**1. Эконометрика как наука. Понятие взаимосвязи между случайными величинами. Ковариация и коэффициент корреляции. Эконометрическая модель**

«Эконометрика позволяет проводить количественный анализ реальных экономических явлений, чтобы охарактеризовать их типичное поведение. Выделяют 2 вида статистического анализа зависимостей:

1. Корреляционный анализ – позволяет оценить значимость и силу взаимосвязи, без указания вида зависимости

2. Регрессионный анализ – позволяет оценить и проанализировать формулу зависимости одной переменной от других.

Коэффициент ковариации показывает направленность взаимосвязи между двумя переменными. Является состоятельной, смещенной оценкой.

Коэффициент корреляции показывает направленность и силу(!) взаимосвязи между двумя переменными. Значения коэффициента корреляции:

- Близкие к 0 – свидетельствуют об отсутствии линейной взаимосвязи

- Близкие к (+1) – о сильной прямой взаимосвязи

- Близкие к (–1) – о сильной обратной взаимосвязи

Эконометрическая модель – это математическое описание экономического явления, отражающее наиболее важные его черты. Модель упрощает, идеализирует изучаемое явление. О правильности построенной модели можно судить по близкому соответствию результатов моделирования и фактических данных.

Общим моментом для любых эконометрических моделей является разбиение зависимой переменной на две части: объясненную и случайную, и є – случайная составляющая. Y=f(x)+E

**2. Понятие регрессии. Модель парной линейной регрессии. Диаграмма рассеяния и проблема выбора линии регрессии. Природа случайного члена регрессии**

Регрессионный анализ – позволяет оценить и проанализировать формулу зависимости одной переменной от других.

Линейная парная модель

Уравнение регрессии:

y = β0 + β1x + ε

β1 – показывает, на сколько изменится значение зависимой переменной y при изменении объясняющей переменной x на единицу.

β0 – показывает среднее значение зависимой переменной y при нулевом значении объясняющей переменной x. Не всегда имеет экономический смысл. (компонент)

Проблема выбора линии заключаеться в том, что надо её максимально описать зависимость, чтобы по минимуму были отклонения.

Природа случайного члена регрессионной модели

1. Невключение объясняющих переменных

2. Неправильная функциональная спецификация модели

3. Случайность поведения рассматриваемых объектов

4. Ошибки измерения

**3. Метод наименьших квадратов. Вывод формул метода наименьших квадратов для парного случая. Суть метода, графическое представление, условия применения**

Метод наименьших квадратов — метод нахождения оптимальных параметров линейной регрессии, таких, что сумма квадратов ошибок (регрессионных остатков) минимальна.

Метод заключается в минимизации евклидова расстояния между двумя векторами -- вектором восстановленных значений зависимой переменной и вектором фактических значений зависимой переменной.

Применение. Задача метода наименьших квадратов состоит в выборе вектора , минимизирующего ошибку .

Метод наименьших квадратов имеет следующие преимущества:

- не требуется знания закона распределения случайного возмущения

- дает оценки по крайней мере состоятельные

- в случае нормального распределения случайного возмущения оценки параметров линейной модели несмещенные и эффективные

Formula: 

4. Свойства оценок коэффициентов регрессии, полученных с помощью метода наименьших квадратов. Теорема Гаусса-Маркова - формулировка, смысл теоремы. Условия Гаусса-Маркова и последствия их нарушения

Теорема Гаусса—Маркова

оценки по обычному методу наименьших квадратов являются не только несмещенными оценками коэффициентов регрессии, но и наиболее эффективными в том случае, если выполнены условия Гаусса—Маркова. С другой стороны, если условия Гаусса—Маркова не выполнены, то, вообще говоря, можно найти оценки, которые будут более эффективными по сравнению с оценками, полученными обычным методом наименьших квадратов. В данной работе не приводится общее рассмотрение этих вопросов. Тем не менее в том случае, если условия Гаусса—Маркова для остаточного члена выполнены, коэффициенты регрессии, построенной обычным методом наименьших квадратов, будут наилучшими линейными несмещенными оценками (best linear unbiased estimators, или BLUE): несмещенными, как уже было показано; линейными, так как они являются линейными функциями значений у; наилучшими, так как они являются наиболее эффективными в классе всех несмещенных линейных оценок. Теорема Гаусса—Маркова доказывает это (краткое изложение, не использующее матричной алгебры, дано в работе Дж. Томаса

для того чтобы регрессионный анализ, основанный на обычном методе наименьших квадратов, давал наилучшие из всех возможных результаты, случайный член должен удовлетворять четырем условиям, известным как условия Гаусса—Маркова.

