_2)^2 - \rho_1^2)}$. Отсюда, учитывая

тот факт, что дисперсия должна быть положительна, получаем еще

необходимые условия стационарности временного ряда, подчиняющегося процессу AR(2): $|\rho_2| < 1$, $\rho_2 - \rho_1 < 1$ и $\rho_1 + \rho_2 < 1$. С последним условием мы уже сталкивались. Можно показать, что в случае выполнения условий стационарности, ACF процесса убывает экспоненциально, в случае, если корни уравнения $1 - \rho_1 L - \rho_2 L^2 = 0$, называемого еще характеристическим уравнением процесса, действительны, и убывает синусоидально, если корни комплексные.

Примеры.

1) $Y_t = 2.5 + 0.9Y_{t-1} - 0.7Y_{t-2} + \varepsilon_t$, $\varepsilon_t \sim iid(0, \sigma^2)$.

Необходимые условия стационарности выполнены. Корни характеристического уравнения комплексные ($0.81 - 4 * 0.7 < 0$).

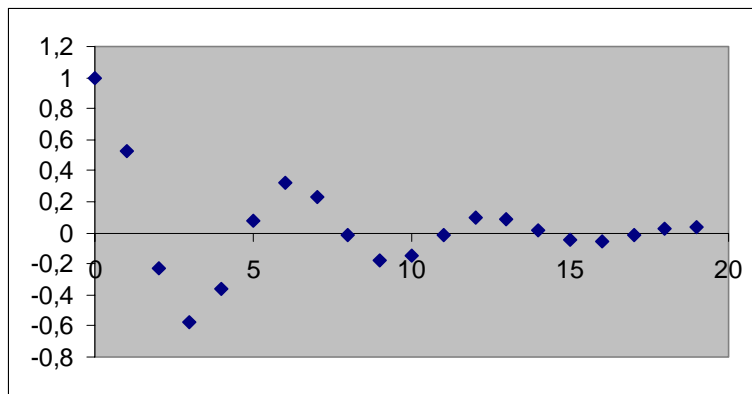


Рис. 3. Коррелограмма ряда $Y_t = 2.5 + 0.9Y_{t-1} - 0.7Y_{t-2} + \varepsilon_t$.

Процесс снова имеет бесконечную память.

$\mu = 3.125$

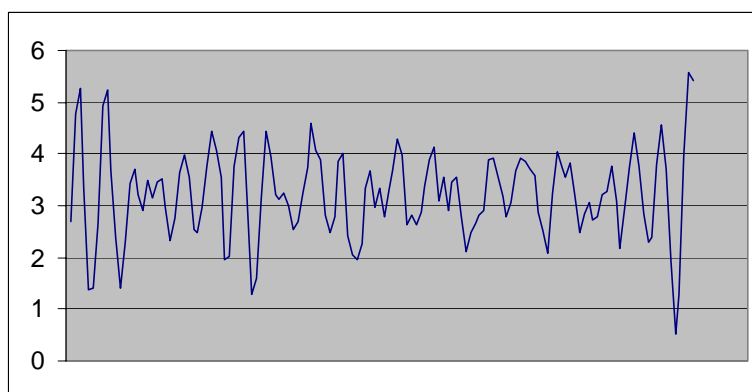


Рис. 4. Типичная реализация процесса $Y_t = 2.5 + 0.9Y_{t-1} - 0.7Y_{t-2} + \varepsilon_t$

2) $Y_t = 2.5 + 0.2Y_{t-1} + 0.7Y_{t-2} + \varepsilon_t$

Корни характеристического уравнения этого процесса действительные (проверить)

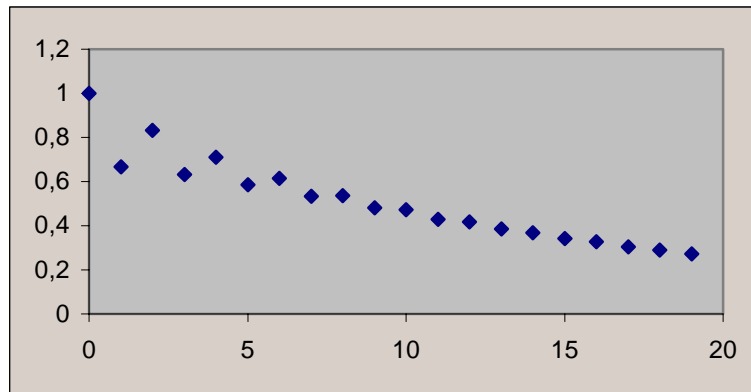


Рис. 5. Коррелограмма ряда $Y_t = 2.5 + 0.2Y_{t-1} + 0.7Y_{t-2} + \varepsilon_t$.

$\mu = 25$

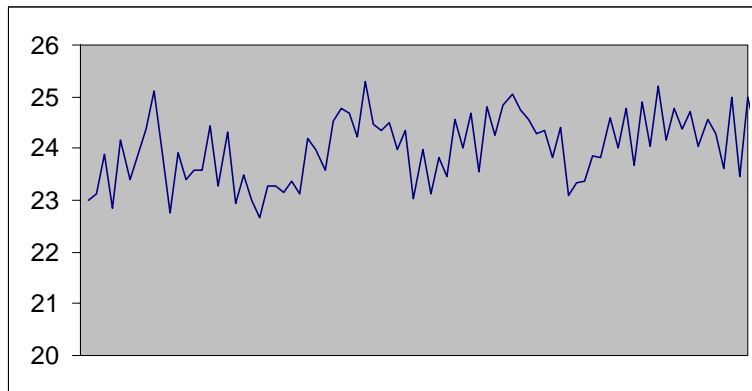


Рис. 6. Типичная реализация процесса $Y_t = 2.5 + 0.2Y_{t-1} + 0.7Y_{t-2} + \varepsilon_t$

II. Процесс скользящего среднего порядка q .

$$Y_t = \alpha + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad \varepsilon_t \sim iid(0, \sigma^2) - MA(q).$$

Нетрудно видеть, что процесс $MA(q)$ всегда стационарен (для любого q и θ_τ). Математическое ожидание процесса $MY_t = \alpha$. Дисперсия процесса:

$$DY_t = v_0 = M(Y_t - \alpha)^2 = M(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q})^2 = \sigma^2(1 + \theta_1^2 + \dots + \theta_q^2)$$

Дисперсия стационарного процесса должна быть конечной. Так у нас вроде все нормально, скажете вы. Однако мы можем рассмотреть модель с конечным числом членов скользящего среднего как аппроксимацию более общей модели. Большинство случайных процессов, которые возникают на практике, требуют бесконечной последовательности скользящих средних в

правой части, следовательно, и бесконечного числа весов θ_τ . Итак, если ряд $MA(\infty)$ является стационарным, дисперсия конечна, следовательно, $\sum_{\tau=1}^{\infty} \theta_\tau^2 < \infty$.

Если ряд сходится, то члены ряда стремятся к 0 с ростом τ , таким образом, θ_τ^2 становятся все меньше и меньше с ростом τ .

Теперь посчитаем ковариацию:

$$v_L = Cov(Y_t - \mu, Y_{t+L} - \mu) = 0, \text{ если } L > q.$$

$$ACF(L) = \begin{cases} 1, L = 0; \\ \dots \\ 0, L > q. \end{cases}$$

Частная автокорреляционная функция процесса $MA(q)$ аналогично автокорреляционной функции для $AR(p)$ процесса экспоненциально убывает. Таким образом, имеет место некая симметрия – пара графиков PACF и ACF для процесса $AR(p)$ имеет тот же вид, что и пара графиков ACF и PACF для $MA(p)$ – процесса.

Примеры процессов скользящего среднего.

а) процесс скользящего среднего порядка 1 – $MA(1)$.

$$Y_t = \alpha + \varepsilon_t - \theta\varepsilon_{t-1} \quad \varepsilon_t \sim iid(0, \sigma^2).$$

Этот процесс имеет среднюю $\mu = \alpha$ и дисперсию $v_0 = \sigma^2(1 + \theta^2)$.

Вычислим ковариацию между двумя соседними уровнями временного ряда, подчиняющегося этому процессу:

$$v_1 = M[(Y_t - \mu)(Y_{t-1} - \mu)] = M[(\varepsilon_t - \theta\varepsilon_{t-1})(\varepsilon_{t-1} - \theta\varepsilon_{t-2})] = -\theta\sigma^2.$$

$$v_2 = M[(Y_t - \mu)(Y_{t-2} - \mu)] = M[(\varepsilon_t - \theta\varepsilon_{t-1})(\varepsilon_{t-2} - \theta\varepsilon_{t-3})] = 0$$

$$v_3 = 0$$

...

Таким образом, процесс $MA(1)$ имеет конечную память в один период. В общем случае ограниченная память процесса скользящего среднего имеет большое значение. Мы можем прогнозировать процесс скользящего среднего только на ограниченное число уровней вперед.

Определим автокорреляционную функцию процесса MA(1):

$$\varphi_L = \begin{cases} 1, & L = 0; \\ -\theta / (1 + \theta^2), & L = 1; \\ 0, & L \geq 2 \end{cases}$$

Рассмотрим пример процесса скользящего среднего порядка 1:

$$Y_t = 2 + \varepsilon_t + 0.8\varepsilon_{t-1}$$

Типичная реализация процесса приведена на рис. 7, автокорреляционная функция – на рис. 8.

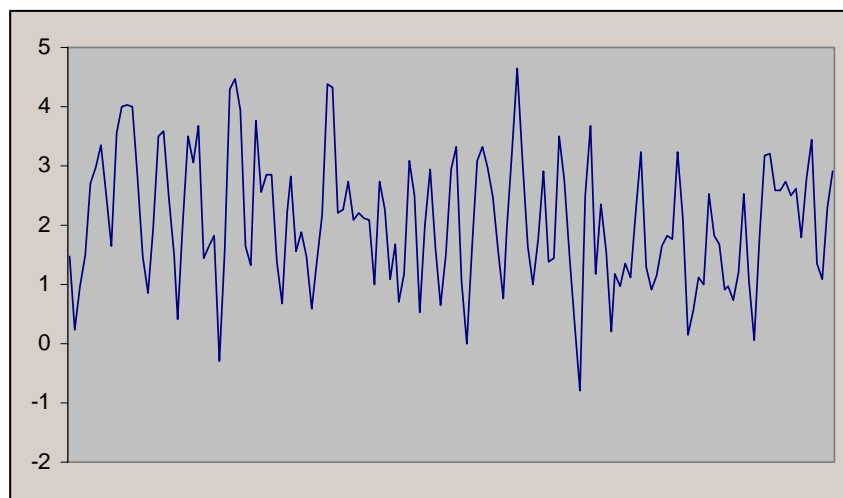


Рис 7. Типичная реализация процесса $Y_t = 2 + \varepsilon_t + 0.8\varepsilon_{t-1}$

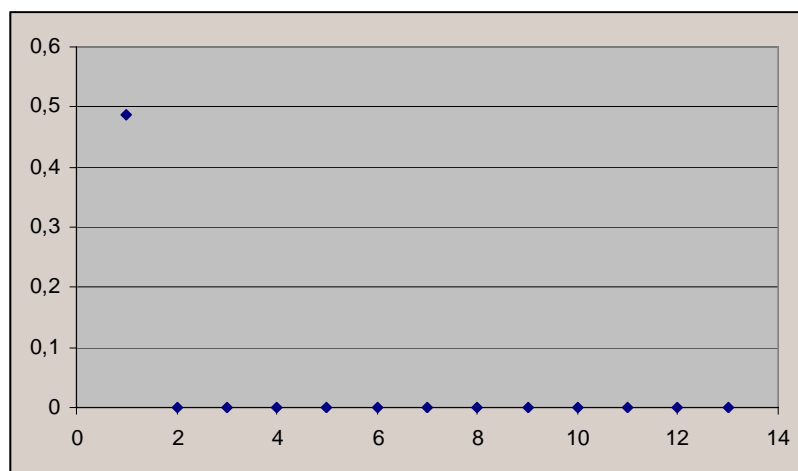


Рис 8. Автокорреляционная функция процесса $Y_t = 2 + \varepsilon_t + 0.8\varepsilon_{t-1}$

б) Процесс скользящего среднего порядка 2 (MA(2)):

$$Y_t = \alpha + \varepsilon_t - \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} \quad \varepsilon_t \sim iid(0, \sigma^2).$$

Этот процесс имеет среднюю $\mu = \alpha$ и дисперсию $v_0 = \sigma^2(1 + \theta_1^2 + \theta_2^2)$.

Вычислим ковариацию между двумя соседними уровнями временного ряда, подчиняющегося этому процессу:

$$v_1 = M[(Y_t - \mu)(Y_{t-1} - \mu)] = M[(\varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2})(\varepsilon_{t-1} - \theta_1\varepsilon_{t-2} - \theta_2\varepsilon_{t-3})] = -\theta_1(1 - \theta_2)\sigma^2$$

$$v_2 = M[(Y_t - \mu)(Y_{t-2} - \mu)] = M[(\varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2})(\varepsilon_{t-2} - \theta_1\varepsilon_{t-3} - \theta_2\varepsilon_{t-4})] = -\theta_2\sigma^2$$

$$v_3 = 0$$

Автокорреляционная функция:

$$\varphi_L = \begin{cases} 1, & L = 0; \\ \frac{-\theta_1(1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2}, & L = 1; \\ \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}, & L = 2; \\ 0, & L \geq 3 \end{cases}$$

Процесс МА(2) имеет память в два периода.

Рассмотрим пример процесса скользящего среднего порядка 2: $Y_t = 2 + \varepsilon_t + 0.6\varepsilon_{t-1} - 0.3\varepsilon_{t-2}$. Среднее процесса $\mu = 2$. Типичная реализация процесса приведена на рис. 9., автокорреляционная функция – на рис. 10.

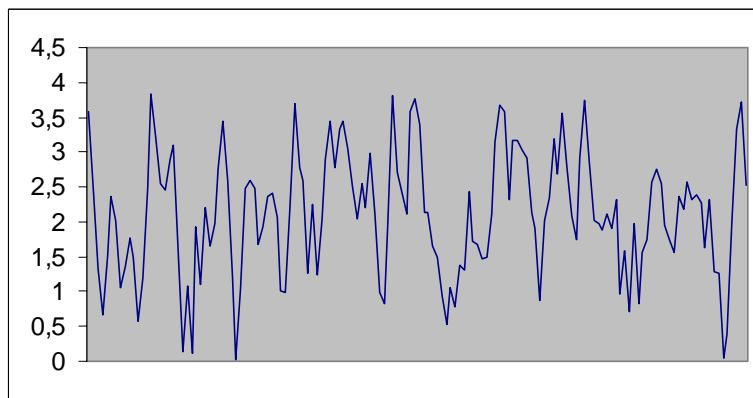


Рис. 9. Типичная реализация процесса $Y_t = 2 + \varepsilon_t + 0.6\varepsilon_{t-1} - 0.3\varepsilon_{t-2}$

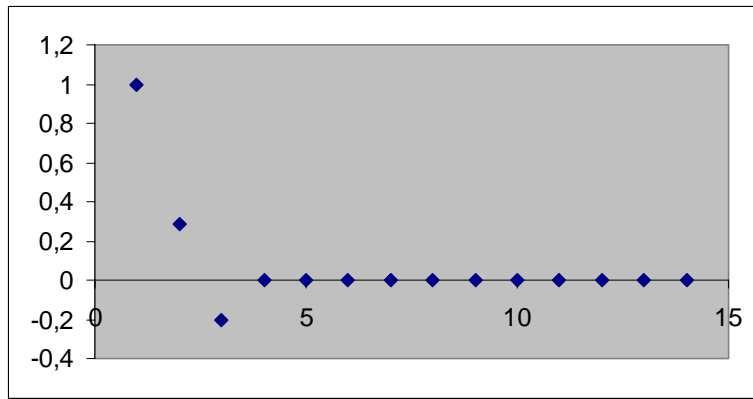


Рис.10. Автокорреляционная функция процесса $Y_t = 2 + \varepsilon_t + 0.6\varepsilon_{t-1} - 0.3\varepsilon_{t-2}$

Упражнение. Показать, что процесс скользящего среднего имеет память в q периодов и его автокорреляционная функция выражается следующим образом:

$$\varphi_L = \begin{cases} 1, & L = 0; \\ \frac{-\theta_L + \theta_1\theta_{L+1} + \dots + \theta_{q-L}\theta_q}{1 + \theta_1^2 + \dots + \theta_q^2}, & L = 1, \dots, q; \\ 0, & L \geq q \end{cases}$$

III. Процесс авторегрессии порядка p и скользящего среднего порядка q – ARMA(p, q).

Множество стационарных процессов нам не удастся смоделировать при помощи чисто как авторегрессионный процесс или как процесс скользящего среднего, поскольку обладают качествами обоих процессов. Рассмотрим следующий процесс:

$$Y_t = \alpha + \rho_1 Y_{t-1} + \dots + \rho_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \text{ где } \varepsilon_t \sim iid(0, \sigma^2).$$

Поскольку ряд стационарен, то $MY_t = MY_{t-1} = \dots = MY_{t-p} = \mu$,

следовательно, $\mu = \frac{\alpha}{1 - \rho_1 - \dots - \rho_p}$. Эта формула дает нам условие

стационарности авторегрессионного ряда. Если ряд стационарный, то μ - конечно, следовательно, $\rho_1 + \dots + \rho_p < 1$.

Дисперсия и ковариация снова рассчитываются при помощи рекурсивных формул.

а) рассмотрим простейший процесс – процесс авторегрессии и скользящего среднего порядка 1 – ARMA(1,1).

$$Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}, \text{ где } \varepsilon_t \sim iid(0, \sigma^2).$$

$$\text{Математическое ожидание процесса: } MY_t = \mu = \frac{\alpha}{1 - \rho}$$

условие стационарности: $|\rho| < 1$.