1-е условие Гаусса—Маркова: E(Ut) = 0 для всех наблюдений. Первое условие состоит в том, что математическое ожидание случайного члена в любом наблюдении должно быть равно нулю. Иногда случайный член будет положительным, иногда отрицательным, но он не должен иметь систематического смещения ни в одном из двух возможных направлений. Vipolnjaetsja avtomaticeski,esli urava soderzit konstantu

2-е условие Гаусса—Маркова: pop. var (u) постоянна для всех наблюдений. Второе условие состоит в том, что дисперсия случайного члена должна быть постоянна для всех наблюдений. Иногда случайный член будет больше, иногда меньше, однако не должно быть априорной причины для того, чтобы он порождал большую ошибку в одних наблюдениях, чем в других. Одна из задач регрессионного анализа состоит в оценке стандартного отклонения случайного члена. Если рассматриваемое условие не выполняется, то коэффициенты регрессии, найденные по обычному методу наименьших квадратов, будут неэффективны. Narushenie privodit k geteroskedasticnosti

3- е условие Это условие предполагает отсутствие систематической связи между значениями случайного члена в любых двух наблюдениях. Например, если случайный член велик и положителен в одном наблюдении, это не должно обусловливать систематическую тенденцию к тому, что он будет большим и положительным в следующем наблюдении (или большим и отрицательным, или малым и положительным, или малым и отрицательным). Случайные члены должны быть абсолютно независимы друг от друга. Narushenie privodit k avtokorreljacii

4-е условие случайный член должен быть распределен независимо от объясняющих переменных В большинстве глав книги мы будем в сущности использовать более сильное предположение о том, что объясняющие переменные не являются стохастическими, т. е. не имеют случайной составляющей. Значение любой независимой переменной в каждом наблюдении должно считаться экзогенным, полностью определяемым внешними причинами, не учитываемыми в уравнении регрессии. Если это условие выполнено, то теоретическая ковариация между независимой переменной и случайным членом равна нулю. Дело в том, что если случайный член и нормально распределен, то так же будут распределены и коэффициенты регрессии.

Предположение о нормальности основывается на центральной предельной теореме. В сущности, теорема утверждает, что если случайная величина является общим результатом взаимодействия большого числа других случайных величин, ни одна из которых не является доминирующей, то она будет иметь приблизительно нормальное распределение, даже если отдельные составляющие не имеют нормального распределения.

**5. Понятие качества регрессии, коэффициент детерминации, скорректированный коэффициент детерминации**

Коэффициент детерминации (R2) Это оценка качества - это квадрат множественного коэффициента корреляции. Он показывает, какая доля [дисперсии](http://ru.wikipedia.org/wiki/%D0%94%D0%B8%D1%81%D0%BF%D0%B5%D1%80%D1%81%D0%B8%D1%8F) результативного признака объясняется влиянием независимых переменных. Формула для вычисления коэффициента детерминации:

где yi — выборочные данные, а fi — соответствующие им значения модели.

Коэффициент принимает значения из интервала [0;1]. Чем ближе значение к 1 тем ближе модель к эмпирическим наблюдениям. R2<50% Построенные при таких условиях регрессионные модели имеют низкое практическое значение.

Значение R2 монотонно возрастает с ростом числа переменных (регрессоров) в регрессии, что зачастую не означает улучшения качества предсказания. Потому правильнее использовать скорректированный коэффициент детерминации, учитывающий число использованных регрессоров и корректировать коэффициент множественной детерминации на потерю степеней свободы вариации.

1. **Модель множественной регрессии. Смысл коэффициентов множественной регрессии. Матричный вид регрессии, метода наименьших квадратов и теоремы Гаусса-Маркова. Вывод формул метода наименьших квадратов в матричном виде**

Идея множественной регрессии состоит в том, что зависимая переменная определяется более чем одной объясняющей переменной. Общий вид множественной регрессии:



Коэф. Регрессии показывают насколько изменится значение зависимой переменной y , если значение соответствующей независимой переменной изменится на 1, при условии, что все остальные переменные останутся неизменными.

Матричная форма записи

Пусть имеется выборка из п наблюдений, а модель включает k peгреccopов и константу. Введем обозначения:

Y=- вектор-столбец наблюдений (размерности п)

X=— матрица значений регрессоров (размерности п на k+1)

- вектор-столбец неизвестных параметров, (размерности k+1)

- вектор-столбец случайных ошибок, (размерности п)

Тогда множественную линейную регрессионную модель можно записать, в матричной форме:



Метод наименьших квадратов

Необходимо найти методом наименьших квадратов оценки неизвестных параметров β.

Они определяются исходя из условия минимизации суммы квадратов остатков по компонентам вектора β.



**7. Проверка гипотез в модели регрессии. Проверка гипотезы о коэффициенте регрессии. Значимость коэффициента, p-значение. Доверительный интервал для коэффициентов регрессии**

Часто на практике необходимо ответить на вопрос: значимо ли отличается коэффициент регрессии от определенного значения С.

Схема тестирования гипотезы (критерий Стьюдента) выглядит следующим образом:

Н0: β = С – нулевая гипотеза

H1: β ≠ С – альтернативная гипотеза

* Вычисляются МНК-оценки коэффициентов регрессии и их стандартные ошибки
* Рассчитывается наблюдаемое значение статистики t: tнабл
* Выбирается требуемый уровень надежности γ (95%, 99%,99,9%) и находится критическое значение статистики Стьюдента с соответствующим количеством степеней свободы: tкрит
* Если |tнабл| > tкрит (по модулю), то нулевая гипотеза отвергается в пользу альтернативной, если нет – нулевая гипотеза не отвергается.