Дисперсия процесса (для простоты положим $\alpha = 0$):

$$\begin{aligned} DY_t = v_0 &= M(Y_t Y_t) = M(Y_t(\rho Y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1})) = M(\rho Y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1})^2 = \\ &= \rho^2 v_0 - 2\rho\theta M(Y_{t-1} \varepsilon_{t-1}) + \sigma^2 + \theta^2 \sigma^2. \end{aligned}$$

Учитывая тот факт, что $M(Y_{t-1} \varepsilon_{t-1}) = \sigma^2$ (показать), получаем для $|\rho| < 1$:

$$v_0 = \frac{1 + \theta^2 - 2\theta\rho}{1 - \rho^2} \sigma^2.$$

Ковариация рассчитывается рекурсивно:

$$\begin{aligned} v_1 &= M(Y_t Y_{t-1}) = M(Y_{t-1}(\rho Y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1})) = \rho v_0 - \theta \sigma^2 = \\ &= \frac{(1 - \rho\theta)(\rho - \theta)}{1 - \rho^2} \sigma^2 \end{aligned}$$

$$v_2 = M(Y_{t-1} Y_{t-2}) = M(Y_{t-2}(\rho Y_{t-2} + \varepsilon_{t-1} - \theta \varepsilon_{t-2})) = \rho v_1.$$

Подобным образом $v_k = \rho v_{k-1}$.

Автокорреляционная функция процесса:

$$\phi_1 = \frac{v_1}{v_0} = \frac{(1 - \rho\theta)(\rho - \theta)}{1 + \theta^2 - 2\theta\rho}.$$

И для $L \geq 2$ получаем: $\phi_L = \rho \phi_{L-1}$.

Таким образом, автокорреляционная функция процесса авторегрессии и скользящего среднего начинается в ϕ_1 , значение которого является функцией ρ

и θ , и затем убывает геометрически. ACF процесса $ARMA(1,1)$ ведет себя точно так же, как ACF процесса $AR(1)$. Хотя выражение для ϕ_1 другое, соотношение между последующими уровнями то же самое. Этот вывод можно обобщить и на случай $ARMA(p,q)$ - процессов. Первые q значений ACF определяются взаимодействием AR и MA компонент, а дальнейшее ее поведение такое же, как и в $AR(p)$ процессе (поскольку процесс скользящего среднего порядка q имеет память только на q периодов).

Аналогичный вывод справедлив и для частной автокорреляционной функции $AR(p,q)$. Она убывает подобно $PACF$ процесса $MA(q)$.

1) Рассмотрим процесс $Y_t = 2 + 0.8Y_{t-1} + \varepsilon_t - 0.9\varepsilon_{t-1}$, $\varepsilon_t \sim iid(0, \sigma^2)$

$\mu = 10$

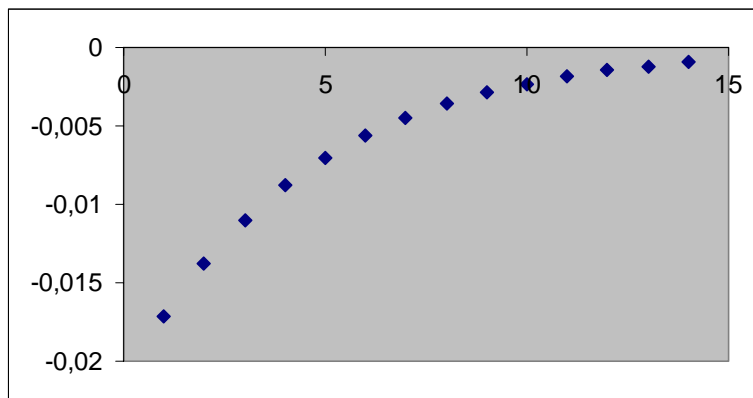


Рис 11. Автокорреляционная функция процесса $Y_t = 2 + 0.8Y_{t-1} + \varepsilon_t - 0.9\varepsilon_{t-1}$.

Рис. 12. Частная автокорреляционная функция процесса.

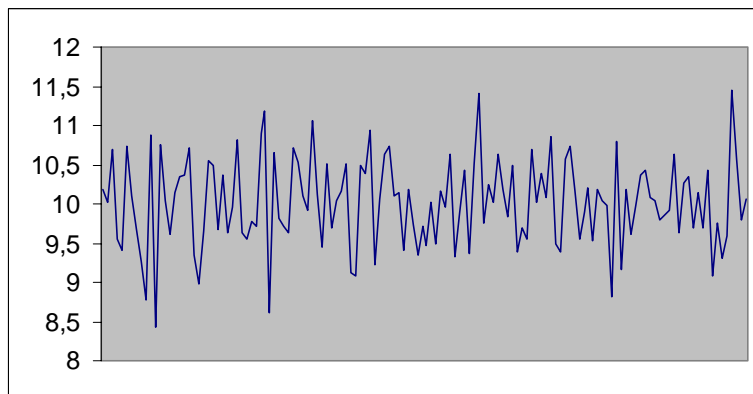


Рис 13. Типичная реализация процесса $Y_t = 2 + 0.8Y_{t-1} + \varepsilon_t - 0.9\varepsilon_{t-1}$

2) Рассмотрим процесс $Y_t = 2 - 0.8Y_{t-1} + \varepsilon_t + 0.9\varepsilon_{t-1}$, $\varepsilon_t \sim iid(0, \sigma^2)$

$$\mu = 1.111$$

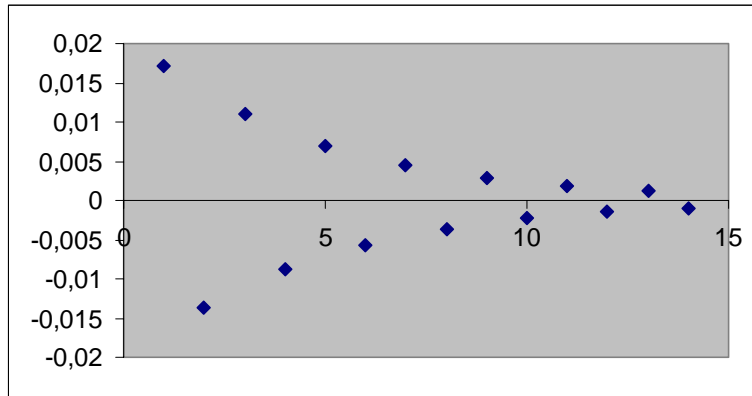


Рис 14. Автокорреляционная функция процесса $Y_t = 2 - 0.8Y_{t-1} + \epsilon_t + 0.9\epsilon_{t-1}$.

Рис 15. Частная автокорреляционная функция процесса

$$Y_t = 2 - 0.8Y_{t-1} + \epsilon_t + 0.9\epsilon_{t-1}.$$

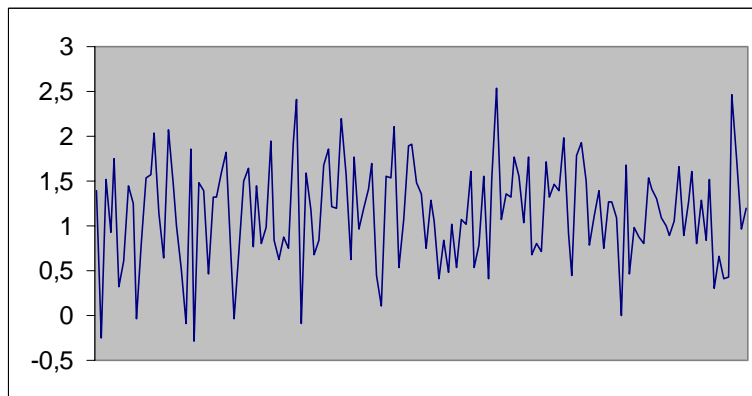


Рис 16. Типичная реализация процесса $Y_t = 2 - 0.8Y_{t-1} + \epsilon_t + 0.9\epsilon_{t-1}$.

Для смешанных процессов авторегрессии и скользящего среднего более высоких порядков дисперсия, ковариация и автокорреляционная функция являются решениями разностных уравнений и не могут быть выражены в явном виде (имеется в виду просто). Однако, можно показать, что

$$v_k = \rho_1 v_{k-1} + \dots + \rho_p v_{k-p} \text{ для } k \geq q+1, \text{ и, таким образом,}$$

$$\phi_k = \rho_1 \phi_{k-1} + \dots + \rho_p \phi_{k-p} \text{ для } k \geq q+1. \text{ Заметим, что } q \text{ — память}$$

авторегрессионного процесса, так что автокорреляционная функция (и ковариация) смешанного процесса проявляют себя так же, как

автокорреляционная функция (и ковариация) чисто авторегрессионного процесса того же порядка.

МОДЕЛИ НЕСТАЦИОНАРНЫХ ГОМОДЕНИЧНЫХ ВРЕМЕННЫХ РЯДОВ

(ARIMA(p,d,q) – процессы).

На практике большинство рядов, с которыми сталкивается исследователь, являются нестационарными, так что характеристики ряда меняются с течением времени. Нестационарный ряд называется гомоденичным, если его можно свести к стационарному, взяв последовательные разности некоторого порядка. Если ряд, составленный из конечных разностей порядка d нестационарного временного ряда Y_t , является стационарным, то говорят, что ряд Y_t является интегрируемым порядка d и обозначают $Y_t \sim I(d)$.

Пусть ряд является интегрируемым порядка d . Обозначим $W_t = \Delta^d Y_t$, где Δ обозначает взятие последовательных разностей, т. е. $\Delta Y_t = Y_t - Y_{t-1}$, $\Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1}$. Ряд W_t является стационарным временным рядом. Зная ряд W_t , мы можем получить исходный ряд Y_t , суммируя ряд W_t d раз. Мы обозначим $Y_t = \Sigma^d W_t$, где Σ - оператор суммирования $\Sigma W_t = \sum_{\tau=-\infty}^t W_\tau$, $\Sigma^2 W_t = \sum_{j=-\infty}^t \sum_{\tau=-\infty}^j W_\tau$, тогда $Y_t = Y_0 + W_1 + \dots + W_t$, Если Y_t был дифференцирован дважды, то мы можем вычислить Y_t из W_t , суммируя W_t дважды.

Упражнение. Если ряд интегрируем порядка 2, получить явное выражение Y_t из W_t .

После того, как мы из нестационарного ряда путем интегрирования его d раз (на практике d не больше двух) получили стационарный ряд W_t , мы можем моделировать ряд W_t как ARMA(p,q). Если ряд Y_t интегрируем порядка d , т. е. $W_t = \Delta^d Y_t$, а W_t – авторегрессионный процесс скользящего среднего, порядок авторегрессии p и скользящего среднего q (ARMA(p,q)), то говорят, что Y_t

является интегрированным процессом авторегрессии и скользящего среднего порядка (p,d,q) или просто $ARIMA(p,d,q)$.

$$\Delta^d Y_t = \alpha + \rho_1 \Delta^d Y_{t-1} + \dots + \rho_p \Delta^d Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

Заметим, что в этом случае $\mu_W = \frac{\alpha}{1 - \rho_1 - \dots - \rho_p}$ и если $\alpha \neq 0$, то ряд Y_t

содержит детерминистический тренд

Анализ гомоденичных временных рядов состоит из следующих шагов:

- I. Спецификация *ARIMA*-моделей.
- II. Оценивание модели.
- III. Проверка адекватности модели.
- IV. Прогнозирование на основе построенной модели.

Рассмотрим каждый шаг более подробно.

I. СПЕЦИФИКАЦИЯ *ARIMA* – МОДЕЛЕЙ.

Как мы видели, многие нестационарные временные ряды моделируются как *ARIMA*-процессы. Проблемой является выбор для данного временного ряда подходящих порядков интегрирования, авторегрессии и скользящего среднего, т. е. выбор d , p и q . Выбор подходящих порядков модели называется спецификацией *ARIMA* – процесса. Эта проблема частично решается визуальным анализом автокорреляционной и частной автокорреляционной функцией временного ряда..

Для данного ряда Y_t первой проблемой является выбор d – порядка интегрирования. Для того, чтобы определить d , используются методы, описанные нами ранее: визуальный анализ ряда, анализ автокорреляционной функции ряда, формальные тесты на стационарность (аугментированный тест Дики-Фуллера) – тесты на единичные корни. Для начала берем временной ряд, если он нестационарный, применяем к нему операцию взятия последовательной разности и смотрим на получившийся ряд. Если он стационарный, то ОК – порядок интегрирования d равен 1 и переходим к следующему пункту, если нет, то интегрируем ряд еще раз и т. д. Таким образом, мы определяем порядок интегрирования d . на практике d редко бывает больше двух.

После определения d мы получаем стационарный временной ряд $W_t = \Delta^d Y_t$ и анализируем его автокорреляционную функцию и частную

автокорреляционную функцию для того, чтобы определить подходящие значения p и q . Для процессов невысокого порядка это не так трудно сделать, поскольку автокорреляционные функции и частные автокорреляционные функции процессов $AR(1)$, $AR(2)$, $MA(1)$, $MA(2)$ и $ARMA(1,1)$ легко опознать. Однако, если ряд W_t не может быть смоделирован $ARMA$ – процессом низких порядков, спецификация p и q становится более сложной и требует более пристального изучения автокорреляционной функции и частной автокорреляционной функции. Например, пики в автокорреляционной функции являются индикаторами члена скользящего среднего, частная автокорреляционная функция может быть использована для определения порядка авторегрессионной части ряда. На этом этапе мы можем сформулировать несколько гипотез относительно возможных значениях порядков p и q . Как мы увидим позднее, существует ряд методов, при помощи которых можно проверить правильность предположения о значениях порядков – проверка адекватности модели. Первый шаг на этом этапе – построить автокорреляционную функцию остатков $ARMA(p,q)$ – модели и посмотреть, похожи ли остатки на «белый шум», если нет, то надо выбрать для ряда новую спецификацию. Остановимся на этом более подробно ниже.

II. ОЦЕНИВАНИЕ МОДЕЛИ.

Предположим, что мы остановились на какой-то спецификации имеющегося временного ряда, т. е. выбрали значения p , d , q для $ARIMA(p,d,q)$ -модели:

$$\Delta^d Y_t = \alpha + \rho_1 \Delta^d Y_{t-1} + \dots + \rho_p \Delta^d Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \text{ где } \varepsilon_t \sim iid(0, \sigma^2).$$

Теперь наша задача состоит в оценке p параметров авторегрессионной части ряда ρ_1, \dots, ρ_p , q параметров части скользящего среднего $\theta_1, \dots, \theta_q$ и свободного члена α . В современные пакеты встроены различные методы оценивания $ARIMA$ – моделей, такие как линейный или нелинейный МНК.

полный или условный метод максимального правдоподобия. Остановившись подробно на этих методах мы не будем.

После того, как оценили модель, необходимо провести этап

III. ПРОВЕРКА АДЕКВАТНОСТИ МОДЕЛИ.

Этот этап осуществляется для того, чтобы проверить правильность предположений относительно параметров модели. Процедура диагностики модели обычно состоит из трех шагов.

1) Оценки коэффициентов модели должны достоверно отличаться от нуля.

2) Сравниваем автокорреляционные функции для ряда, полученного симуляцией, т. е. временного ряда, сгенерированного моделью, с выборочной автокорреляционной функцией анализируемого ряда. Если эти автокорреляционные функции сильно отличаются, мы должны выбрать другую спецификацию для нашего ряда. Если же автокорреляционные функции отличаются не очень сильно, необходимо провести анализ остатков.

3) Поскольку мы с вами ожидаем, что остатки e_t представляют ошибки, которые по условию должны быть некоррелированы, мы можем проверить на автокорреляцию ряд, составленный из остатков. Если они коррелируют, мы должны выбрать другую спецификацию, т. е. выбрать другие значения для p , q и d . После того, как мы оценили новую модель, снова проверяем остатки. После того, как анализ остатков нас удовлетворил, мы можем использовать выбранную модель для прогнозирования.

Обозначим автокорреляционную функцию остатков следующим образом:

$$r_L = \frac{\sum e_t e_{t-L}}{\sum e_t^2}$$

Если модель правильно специфицирована, то остатки являются «белым шумом» и для больших величин смещения L (для моделей невысокого порядка $L \geq 5$) автокорреляция остатков r_L имеет нормальное распределение со средним

0 и дисперсией $1/T$. Поэтому, для того, чтобы проверить нулевую гипотезу о том, что коэффициент корреляции между уровнями ряда остатков равен нулю, нам надо посмотреть, лежит ли r_L внутри интервала $0 \pm \frac{u_{kp}}{\sqrt{T}}$, если нет, то мы с заданной вероятностью можем отвергнуть нулевую гипотезу.

Кроме этого теста на практике для того, чтобы убедиться в том, что остатки некоррелируют друг с другом, используют еще ряд тестов:

1. Рассмотрим статистику Q , которая вычисляется следующим образом: $Q = T \sum_{i=1}^K r_i^2$ - Q-статистика Бокса-Пирса. Эта статистика представляет собой сумму K независимых нормально распределенных случайных величин со средними 0 и дисперсиями $1/T$, поэтому Q распределена примерно (приблизенно) как «хи-квадрат» с числом степеней свободы $K - p - q$. Мы сказали «приблизенно», потому что первые несколько членов будут иметь дисперсию несколько меньше, чем $1/T$ и могут коррелировать между собой. Для того, чтобы принять решение о справедливости нулевой гипотезы, значение Q-статистики сравнивают с критическим значением распределения «хи-квадрат» с числом степеней свободы $K - p - q$ для выбранного уровня значимости.

Пример. Пусть для данного временного ряда мы выбрали спецификацию $ARMA(1,1)$, оценили модель, посчитали Q-статистику и получили, что $Q=33,5$ для $K=20$. По таблице критических значений распределения «хи-квадрат» с числом степеней свободы 18 мы нашли $\chi_{kp}^2(0.05;18) = 28.9$. Таким образом, Q-статистика слишком велика и мы не принимаем для нашего ряда модель $ARMA(1,1)$, поскольку вероятность того, что ошибки не являются «белым шумом», как минимум 95%.