P-значение

Часто удобнее рассматривать непосредственно вероятность того, что наблюдаемое значение не превысит критическое: P-значение или p-value – это вероятность принятия гипотезы, т.е. если p-значение < уровня значимости, который равен 0,01; 0,05 или 0,10 (чаще всего это 0,05), то нулевая гипотеза Н0 – отвергается. Часто проверяется гипотеза H0: β = 0, которую в этом случае называют гипотезой о незначимости коэффициента.

Доверительные интервалы для коэффициентов регрессии

Доверительный интервал – это вычисленный на данных интервал, который с заданной вероятностью покрывает интересующий нас неизвестный параметр генеральной совокупности. В его основе используется стандартная ошибка оцениваемого параметра.

Приведенный интервал называют γ-процентным (90-, 95- или 99-процентным) доверительным интервалом для истинного значения коэффициента β.

1. **Проверка гипотез в модели регрессии. Общая линейная гипотеза. Гипотеза о качестве регрессионной модели. Сравнение длинной и короткой регрессий. Тест Чоу**

Общая линейная гипотеза

Гипотезу Hβ = r называют общей линейной гипотезой. Линейные гипотезы обычно вытекают из знаний экспериментатора или его предположений относительно возможных моделей. Проблема сравнения двух подвыборок является частным случаем общей линейной гипотезы.

Гипотеза о качестве регрессионной модели ни один регрессор не оказывает влияние на зависимую переменную. Гипотеза о равенстве нулю каждого из коэффициентов регрессии в отдельности Н0: bi=0. Для этого вычисляется Р-значение. Здесь следует подчеркнуть, что принятие Hо (высокое Р-значение) еще не говорит о том, что рассматриваемый признак xi нужно исключить из модели. Этого делать нельзя, поскольку суждение о ценности данного признака может выноситься, исходя из анализа совокупного взаимодействия в модели всех признаков. Поэтому высокое p-значение служит только «сигналом» о возможной неинформативности того или иного признака. Для проверки значимости модели регрессии используется F-критерий Фишера.

Сравнение длинной и короткой регрессий. Рассчитываем F-статистику и если это значение превышает критическое – делаем выбор в пользу неограниченной (длинной) регрессии, иначе – в пользу ограниченной (короткой).

Тест Чоу. Тест на равенство коэффициентов регрессии в двух выборках, называют тестом Чоу. Нулевая гипотеза проверяется с помощью F-статистики для гипотезы о том, что коэффициенты при всех добавленных переменных равны нулю.Выборку делят на части, у различных интервалов различный У, строят много интервалов, который наиболее значим по тесту Чоу, где используют сумму квадратов остатков модели для н-подмножеств.

**9. Нелинейные регрессионные модели. Построение нелинейной модели. Оценивание эластичности с помощью регрессионной модели**

Pri mnogoobrazii ekonom rpocessov casto zavisimostj budet nelinejnoj (spros,elasticnostj). Для оценки параметров нелинейных моделей используются два подхода. Первый подход основан на линеаризации модели и заключается в том, что с помощью подходящих преобразований исходных переменных исследуемую зависимость представляют в виде линейного соотношения между преобразованными переменными. Второй подход обычно применяется в случае, когда подобрать соответствующее линеаризующее преобразование не удается. В этом случае применяются методы нелинейной оптимизации на основе исходных переменных. Таким образом, функции, которые показывают изменение одной переменной от другой в процентах или в несколько раз являются функциями, отражающими эластичность.

1. **Сравнение регрессионных моделей. Тест Бокса-Кокса, процедура Зарембки**

1. Критерий Хоэла. Estj dve pohozie modeli, Строят тестовую линейную зависимость в виде уравнения в параметрической форме. Проверка сводится к оценке в уравнении углового коэффициента . Если  значимо положителен (1), от модели отказываются в пользу второй модели . Если  незначимо положителен (<1), то нельзя определить, какая из моделей лучше. Критерий Хоэла называется несимметричным, так как он может использоваться только при >0.

2. Критерий Вильяма и Клута. Для сравнения двух регрессионных моделей, которые, по крайней мере, первоначально представляются равноценными, можно использовать симметричный критерий Вильяма и Клута. Проверка осуществляется путем оценки параметра  в тестовом уравнении корреляция регрессия гетероскедастичность логарифмирование

Sravnivaem aljternativnie modeli po R2, esli odinakovie, to Тогда следует применять стандартную процедуру в виде теста Бокса — Кокса. Если нужно всего лишь сравнить модели с использованием результативного фактора и его логарифма в виде варианта зависимой переменой, то применяют вариант теста Зарембки- eto сравнения среднеквадратичной ошибки (СКО) в линейной и логарифмической моделях. Соответствующая процедура включает следующие шаги.