Иногда рассматривают еще модификацию Q-статистики Бокса-Пирса – тест Льюнга-Бокса. Его распределение ближе к «хи-квадрат» на конечных выборках.

Если же мы с вами находимся в ситуации, когда несколько *ARMA* моделей оказываются адекватными исходному временному ряду, можно дать несколько рекомендаций:

1. Выбрать наиболее простую модель.
2. Выбрать модель, наилучшую с точки зрения прогноза – см. ниже.
3. Провести дополнительные тесты. В компьютерные пакеты среди результатов оценивания приводится информационный критерий Акаике (Akaike):

$$AIC = \frac{p+q}{T} + \ln \left(\frac{\sum_{t=1}^T e_t^2}{T} \right).$$

Критерий Акаике является попыткой свести в один показатель два требования: уменьшение параметров модели и улучшение качества подгонки модели. Из двух моделей выбирают ту, для которой значение *AIC* является меньшим.

В пакете Eviews приводится также значение критерия Шварца (Schwarz), отличие которого от информационного критерия Акаике состоит в большем штрафе за количество параметров:

$$\frac{(p+q) \ln T}{T} + \ln \left(\frac{\sum_{t=1}^T e_t^2}{T} \right)$$

16. СИСТЕМЫ ОДНОВРЕМЕННЫХ УРАВНЕНИЙ.

Цель любого эконометрического исследования – изучение связей между экономическими переменными вне зависимости оттого, что является конечной целью прикладного исследования: прогноз или управление. До сих пор мы рассматривали модели и методы, предназначенные для анализа ситуации, описываемой одним автономным уравнением. Теперь же интересующие нас зависимости будут описываться целой системой взаимосвязанных соотношений. В связи с этим возникает ряд новых вопросов, связанных с правильной идентификацией системы. Кроме того, мы должны выяснить, можно ли использовать методы, которые мы рассматривали при изучении автономных уравнений, последовательно к каждому уравнению системы. Или же нам необходимы специальные методы для оценивания одновременно всех уравнений системы.

В нашей лекции мы ограничимся рассмотрением систем одновременных уравнений (СОУ) (simultaneous equations), которые наиболее часто возникают в экономической практике – есть еще, например, системы внешне не связанных уравнений. Простейший и наиболее общий пример СОУ, который мы с вами тоже будем рассматривать для иллюстрации методов и моделей – уравнение спроса и предложения на рынке какого-нибудь товара:

$$\text{уравнение предложения: } Q_t^S = \alpha_1 + \alpha_2 P_t + \alpha_3 P_{t-1} + \varepsilon_t;$$

$$\text{уравнение спроса: } Q_t^D = \beta_1 + \beta_2 P_t + \beta_3 Y_t + u_t;$$

$$\text{условие равновесия на рынке: } Q_t^S = Q_t^D,$$

где P – цена на товар, Y – доход, Q – количество товара.

Для иллюстрации того факта, что обычный МНК, примененный к каждому отдельному уравнению системы приводит к смещенности и несостоятельности оценок, мы будем использовать несколько более простую модель.

Для начала нам надо ввести ряд новых терминов вместо зависимых и независимых переменных, поскольку в СОУ одна и та же переменная может находиться как в правой части одного уравнения (быть объясняющей), так и в левой части другого уравнения (быть объясняемой).

Рассмотрим следующую модель, которая содержит функцию спроса тождество, определяющее доход:

$$C_t = \alpha + \beta Y_t + \varepsilon_t$$

$$Y_t = C_t + Z_t$$

где C – потребительские расходы, Y – доход, Z – непотребительские расходы. Конечно, это простейшая модель. Ее можно расширить, добавив, например, лаговые значения переменной Y , другие переменные.

Случайный член в первом уравнении удовлетворяет условиям гомоскедастичности и отсутствия автокорреляции ошибок. Как мы видим, СОУ могут содержать как уравнения (содержат параметры, подлежащие оценке и случайную составляющую) и тождества. Почему же мы не можем применить в лоб МНК для оценки параметров уравнения (1). Вспомним теорему Гаусса-Маркова. Одно из условий – отсутствие корреляции объясняющей переменной, т. е. той переменной, которая стоит в правой части уравнения, со случайным членом. Что мы имеем в нашем случае?

Подставим из уравнения (1) выражение для C_t в уравнение (2):

$$C_t = \alpha + \beta Y_t + \varepsilon_t$$

$$Y_t = C_t + Z_t, \text{ получим}$$

$$Y_t = \alpha + \beta Y_t + \varepsilon_t + Z_t \text{ или } Y_t = \frac{\alpha}{1-\beta} + \frac{1}{1-\beta} Z_t + \frac{1}{1-\beta} \varepsilon_t$$

если объем непотребительских расходов (что это такое) возрастает на 1, то доход увеличивается на $\frac{1}{1-\beta}$.

Таким образом, величина Y_t включает в себя случайную составляющую ε_t , следовательно, Y_t в уравнении (1.1) коррелирована с ε_t автоматически. Чем это плохо? В этом случае, оценки коэффициентов уравнения (1.1), построенные

при помощи обычного МНК, становятся смещенными и несостоятельными, т. е. смещение не элиминируется с ростом выборки. В общем случае мы даже ничего не можем сказать про направление смещения. Кроме того, рассчитанные стандартные ошибки будут некорректными, т. е. мы не сможем проверять различные гипотезы. Проблема смещения, порождаемая СОУ, может быть разрешена путем замены ОМНК на другой метод оценивания. Об этом поговорим немного позже.

Теперь несколько определений.

Переменные в СОУ делятся на эндогенные и предопределенные. Предопределенные переменные отличаются от экзогенных тем, что предопределенные переменные некоррелированы с остаточным членом. В случае отсутствия автокорреляции ошибок, в качестве предопределенных переменных выступают лагированные значения эндогенных переменных (объяснение) и экзогенные переменные. В нашем примере C и Y – эндогенные переменные, Z – экзогенная переменная.

Рассмотрим другой пример:

$$C_t = \alpha + \beta Y_t + \varepsilon_t$$

$$Y_t = C_t + Z_t$$

Предположим теперь, что на непотребительские расходы влияют, скажем, последние изменения в уровне дохода и величина процентной ставки r . Включим в модель третье уравнение:

$$Z_t = \gamma(Y_{t-1} - Y_{t-2}) + \delta r_t + u_t$$

Теперь у нас переменные C , Y и Z эндогенные, r – экзогенная, кроме того, третье уравнение содержит лаговые значения эндогенных переменных, которые вместе с единственной экзогенной переменной являются предопределенными. Таким образом, значения эндогенных переменных, как особенно четко мы увидим позднее, определяются в результате одновременного воздействия образующих модель соотношений. Значение же экзогенных переменных считаются заданными извне, нашей системой они не определяются, в модели ничего не говорится о том, как образуются значения этих переменных.

Уравнения, составляющие исходную модель, называются структурными уравнениями модели. Как мы видели, структурные уравнения могут быть двух типов: поведенческие – описывают эмпирические взаимодействия между переменными, и уравнения-тождества. Вторые не содержат каких-либо подлежащих оценке параметров и не включают остаточный член. Структурные уравнения могут содержать эндогенные переменные как в правой, так и в левой части. Вернемся к нашему исходному примеру. Уравнения (1) и (2) – структурные уравнения модели. Путем несложных преобразований мы получили уравнение для Y_t , которое содержало в правой части лишь predetermined переменные и случайный член. Аналогичное уравнение можно получить и для C_t :

$$C_t = \frac{\alpha}{1-\beta} + \frac{\beta}{1-\beta} Z_t + \frac{1}{1-\beta} \varepsilon_t$$

Такие уравнения называются уравнениями в приведенной форме. В правой части таких уравнений могут быть и лаговые значения эндогенных переменных.

Упражнение. Составить приведенные уравнения для примера с процентной ставкой.

Остаточный член в приведенных уравнениях по-прежнему удовлетворяет условию Гаусса-Маркова. Следовательно, мы можем применять ОМНК к приведенным уравнениям – косвенный метод наименьших квадратов (КМНК).

Рассмотрим приведенное уравнение для Y_t :

$$Y_t = \frac{\alpha}{1-\beta} + \frac{1}{1-\beta} Z_t + \frac{1}{1-\beta} \varepsilon_t$$

Переобозначим:

$$Y_t = \alpha' + \beta' Z_t + u_t$$

где $\alpha' = \frac{\alpha}{1-\beta}$, $\beta' = \frac{1}{1-\beta}$, $u_t = \frac{\varepsilon_t}{1-\beta}$.

Оценивая (1.4) при помощи ОМНК, мы получим оценки a' и b' , которые будут являться несмещенными оценками для α' и β' . Возвращаясь к исходному уравнению, получим оценки для параметров α и β :

$$a' = \frac{a}{1-b}, \quad b' = \frac{1}{1-b}.$$

Получили два уравнения для двух неизвестных.

$$a = \frac{a'}{1+b'}, \quad b = \frac{1}{1+b'}$$

Поскольку мы можем получить единственное выражение для a и b через оценки a' и b' , уравнения называются однозначно определенными (идентифицируемыми). К сожалению, так бывает не всегда:

- 1) нельзя получить единственные значения коэффициентов структурного уравнения. В этом случае уравнение называют неопределенным – неидентифицируемым;
- 2) Нельзя получить никакого решения – в случае переопределенного (сверхидентифицируемого) уравнения.

Неидентифицируемость.

Модель 1. Рассмотрим модель, состоящую из двух поведенческих уравнений. Допустим, что предложение товара на душу населения и спрос на него задаются следующими уравнениями:

$$Q^D = \alpha + \beta P + \gamma X + u_D - \text{уравнение спроса,}$$

$$Q^S = \delta + \varepsilon P + u_S - \text{уравнение предложения.}$$

Здесь P – цена товара, X – доход на душу населения.

Переменная X экзогенная, а P и Q – эндогенные, их значения определяются в процессе установления рыночного равновесия. Когда рынок находится в равновесии $Q_D = Q_S = Q$. Выразив P и Q через X и остаточные члены (проделать самостоятельно), получим уравнения в приведенной форме:

$$P = \frac{\alpha - \beta}{\varepsilon - \beta} + \frac{\gamma}{\varepsilon - \beta} X + \frac{u_D - u_S}{\varepsilon - \beta},$$

$$Q = \frac{\alpha\varepsilon - \beta\delta}{\varepsilon - \beta} + \frac{\gamma\varepsilon}{\varepsilon - \beta} X + \frac{\varepsilon u_D - \beta u_S}{\varepsilon - \beta}.$$

Перепишем уравнения в приведенном виде как следующее:

$$P = \alpha' + \beta' X + v_P$$

$$Y = \gamma' + \delta' X + v_P, \text{ где}$$

$$\alpha' = \frac{\alpha - \delta}{\varepsilon - \beta}, \beta' = \frac{\gamma}{\varepsilon - \beta}, \gamma' = \frac{\alpha\varepsilon - \beta\delta}{\varepsilon - \beta}, \delta' = \frac{\gamma\varepsilon}{\varepsilon - \beta},$$

$$v_P = \frac{u_D - u_S}{\varepsilon - \beta}, v_D = \frac{\varepsilon u_D - \beta u_S}{\varepsilon - \beta}.$$

Применим для оценки коэффициентов получившейся системы КМНК:

$$\hat{P} = a' + b' X$$

$$\hat{Y} = c' + d' X. \text{ Получили четыре уравнения для пяти неизвестных:}$$

$$a' = \frac{a - d}{e - b}, b' = \frac{c}{e - b}, c' = \frac{ae - bd}{e - b}, d' = \frac{ce}{e - b}.$$

Однако здесь мы можем достичь некоторых результатов:

1. мы можем получить оценку e из второго и четвертого соотношений:

$$\frac{d'}{b'} = \frac{ce}{e - b} / \frac{c}{e - b} = e$$

2. первое и третье соотношение дают нам оценку d :

$$c' - ea' = \frac{ae - bd}{e - b} - e \frac{a - d}{e - b} = d$$

Итак, в ходе рассмотрения этой СОУ, при помощи КМНК мы смогли оценить уравнение предложения. Для получения оценок параметров структурного уравнения спроса у нас осталось два уравнения на три переменные. Следовательно, уравнение спроса остается неопределенным. Вообще-то мы могли придать произвольное значение одному из параметров, например, c . Однако, это привело бы к наличию бесконечного числа кривых спроса, удовлетворяющих нашим данным.

Модель 2. Рассмотрим другую модель.

$Q_t^S = \alpha + \beta P_t + u_t^S$ - уравнение предложения;

$Q_t^D = \gamma + \delta P_t + u_t^D$ - уравнение спроса;

$Q_t^S = Q_t^D$ - условие равновесия на рынке.

или с подстановками, учитывая равновесие:

$$Q_t = \alpha + \beta P_t + u_t^S$$

$$Q_t = \gamma + \delta P_t + u_t^D .$$

Приведенная форма уравнений:

$$Q_t = \frac{\alpha\delta - \beta\gamma}{\delta - \beta} + \frac{\delta u_t^S - \beta u_t^D}{\delta - \beta}$$

$$P_t = \frac{\alpha - \gamma}{\delta - \beta} + \frac{u_t^S - \beta u_t^D}{\delta - \beta}$$

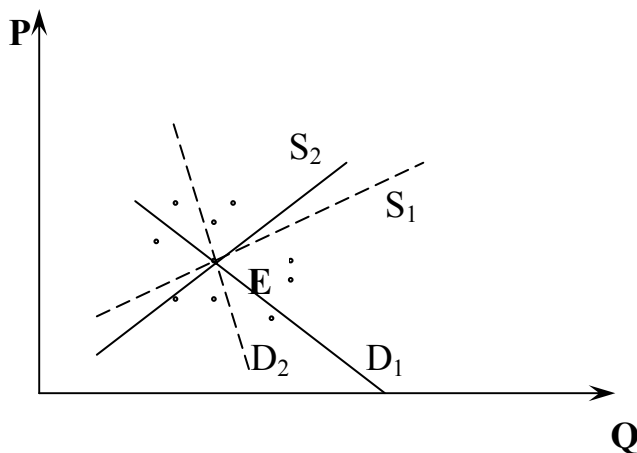


Рис. 1. Модель спроса-предложения

Проинтерпретируем уравнение графически на графике кривых спроса и предложения. В каждый момент времени t мы наблюдаем только равновесные значения P и Q . Пусть E – истинная точка равновесия спроса и предложения на рынке рассматриваемого товара. Эта точка является точкой пересечения истинных кривых спроса и предложения, которых мы не знаем – является точным решением соответствующей системы уравнений. То, что точки (P_t, Q_t) не совпадают с точкой E , объясняется наличием случайных членов (шума) в обоих уравнениях. Если бы их не было, то в каждый момент времени t

наблюдаемое равновесие совпало бы с E . Тем не менее, наблюдаемые точки равновесия близки к E . Итак, чтобы оценить уравнения спроса и предложения, у нас есть информация только о точках равновесия в каждый момент времени. Понятно, что этой информации явно недостаточно для оценки кривых. Оценка отдельно кривых спроса и предложения возможна только из-за наличия в обоих уравнениях случайного члена.

КМНК: Модель, которую мы рассматриваем, является примером модели, в которой оба уравнения неидентифицируемы. Любая пара кривых спроса и предложения, пересекающихся в точке E , подойдут на роль «истинных» кривых спроса и предложения. При помощи КМНК мы получим две оценки для четырех параметров - ?

Что же касается первой из рассмотренных моделей (Модель 1), ее тоже можно проиллюстрировать графически.

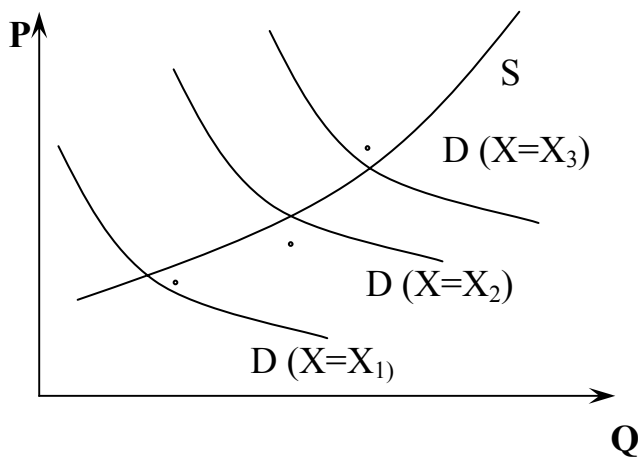


Рис. 2. Кривая спроса идентифицируема.

В этом случае мы по-прежнему в каждый момент времени на рынке наблюдается равновесие и P_t и Q_t — равновесные значения цены на товар и количество товара. Но теперь мы наблюдаем и значение дохода в момент времени t — X_t . Поэтому мы не можем изобразить одну кривую спроса. Если значение дохода варьируется во времени, то кривая спроса так же будет перемещаться по плоскости со временем. Кривая предложения у нас по-прежнему одна (Экономический смысл такой ситуации спросить у студентов). •

- наблюдаемые точки равновесия для различных уровней дохода. Точки равновесия расположены вдоль кривой предложения, что позволит нам ее оценить именно за счет движения равновесия P и Q . Изменение дохода от одного момента времени к другому существенно необходимо для идентификации уравнения предложения, т. е. условие $\gamma \neq 0$ существенно. Выполнение же ограничения $\gamma = 0$ будет препятствовать идентификации уравнения предложения.

Сверхидентифицируемость.