1. Вычисляется среднее геометрическое значений у в выборке, совпадающее с экспонентой среднего арифметического значений логарифма от у.
2. Пересчитываются наблюдения у таким образом, что они делятся на полученное на первом шаге значение.
3. Оценивается регрессия для линейной модели с использованием пересчитанных значений у вместо исходных значений у и для логарифмической модели с использованием логарифма от пересчитанных значений у. Теперь значения СКО для двух регрессий сравнимы, и поэтому модель с меньшей суммой квадратов отклонений обеспечивает лучшее соответствие с истинной зависимостью наблюденных значений.
4. Для проверки того, что одна из моделей не обеспечивает значимо лучшее соответствие, можно использовать произведение 1/2 числа наблюдений на логарифм отношения значений СКО в пересчитанных регрессиях с последующим взятием абсолютного значения этой величины. Такая статистика имеет распределение χ2 с одной степенью свободы (обобщение нормального распределения).

# 11. Спецификация регрессионной модели. Ошибки спецификации модели – включение незначимых и невключение значимых переменных. Замещающие переменные

Возможные ошибки спецификации регрессионной модели:

- Невключение значимых переменных

- Включение незначимых переменных

Невключение значимых переменных

• (–) Смещенность оценок коэффициентов регрессии

• (–) Смещенность оценки дисперсии ошибок регрессии

• (+) Меньшая вариация оценок коэффициентов регрессии

Включение незначимых переменных

• (+) Несмещенность оценок коэффициентов регрессии

• (+) Несмещенность оценки дисперсии ошибок регрессии

• (–) Большая вариация оценок коэффициентов регрессии

Замещающие переменные, причины:

1. Необходимость показателя не была учтена при составлении выборки

2. Переменная трудноизмерима (например, уровень образования)

3. Сбор данных о переменной x1 требует значительных затрат

При оценивании модели без переменной x1 полученные оценки будут смешенными.

Последствия использования замещающих переменных:

1. Оценки коэффициентов при переменных x2,…, xk становятся несмещенными

2. Стандартные ошибки и t-статистики коэффициентов te ze

3. R2 имеет такое же значение, как и при оценивании с переменной x1

4. Коэффициент β1 нельзя оценить (оценивается только β1δ1), но его стандартная ошибка и t-статистика позволяет оценить значимость x1

5. Получить оценку свободного члена модели невозможно (но она часто и не особенно важна) последствия справедливы приблизительно

**12. Мультиколлениарность в регрессионной модели: понятие, причины, последствия**

Мультиколлинеарность — это понятие, которое используется для описания проблемы, когда нестрогая линейная зависимость между объясняющими переменными приводит к получению ненадежных оценок регрессии. Оценка любой регрессии будет страдать от нее в определенной степени, если только все независимые переменные не окажутся абсолютно некоррелированными.

Различные методы, которые могут быть использованы для смягчения мультиколлинеарности, делятся на две категории: к первой категории относятся попытки повысить степень выполнения четырех условий, обеспечивающих надежность оценок регрессии; ко второй категории относится использование внешней информации, но можно привнести или усилить автокорреляцию, но она может быть нейтрализована. Кроме того, можно привнести (или усилить) смещение, вызванное ошибками измерения, если поквартальные данные измерены с меньшей точностью, чем соответствующие ежегодные данные.

**13. Включение качественных признаков с регрессионную модель: фиктивные переменные, множественные фиктивные переменные, перекрестные фиктивные переменные**

Качественные факторы включаются в эконометрические модели с помощью фиктивных переменных. Под фиктивной переменной понимают переменную, которая равна 1 для конкретной части выборочной совокупности, и 0 - для оставшейся части.

Переменные, полученные путем перемножения с фиктивными, обычно называют перекрестными фиктивными переменными.

Множественные фиктивные переменные – фиктивные переменные, которые принима.т не 2, а более значений.

Достаточно часто качественные объясняющие переменные принимают не два, а несколько (m) значений.

**14. Прогнозирование с помощью регрессионной модели. Точечный и интервальный прогнозы**

Прогноз получают путем подстановки в регрессионное уравнение Y = a0 + a1x.

Здесь a0, a1 — параметры, которые оцениваются из статистических [данных](http://slovari.yandex.ru/dict/lopatnikov/article/lop/lop-0271.htm). Они называются коэффициентами регрессии.

В случае же совместного влияния на Y нескольких факторов (x1, x2, ..., xn) уравнение принимает вид

Y = a0 + a1x1 + ... + anxn.

В первом случае имеем парную Р., во втором — множественную) переменных. Результат представляет собой оценку среднего значения зависимой переменной при данных уровнях фактор-аргументов. Для уравнения регрессии обычно определяют доверительные интервалы, которые также можно использовать в прогнозировании.

точечный прогноз — прогноз, которым указывается единственное значение прогнозируемого [показателя](http://slovari.yandex.ru/dict/lopatnikov/article/lop/lop-1101.htm). Конкретное значение.

интервальный прогноз — прогноз, которым указывается не единственное значение прогнозируемого [показателя](http://slovari.yandex.ru/dict/lopatnikov/article/lop/lop-1101.htm) (или вектор значений), а некоторый интервал. Пример И. п.: “Население города N. составит в 2000 г. от 30 до 35 тыс. человек”.