Модель 3. Опять рассмотрим равновесную модель спроса-предложения на рынке определенного товара. Только теперь добавим в уравнение предложения еще одну переменную, любую по смыслу.

$$\text{уравнение предложения: } Q_t = \alpha + \beta P_t + u_t^S$$

$$\text{уравнение спроса: } Q_t = \gamma + \delta P_t + \varepsilon X_t + \rho W_t + u_t^D.$$

В этом уравнении у нас по-прежнему две эндогенные переменные, а предопределенных переменных теперь две – X и W . В этом случае кривая спроса перемещается как результат вариации двух переменных. Уравнение предложения переопределено. Наличие двух экзогенных переменных, которые не присутствуют в уравнении предложения, влечет за собой наличие, по крайней мере, двух способов получения оценок структурных параметров уравнения предложения. Причем уравнение спроса по-прежнему неопределенно.

Уравнение в приведенной форме для нашей модели:

$$P_t = \frac{\gamma - \alpha}{\beta - \delta} + \frac{\varepsilon}{\beta - \delta} X_t + \frac{\rho}{\beta - \delta} W_t + \frac{u_t^D - u_t^S}{\beta - \delta}$$

$$Q_t = \frac{\gamma}{\beta - \delta} + \frac{\delta}{\beta - \delta} X_t + \frac{\rho}{\beta - \delta} W_t + \frac{\gamma}{\beta - \delta}$$

Перепишем уравнение следующим образом:

$$P_t = \alpha' + \beta' X_t + \gamma' W_t + v_t^P$$

$$Q_t = \delta' + \varepsilon' X_t + \phi' W_t + v_t^Q.$$

Оценим это уравнение. Получим a', b', c', d', e', f' . Всего 6 соотношений для 5 переменных. Как можно видеть, e'/b' и f'/c' дают оценку β . Они могут совпасть только случайным образом. Мы получили две разные оценки для одного параметра. Аналогичная ситуация и для α . Уравнение предложения переопределено. На оценку же уравнения спроса у нас осталось два соотношения на четыре переменные. Уравнение спроса по-прежнему неопределенно.

Можно предложить модель равновесия спроса-предложения, в которой оба уравнения идентифицируемы. Например, следующую (**Модель 4**):

$$\text{уравнение предложения: } Q_t^S = \alpha + \beta P_t + \gamma T_t + u_t^S$$

$$\text{уравнение спроса: } Q_t^D = \delta + \varepsilon P_t + \phi X_t + u_t^D$$

$$Q_t^S = Q_t^D,$$

где T_t – температура воздуха.

Упражнение. Придумать экономическую интерпретацию модели. Убедиться в ее идентифицируемости через приведенные уравнения и через геометрическую интерпретацию.

Из всего вышеизложенного возникает следующее желание. Нам хотелось бы иметь некоторый критерий идентифицируемости уравнения. Вообще-то, если наша цель – прогнозирование, то мы можем ограничиться рассмотрением приведенных уравнений, не заботясь об идентификации, но если наша цель – именно идентификация модели, то такой критерий необходим. Одно из условий идентификации может быть выражено следующим образом. Если уравнение в СОУ идентифицируемо, то число предопределенных переменных, не присутствующих в уравнении, должно быть не меньше (больше или равно) числу входящих в уравнение эндогенных переменных минус 1. В это число

входят как левосторонние, так и правосторонние эндогенные переменные. Это условие необходимое, но не достаточное. Можно привести пример модели, в которой уравнение удовлетворяет условию идентификации, но на самом деле уравнение неидентифицируемо. Тем не менее, необходимое и достаточное условие идентифицируемости уравнения существует, однако, его изложение требует сложного аппарата из теории матриц.

Методы оценивания систем одновременных уравнений.

1. Стандартный метод оценивания уравнений со случайным членом, коррелирующим с регрессорами – метод инструментальных переменных (МИП). Проблема в случае, когда регрессор коррелирует с остаточным членом следующая. Рассмотрим стандартную парную линейную модель:

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad (1.5),$$

где X и ε коррелированы. В этом случае:

$$\hat{\beta}_{МНК} = \frac{\frac{\sum_{i=1}^N X_i Y_i}{N} - \bar{X}\bar{Y}}{\frac{\sum_{i=1}^N X_i^2}{N} - \bar{X}^2} = \frac{Cov(X, Y)}{\hat{\sigma}_X^2} = \frac{Cov(X, \alpha + \beta X + \varepsilon)}{\hat{\sigma}_X^2} = \beta + \frac{Cov(X, \varepsilon)}{\hat{\sigma}_X^2}.$$

Возьмем математическое ожидание обеих частей, получим, что оценка $\hat{\beta}_{МНК}$ будет смещена. Что со состоятельностью?

В этом случае $Cov(X, \varepsilon)$ не будет стремиться к нулю по вероятности даже в больших выборках, поскольку $Cov(X, \varepsilon)$ является состоятельной оценкой теоретической ковариации м/у X и ε , а она не равна нулю из-за того, что X и ε коррелируют. Таким образом, МНК-оценка β смещена и несостоятельна, т. е. Смещение не элиминируется с ростом выборки. Направление смещения зависит от характера зависимости между X и ε .

В этой ситуации, как было сказано выше, на помощь приходит метод инструментальных переменных. Он может принести пользу не только в случае

СОУ, но и во многих других случаях, это достаточно мощный метод решения некоторых возникающих проблем, так что мы уделим ему некоторое время.

Суть метода – частичная замена непригодной объясняющей переменной той переменной, которая некоррелирована со случайным членом. Для иллюстрации вернемся к модели (1.5). Пусть нам удалось найти переменную Z , которая коррелирует с X , но некоррелирует с ε . Тогда оценка β по методу ИП определяется следующим образом:

$$\hat{\beta}_{ИП} = \frac{\frac{\sum_{i=1}^N Z_i Y_i}{N} - \bar{Z}\bar{Y}}{\frac{\sum_{i=1}^N X_i Z_i}{N} - \bar{X}\bar{Z}} = \frac{Cov(Z, Y)}{Cov(Z, X)} = \frac{Cov(Z, \alpha + \beta X + \varepsilon)}{Cov(Z, X)} = \beta + \frac{Cov(Z, \varepsilon)}{Cov(Z, X)}.$$

Когда эта оценка состоятельна:

$$p \lim \hat{\beta}_{ИП} = \beta + \frac{p \lim Cov(Z, \varepsilon)}{p \lim Cov(Z, X)} = \beta + \frac{0}{?} = \begin{cases} 0, & \text{если } p \lim Cov(Z, X) \neq 0, \\ ?, & \text{если } p \lim Cov(Z, X) = 0. \end{cases}$$

Для того чтобы оценка $\hat{\beta}_{ИП}$ была состоятельной, необходимо, чтобы $p \lim Cov(Z, X) \neq 0$, т. е. необходимо наличие «сильной» корреляции между переменными X и Z .

Трудность, возникающая при применении данного метода – отыскание переменных, пригодных к роли инструментов для X . Они должны быть сильно коррелированы с X и совсем некоррелированы с ε . К счастью, для некоторых случаев разработаны приемы, которые позволяют добиться успеха.

Вернемся к СОУ. В этом случае сама структура модели подсказывает нам, что в качестве инструментальных переменных можно взять predetermined переменные. Они коррелируют с эндогенными переменными, поскольку являются частью модели. Тот факт, что они являются predetermined, гарантирует (по определению) их некоррелированность со случайным членом.

Рассмотрим Модель 1:

$$Q_t = \alpha + \beta P_t + \gamma X_t + u_t^D - \text{уравнение спроса,}$$

$$Q_t = \delta + \varepsilon P_t + u_t^S - \text{уравнение предложения. (Рыночное равновесие)}$$

В этой модели одна предопределенная переменная – X . Ее мы и возьмем в качестве инструментальной для P в уравнении предложения. Получим состоятельные оценки коэффициентов уравнения предложения. Что же касается уравнения спроса, то мы не можем взять в качестве инструментальной для P ту же переменную X , поскольку эта переменная уже присутствует в правой части уравнения. Использование X в качестве инструментальной приведет к совершенной мультиколлинеарности. Поэтому уравнение спроса останется неидентифицируемым.

Рассмотрим Модель 4. В этой модели у нас две предопределенные переменные – X и T . Переменную X мы можем использовать в качестве инструментальной для P в уравнении спроса, а переменную T – в уравнении предложения. Получим состоятельные оценки коэффициентов обоих уравнений.

Рассмотрим модель 3. В этой модели у нас тоже две предопределенные переменные – X и Z . Обе они фигурируют в уравнении спроса, поэтому мы не можем использовать ни одну из них в качестве инструментальной для P в этом уравнении. Оно по-прежнему остается неопределенным. Зато в качестве инструмента для P в уравнении предложения мы можем использовать как X , так и Z . В обоих случаях получим состоятельные оценки α и β . Какую из них выбрать? Вопрос студентам. Можно, конечно посчитать коэффициент корреляции между P и X и между P и Z и выбрать ту, значение которого для нее больше. Я бы взяла линейную комбинацию этих переменных. Построение такой инструментальной переменной – суть двухшагового метода наименьших квадратов, который мы рассмотрим ниже.

Итак, метод инструментальных переменных может нам помочь в случае идентифицируемых и сверхидентифицируемых уравнений. В случае сверхидентифицируемых уравнений возникает проблема выбора инструмента.