**15. Нарушение условия гомоскедастичности: гетероскедастичность, автокорреляция**

Гетероскедастичность и автокорреляция это нарушения условия гомоскедастичности (второго и третьего условия Гаусса-Маркова). Обычно рассматривают два варианта нарушения условия гомоскедастичности:

1. Ошибки имеют различную дисперсию для различных наблюдений - Гетероскедастичность. Гетероскедастичность – «неодинаковый разброс» теоретическое распределение случайного члена является разным для различных наблюдений в выборке

Например, если рассматривать зависимость расходов на отдых от заработной платы, то логично предполагать, что разброс будет выше для более обеспеченных людей.

2. Ошибки имеют постоянную дисперсию, но неявляются независимыми Данное явление носит название автокорреляции

Общий случай простое преобразования для того, чтобы добиться гомоскедастичности В общем случае:

1. Рассчитывают МНК-оценки коэффициентов регрессии

2. Находят остатки ei и их квадраты ei2

3. Находят логарифмы квадратов остатков ln(ei2)

4. Рассчитывают регрессию ln(ei2)

5. Получают прогноз ln(ei2)прог

6. Находят веса наблюдений wi= exp(ln(ei2)прог)

7. Полученные веса wi используют во взвешенном методе наименьших квадратов

Последствия гетероскедастичности и автокорреляции приводит к неэффективности получаемых коэфициентов регресии (но они остаются несмещенны) и к неправильному расчету наблюдаемых t и F-статистик.

**16. Обобщенный метод наименьших квадратов, теорема Айткена**

Применение обычного метода наименьших квадратов при нарушении условия гомоскедастичности приводит к следующим отрицательным последствиям:

1. оценки неизвестных коэффициентов β неэффективны, то есть существуют другие оценки, которые являются несмещенными и имеют меньшую дисперсию.

2. стандартные ошибки коэффициентов регрессии будут занижены, а, следовательно, t -статистики – завышены, и будет получено неправильное представление о точности уравнения регрессии.

Обобщенный метод наименьших квадратов

Рассмотрим метод оценивания при нарушении условия гомоскедастичности, матрица имеет вид β= (ХТ Ω-1 Х)-1 ХТ Ω-1у

Расчёт неизвестных коэффициентов регрессии по данной формуле называют обобщённым методом наименьших квадратов (ОМНК).

Теорема Айткена: при нарушении предположения гомоскедастичности оценки, полученные обобщенным методом наименьших квадратов, являются несмещенными и наиболее эффективными (имеющими наименьшую вариацию). На практике матрица Ω практически никогда не известна. Поэтому часто пытаются каким-либо методом оценить оценки матрицы Ω и использовать их для оценивания. Этот метод носит название доступного обобщенного метода наименьших квадратов.

**17. Тесты на гетероскедастичность: Спирмена, Бреуша-Пагана, Уайта, Голдфельда-Квандта**

Ранговая корреляция. Тест ранговой корреляции Спирмена

* Ранг наблюдения переменной - номер наблюдения переменной в упорядоченной по возрастанию последовательности.
* Тест ранговой корреляции Спирмена тест на гетероскедастичность, устанавливающий, что стандартное отклонение остаточного члена регрессии имеет нестрогую линейную зависимость с объясняющей переменной.

При выполнении теста ранговой корреляции Спирмена предполагается, что дисперсия случайного члена будет либо увеличиваться, либо уменьшаться по мере увеличения x, и поэтому в регрессии, оцениваемой с помощью МНК, абсолютные величины остатков и значения х будут коррелированны. Данные по х и остатки упорядочиваются. Если предположить, что соответствующий коэффициент корреляции генеральной совокупности равен нулю, т.е. гетероскедастичность отсутствует, то коэффициент ранговой корреляции имеет нормальное распределение с математическим ожиданием 0 и дисперсией 1/(n - 1) в больших выборках.

Тест Голдфелда-Квандта

* Наиболее популярным формальным критерием является критерий, предложенный С. Голдфелдом и Р. Квандтом.
* При проведении проверки по этому критерию предполагается, что стандартное отклонение распределения вероятностей u пропорционально значению х в этом наблюдении.
* Предполагается также, что случайный член распределен нормально и не подвержен автокорреляции.
* Иными словами тест Голдфелда- Квандта - тест на гетероскедастичность, устанавливающий, что стандартное отклонение остаточного члена регрессии растет, когда растет объясняющая переменная.
* Все n наблюдений в выборке упорядочиваются по величине х, после чего оцениваются отдельные регрессии для первых n’ и для последних n’ наблюдений;
* Средние (n- 2n’) наблюдений отбрасываются.
* Если предположение относительно природы гетероскедастичности верно, то дисперсия и в последних n’ наблюдениях будет больше, чем в первых n’ , и это будет отражено в сумме квадратов остатков в двух указанных "частных" регрессиях.
* Обозначим суммы квадратов остатков в регрессиях для первых n’ и последних n’ наблюдений соответственно через RSS1, и RSS2,
* рассчитаем отношение RSS2/RSS1, которое имеет распределение F с (n’ - k - 1 ) и (n’ - k - 1 ) степенями свободы, где k -число объясняющих переменных в регрессионном уравнении.
* Метод Голдфелда-Квандта может также использоваться для проверки на гетероскедастичность при предположении, что σ , обратно пропорционально х,.
* Используется та же процедура, что и описанная выше, но тестовой статистикой теперь является показатель RSS1/RSS2, который вновь имеет F-распределение с ( n’- k - 1) и (n’ - k -1) степенями свободы.
* Обозначим суммы квадратов остатков в регрессиях для первых n’ и последних n’ наблюдений соответственно через RSS1, и RSS2,
* рассчитаем отношение RSS2/RSS1, которое имеет распределение F с (n’ - k - 1 ) и (n’ - k - 1 ) степенями свободы, где k -число объясняющих переменных в регрессионном уравнении
* Таким образом, тест Голдфелда-Квандта состоит из трех этапов:
1. все наблюдения в выборке упорядочиваются по возрастанию х.
2. берутся первые и последние n наблюдений (треть от всех), оцениваются две различные регрессии и находятся RSS1 и RSS2
3. Для отношения RSS2/RSS1, проводят тест Фишера с (n’ - k - 1) верхними и (n’ - k - 1) нижними степенями свободы, где k - количество объясняющих переменных в регрессиях.

Тест Бреуш-Пагана

Тест применим в предположении, что: Дисперсии зависят от некоторых дополнительных переменных:

1. Строится уравнение регрессии: и вычисляются остатки:

2. Вычисляют оценку дисперсии остатков:

3. Строят вспомогательное уравнение регрессии

4. Для вспомогательного уравнения регрессии определяют объясненную часть вариации RSS.

5. Находим тестовую статистику:

6. Если верна гипотеза H0: гомоскедастичность остатков, то статистика BP имеет распределение. Т.е. о наличии гетероскедастичности остатков на уровне значимости α свидетельствует:

Замечания

При гетероскедастичность может быть скорректирована:

Тест Уайта

Предполагается, что дисперсии связаны с объясняющими переменными в виде: 

Т.к. дисперсии неизвестны, то их заменяют оценками квадратов отклонений ei2.

Алгоритм применения (на примере трех переменных)

1. Строится уравнение регрессии: и вычисляются остатки.

2. Оценивают вспомогательное уравнение регрессии:

3. Определяют из вспомогательного уравнения тестовую статистику

4. Проверяют общую значимость уравнения с помощью критерия χ2.

Замечания

Тест Уайта является более общим чем тест Голдфелда-Квандта.

Неудобство использования теста Уайта: Если отвергается нулевая гипотеза о наличии гомоскедастичности то неясно, что делать дальше.

**18. Коррекция гетероскедастичности: логарифмирование, взвешенная регрессия, общий случай**

Часто наличие гетероскедастичности в остатках регрессии свидетельствует о неправельной спецификации модели.

Рассмотрим две модели – линейную

yi = ß0 + ß1xi + εi

и логарифмическую

yi = eßoxiß1eεi

В линейной модели случайный член присудствует в аддитивной форме, а в логарифмической модели – в мультипликативной.

Мультипликативная форма отражает более сильное влияние случайного члена для больших значений регрессоров и более слабое – для маленьких.

Следовательно, если в линейной модели наблюдается такой вид гетероскедастичности, то вполне возможно, что в логарифмической модели гетероскедастичности не будет.



Логарифмическая регрессия не всегда позволяет избавится от гетероскедастичности. Кроме того, логарифмическая модель не всегдя удовлетворяет целям исследования (требуется оценить зависимость в абсолютных величнах, а не эластичность)

В этих случаях используют другой подход – взвешенную регрессию.

Рассмотрим модель

yi = ß0 + ß1xi + εi

Пусть в моделе пресудствует гетероскедастичность

D(εi) = σi2

И нам известно точное значения дисперсий ошибок модели σi2

(далее идут формулы и решения, не думаю что их придется расписывать, поэтому не буду забивать ваши светлые головы всякими решениями)

После всех вычеслений...таким образом случайный член модели имеет постоянную дисперсию (по расчетам она равна 1 ), следовательно обычные МНК-оценки неизвестных коэфицентов будут несмещенными и эффективными.

На практике дисперсии ошибок почти никогда не бывает. Однако иногда можно предположить, что σi2 пропорциональны некоторой переменной zi.



Тогда в качестве весов наблюдений следует использовать величину 1/zi:



Дисперсия случайного члена такой модели также постоянна



Достаточно часто в качестве переменной, взаимосвязанной с дисперсией случайного члена можно использовать регрессор:

σi = λxi

в этом случае взвешенная модель имеет вид:



Коэфицент ß1 в преобразованной модели соответствует свободному члену.

# Общий случай

Подобрать простое преобразование для того, чтобы добиться гомоскедастичности удается не всегда.

В общем случае используют следующую процедуру

1. Расчитываются МНК-оценки коэффицентов регресии
2. Находят остатки еi и их квадраты
3. Находят логарифмы отстатков
4. Расчитывают регрессию
5. Плучают прогноз
6. Находят веса наблюдений wi
7. Полученные веса wi используют во взвешенном методе наименьших квадратов
8. **Коррекция гетероскедастичности: логарифмирование, взвешенная регрессия, общий случай**



Часто наличие гетероскедастичности в остатках регрессии свидетельствует о неправельной спецификации модели. Если в линейной модели наблюдается такой вид гетероскедастичности, то вполне возможно, что в логарифмической модели гетероскедастичности не будет.