2. КМНК – проблема идентифицируемости. Если интересует прогноз значений эндогенных переменных, то можно применять.

Упражнение. Показать, что в случае идентифицируемого уравнения КМНК-оценки совпадут с оценками по методу инструментальных переменных.

Из упражнения следует, что КМНК-оценки для идентифицируемых уравнений, хотя в общем случае могут быть смещены ($M(\frac{X}{Y}) \neq \frac{M(X)}{M(Y)}$), но состоятельны.

3. Двухшаговый метод наименьших квадратов (ДМНК).

ДМНК представляет собой очень полезную процедуру оценивания структурных уравнений модели.

Рассмотрим Модель 3. В этой модели уравнение предложения переопределено. В МИП у нас возникла проблема выбора инструмента для P в уравнении предложения из двух подходящих переменных. Этой проблемы не возникает при использовании ДМНК. Идея метода – вместо отдельного использования подходящих для инструмента переменных построить их линейную комбинацию таким образом, чтобы максимизировать значение коэффициента корреляции между этим новым инструментом и заменяемой переменной (в нашем случае это P). Обозначим такую комбинацию Z : $Z = \theta_0 + \theta_1 X + \theta_2 W$. Эта задача не так сложна, как кажется на первый взгляд. Наиболее разумным здесь является взять в качестве такого инструмента прогнозные значения P в регрессии P на X и W , т.е. $P = \theta_0 + \theta_1 X + \theta_2 W + u$ и $\hat{P}_i = \hat{\theta}_0 + \hat{\theta}_1 X_i + \hat{\theta}_2 W_i$. Беря в качестве инструмента \hat{P} , мы, таким образом, удовлетворим обоим требованиям, накладываемым на желательную комбинацию. Строя регрессию P на имеющиеся переменные, мы максимизируем R^2 , тем самым максимизируя коэффициент корреляции между P и \hat{P} , что нам и нужно.

Итак, ДМНК состоит из следующих двух шагов:

1. Построение инструмента для эндогенной переменной. Для этого нам надо осуществить регрессию по методу наименьших квадратов этой переменной на все предопределенные переменные в модели и вычислить

прогнозные значения для этой переменной. По построению новый инструмент и случайные члены исходной модели не будут коррелировать (строго говоря, это справедливо лишь для достаточно больших выборок, так что здесь мы можем завести разговор о состоятельности ДМНК).

Таким образом, на первом шаге мы построили новую переменную, которая линейно связана со всеми предопределенными переменными и очищена от корреляции с ошибками во всех уравнениях структурной модели.

2. Эндогенная переменная в структурном уравнении заменяется инструментальной переменной, полученной на первом шаге. Вычисляются оценки коэффициентов структурного уравнения по обычному методу наименьших квадратов. Эти оценки состоятельны.

ДМНК справляется проблемой переопределенности. В случае же идентифицируемости оценки по ДМНК совпадают с оценками КМНК и МИП.

Упражнение. Показать это для **Модели 3**.

Что же в случае неидентифицируемости происходит с ДМНК. Вспомним модель 1. Если мы попытаемся осуществить регрессию P на X , при подстановке прогнозных значений P в уравнение спроса получим полную мультиколлинеарность. Эта проблема делает невозможным дальнейшие вычисления.

Кроме рассмотренных методов, существуют и другие методы оценивания СОУ: трехшаговый метод наименьших квадратов, предложенный Зельнером и Тейлом. Этот метод предназначен для оценивания всех уравнений модели и в определенных обстоятельствах может оказаться асимптотически эффективнее двухшагового метода. Метод требует трудоемких вычислений и реже других применяется для оценки систем одновременных уравнений. Кроме этого, ранее мы рассматривали модели с ошибкой, удовлетворяющей условиям Гаусса-Маркова. В случае присутствия автокорреляции в ошибках, лаговые переменные уже не будут предопределенными. Автокорреляция достаточно часто присутствует во временных рядах. При этом ни МНК, ни ДМНК не будут состоятельными. Для такого случая (лаговая эндогенная переменная +

автокорреляция) предложена альтернативная процедура получения состоятельных оценок. В нашем курсе мы их рассматривать не будем.

2. СИСТЕМЫ ВНЕШНЕ НЕ СВЯЗАННЫХ УРАВНЕНИЙ.

(seemingly unrelated model – SUR).

Рассмотрим следующий пример. Пусть исследуется зависимость между расходами на некоторые товары (Y) и доходами домашних хозяйств (X) или между

В общем случае мы имеем систему из m уравнений:

$$Y_1 = \beta_{01} + \beta_{11}X_{11} + \dots + \beta_{1k_1}X_{1k_1} + \varepsilon_1$$

...

$$Y_m = \beta_{0m} + \beta_{m1}X_{m1} + \dots + \beta_{mk_m}X_{mk_m} + \varepsilon_m$$

Модель записана в векторной форме, т. е. Y_i , X_{ij} и ε_i есть вектора, состоящие из N компонент. Ошибка в каждом уравнении системы удовлетворяет условиям Гаусса-Маркова, однако, ошибки разных уравнений коррелируют между собой следующим образом:

$$M(\varepsilon_{is}\varepsilon_{jt}) = \begin{cases} \sigma_{ij}, t = s; \\ 0, t \neq s. \end{cases}$$

Если данные имеют структуру временных рядов, то говорят, что ошибки, относящиеся к разным индивидуумам коррелируют в один и тот же момент времени и не коррелируют для разных моментов времени.

Это предположение может и не соответствовать действительности. Например, я пошла сегодня в магазин и приобрела там очень хорошие босоножки. Моя подруга, узнав об этом, на следующий день пошла и купила такие же. Наши ошибки для разных моментов времени уже будут коррелировать.

Если случайные члены этих двух уравнений не коррелируют друг с другом, то мы можем получить эффективные оценки, используя МНК для каждого уравнения по отдельности, используя N наблюдений для каждой регрессии. Однако, если мы предположим коррелированность ошибок, то мы можем получить более эффективные оценки, рассматривая эти уравнения как систему или одно уравнение с Nm наблюдениями и $k_1 + \dots + k_m$ регрессорами и используя для оценивания обобщенный метод наименьших квадратов.

Рассмотрим структуру матрицы ковариаций ошибок. Она имеет следующую структуру:

$$\Omega = \begin{pmatrix} \sigma_{11} & & \sigma_{12} & & \dots & \sigma_{1N} \\ & \dots & & \dots & & \dots \\ & & \sigma_{11} & & \sigma_{12} & & \sigma_{1N} \\ \sigma_{21} & & \sigma_{22} & & \dots & \sigma_{2N} \\ & \dots & & \dots & & \dots \\ & & \sigma_{21} & & \sigma_{22} & & \sigma_{2N} \\ & & & \dots & & & \dots \\ \sigma_{N1} & & \sigma_{N2} & & \dots & \sigma_{NN} \\ & \dots & & \dots & & \dots \\ & & \sigma_{N1} & & \sigma_{N2} & & \sigma_{NN} \end{pmatrix}$$

Оценка по обобщенному методу наименьших квадратов записывается следующим образом: $\beta_{ОМНК} = (X\Omega^{-1}X)^{-1}X\Omega^{-1}Y$. В общем случае эта оценка не будет совпадать с оценками, полученными по методу наименьших квадратов для каждого уравнения в отдельности. Они совпадают, если:

- 1) уравнения действительно не связаны друг с другом, т. е. $\sigma_{ij} = 0, i \neq j$;
- 2) все уравнения имеют один и тот же набор объясняющих переменных.

Эффективность этой оценки по сравнению с МНК-оценками тем выше, чем сильнее корреляция между ошибками.

Оценку матрицы Ω можно получить следующим образом. Применяем обычный МНК к каждому уравнению, получая векторы остатков e для каждого

уравнения. В качестве оценки σ_{ij} берем величину $\frac{\sum_{k=1}^N e_{ik} e_{jk}}{N}$. Можно показать, что эти оценки являются состоятельными.

17. Литература.

1. Магнус Я.Р., Катышев К.К., Пересецкий А.А. Эконометрика. Начальный курс. Дело. 2005.
2. Доугерти К. Введение в эконометрику. Пер. с англ. М.:ИНФРА_М, 1997.
3. Бородич С. А. Эконометрика. Минск, Новое знание, 2001.
4. Джонстон Дж. Эконометрические методы. М., Статистика, 1980.
5. Елисеева И. И. Эконометрика. М., Финансы и статистика, 2001.
6. Елисеева И. И., Практикум по эконометрике. М., Финансы и статистика, 2001
7. Кремер Н. Ш., Путко Б. А. Эконометрика. М. ЮНИТИ, 2002
8. Э. Берндт. Практика эконометрики: классика и современность. М. ЮНИТИ, 2005
9. R.S. Pindyck & D.L. Rubinfeld, Econometric Models and Economic Forecasts, 3rd edition, McGraw Hill, 1991.
10. W.H.Greene, Econometric Analysis, 3rd edition, Prentice Hall, 1997.
11. J.Johnston, J.DiNardo, Econometrics Methods, 4th edition, McGraw-Hill, 1997.