Логарифмическая регрессия не всегда позволяет избавится от гетероскедастичности. Кроме того, логарифмическая модель не всегдя удовлетворяет целям исследования (требуется оценить зависимость в абсолютных величнах, а не эластичность)

В этих случаях используют другой подход – взвешенную регрессию. Пусть в моделе пресудствует гетероскедастичность D(εi) = σi2 После всех вычеслений случайный член модели имеет постоянную дисперсию, следовательно обычные МНК-оценки неизвестных коэфицентов будут несмещенными и эффективными. На практике дисперсии ошибок почти никогда не бывает.

Общий случай

Подобрать простое преобразование для того, чтобы добиться гомоскедастичности удается не всегда. В общем случае используют следующую процедуру

* Расчитываются МНК-оценки коэффицентов регресии
* Находят остатки еi и их квадраты
* Находят логарифмы отстатков
* Расчитывают регрессию
* Плучают прогноз
* Находят веса наблюдений wi
* Полученные веса wi используют во взвешенном методе наименьших квадратов

**20. Автокорреляция: понятие, виды, последствия**

Автокорреляция - случайные члены регрессии в разных наблюдениях являются зависимыми. Автокорреляция приводит к неэффективности получаемых МНК-оценок и к неправильному расчету наблюдаемых t и F-статистик и ошибочным решениям при тестировании гипотез. Первый тип автокорреляции – положительная автокорреляция.-это когда после положительных ошибок более вероятны положительные ошибки, после отрицательных – отрицательные. То есть ошибки имеют тенденция к сохранению своего знака. Противоположный случай –это отрицательной автокорреляцией: когда после положительных ошибок более вероятны отрицательные ошибки, после отрицательных – положительные.

Виды автокорреляции:

А. первого порядка: Ошибка зависит от ее значений в предыдущие p периодов времени и от случайного члена μt (называемого инновацией в момент времени t).

Автокорреляцией со скользящим средним q-oго порядка, обозначаемойMA(q), то есть ошибка в момент времени t зависит от инноваций в текущий и предыдущие q моментов времени. Автокорреляция со скользящим средним первого порядка, MA(1):

Последствия автокорреляции:

Потеря эффективности оценок

Смещение дисперсии

t- и F-статистики неправильные.

**20. Автокорреляция: тест Дарбина-Уотсона, исправление автокорреляции**

Обнаружение автокорреляции:

1. Графический метод.

2. Метод рядов.

3. Специальные тесты.

Большинство тестов на наличие автокорреляции в модели основаны на идеи: если корреляция есть у ошибок t, то она будет и в остатках регрессионной модели еt. Наиболее распространённый тест для обнаружения автокорреляции первого порядка: тест Дарбина- Уотсона. Он основан на d статистике: сравнивается среднеквадратичная разность соседних значений с дисперсией остатков.

Для процесса первого порядка:



Формула:

 ,

для больших выборок d=2-2p

Статистика Д-У применяется для проверки нулевой гипотезы о том, что в ряду не существует корреляции первого порядка (автокорреляции) между коэффициентами. Суть проверки: в сравнении расчетных значений d с критическими значениями из таблицы. Результат проверки:

|  |  |  |
| --- | --- | --- |
| P -> 0 | d->2 | Нет автокорреляции |
| P -> 1 | d->0 | Положительная автокорреляция |
| P -> -1 | d->4 | Отрицательная автокорреляция |

Если автокорреляция отсутствует, то , и значение d должно быть близким к 2. При наличии положительной автокорреляции d, будет меньше 2; при отрицательной автокорреляции d будет больше 2.

Критическое значение d при данном уровне значимости зависит от количества объясняющих переменных в уравнении регрессии и от количества наблюдений. К сожалению, оно зависит еще и от конкретных значений, принимаемых объясняющими переменными. Поэтому невозможно составить таблицу с точными критическими значениями для всех возможных выборок.

Тест Д-У ловит только определённую автокорреляцию t от t-1 . Поэтому существуют промежутки неопределённости, где мы не можем утверждать есть автокорреляция или нет. Для уменьшения промежутка неопределённости нудно увеличить число наблюдений.

Ограничения:

1. Тест не предназначен для обнаружения других видов автокорреляции (более чем первого).

2. В модели должен присутствовать свободный член.

3. Данные должны иметь одинаковую периодичность (не должно быть пропусков в наблюдениях).

Устранение автокорреляции.

Если р известно:, где t инновация, которая патологически тоже может содержать ошибку, но мы считаем. что она ошибку не содержит. Т.к. ошибка в данный момент времени зависит от ошибки в предыдущий момент времени, можно её исправить, сдвинув все ошибки на 1 момент времени назад новой переменной. Если р известно, то применение обобщённого метода наименьших квадратов позволяет получить несмещенные, эффективные оценки неизвестных коэффициентов регрессии.. Матрица выглядит следующим образом: главная диагональ =1, вторая =р, далее pn\*var(t ), и т.д. Проблема автокорреляции устранена. На практике значения коэффициента автокорреляции ρ обычно неизвестны и его оценивают другим способом.

Если р неизвестно: Нужно умножить уравнение t -1 на ρ и вычесть из t.,т.е. 

Полученная модель будет эквивалентна модели со случайным членом. Примечание: модель содержит значение регрессора и зависимой переменной в предыдущий момент времени.

1. **Временные ряды, характеристики временных рядов, декомпозиция**

Временно́й ряд— это упорядоченная (по времени) последовательность значений некоторой произвольной переменной величины. Каждое отдельное значение данной переменной называется отсчётом временного ряда. Тем самым, временной ряд существенным образом отличается от простой [выборки данных](http://ru.wikipedia.org/wiki/%D0%92%D1%8B%D0%B1%D0%BE%D1%80%D0%BA%D0%B0).

Ана́лиз временны́х рядо́в — совокупность математико-статистических методов анализа, предназначенных для выявления структуры временных рядов и для их прогноза.

Временные ряды состоят из двух элементов:

* 1. периода времени;
	2. числовых значений того или иного показателя, называемых уровнями ряда.

Временные ряды классифицируются по следующим признакам:

* по форме представления уровней
* по характеру временного параметра: моментные и интервальные временные ряды.
* по расстоянию между датами и интервалами времени выделяют полные (равноотстоящие) – когда даты регистрации или окончания периодов следуют друг за другом с равными интервалами и неполные (неравноотстоящие) – когда принцип равных интервалов не соблюдается.
* временные ряды бывают детерминированными и случайными: первые получают на основе значений некоторой неслучайной функции (ряд последовательных данных о количестве дней в месяцах); вторые есть результат реализации некоторой случайной величины.
* в зависимости от наличия основной тенденции выделяют стационарные ряды – в которых среднее значение и дисперсия постоянны и нестационарные – содержащие основную тенденцию развития.
* Типичным примером временного ряда можно назвать биржевой курс, при анализе которого пытаются определить основное направление развития (тенденцию или [тренда](http://ru.wikipedia.org/wiki/%D0%A2%D1%80%D0%B5%D0%BD%D0%B4)).
1. **Стохастические регрессоры. Двухшаговый метод наименьших квадратов. Тест Хаусмана**

Не всегда допустимо, что регрессоры не являются случайными величинами (и, соответственно, некоррелированы со случайным членом модели). Возможные причины:

1.При измерении значений регрессоров допускается возможность случайных ошибок (ошибок измерения) 2.В состав регрессоров входят лаги зависимой переменной, которые являются случайными величинами.

Рассмотрим сначала ситуацию, когда регрессоры являются стохастическими, но не взаимосвязаны со случайным членом модели:

Cov(Xj,ε) = 0, j=1,…, k

Пусть также матрица X имеет полный ранг (то есть ни одна из реализаций случайной матрицы не имеет линейно зависимых столбцов). В этом случае выполняются условия Гаусса-Маркова, а, следовательно, обычный метод наименьших квадратов позволяет получить несмещенные эффективные оценки неизвестных параметров модели. Если условие независимости регрессоров и случайного члена модели не выполняется, то оценки, полученные с помощью метода наименьших квадратов, будут:

1.Смещенными

2.Несостоятельными

Одним из возможных вариантов получения более хороших оценок параметров модели является использование инструментальных переменных

Двухшаговый метод наименьших квадратов (ДМНК) использует следующую центральную идею: на основе приведенной формы модели получают для сверхидентифицируемого уравнения теоретические значения эндогенных переменных, содержащихся в правой части уравнения. Затем они подставляются вместо фактических значений и применяют обычный МНК к структурной форме сверхидентифицируемого уравнения. В свою очередь, сверхидентифицируемая структурная модель может быть двух типов: либо все уравнения системы сверхидентифицируемы, либо же система содержит наряду со сверхидентифицируемыми и точно идентифицируемые уравнения. В первом случае, если все уравнения системы сверхидентифицируемые, для оценки структурных коэффициентов каждого уравнения используется ДМНК. Если в системе есть точно идентифицируемые уравнения, то структурные коэффициенты по ним находятся из системы приведенных уравнений.

На первом шаге с помощью обычного метода наименьших квадратов оценивают зависимость X от Z:

Прогнозные значения этой модели используются на втором шаге, для получения оценок неизвестных коэффициентов. Таким образом, необходимо выбрать между возможно несостоятельными, но эффективными МНК-оценками, и неэффективными, но состоятельными ИП-оценками.

Выбор между такими двумя оценками осуществляется на основе теста Хаусмана.

Данная статистика имеет распределение Хи-квадрат с m степенями свободы (m – количество инструментальных переменных) χ2(m) при выполнении нулевой гипотезы об отсутствии корреляции между регрессорами и случайным членом.

Соответственно, если наблюдаемое значение статистики не превысит критическое, то нулевая гипотеза не отклоняется и следует предпочесть обычные МНК-оценки, в противном случае – ИП- оценки.