Семинар №16

Тема

План

На данном занятии мы обсудим тест Jarque–Bera о нормальном законе случайного возмущения на примере нелинейной модели производства товаров и услуг России, которую обсудили и оценили на занятии №15.

Задача №1. Исследовать предполождение о нормальном законе распредердения случайного возмущения в линеарилизованной модели товаров и услуг в России.

Решение:

Шаг 1. Открываем файл с оценённой с линеаризованной и оценённой моделью и рассчитаем оценки случайных возмущений в уравнениях набдюдений по правилу:

$$u = \ln \frac{Y}{L}(\text{произв. труда}) - \left(a_0 + \alpha \cdot \ln \frac{K}{L}\,(\text{капитало вооружённость})\right)$$

u=LN(Y/L)-(a0+alpha*LN(K/L))
0.001180901
0.001571932
0.000576792
-0.002004605
-0.002583087
0.00032376
-0.001757011
0.000190067
0.005106452
-0.000854872
-0.000850503
-0.000899826

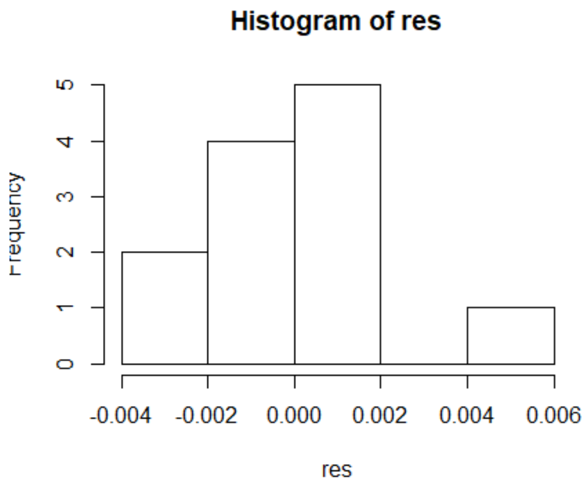
Шаг 2. В `Rstudio` выполняем следующий код:

```
library(ggplot2)
library(lmtest)
library(dplyr)
library(tseries)
file.show("normMain.txt")
pm<- read.table("normMain.txt", sep="", dec=".", header = TRUE)
pm

pmmodel<-lm(data=pm, LN.Y.L.~LN.K.L.)
summary(pmmodel)
res<-residuals(pmmodel)
jarque.bera.test(res)
```

```
  Jarque Bera Test
data:  res
X-squared = 3.4529, df = 2, p-value = 0.1779
```

Поскольку `p-value` больше чем 0.05, то гипотеза о нормальном распределении принимается.

Используя, функцию `hist` построим гистограмму



$\boxed{\text{ДЗ}}$ Исследовать на нормальность модель полученную в ДТ3.

Лекция №1
Эконометрика её задача и методы

$\boxed{\text{Д/з}}$ выполняется на отдельных листках.

План

1. Структура экономических задач;
2. Эконометрика, её задача и методы;
3. Первый принцип спецификации эконометрических моделей и эконометрическая теория;
4. Второй принцип спецификации моделей и алгебра
5. Третий принцип спецификации эконометрических моделей отражение фактора времени;
6. Приведённая форма модели, как инструмент анализа экономического объекта;

Структра эконометрических задач

1. Исходные даннные (значения известны):
$$x_1, x_2, \cdots, x_n \tag{1}$$
2. Искомые неизвестные:
$$y_1, y_2, \cdots, y_m \tag{2}$$
3. Взаимосвязи велечин (1) и (2);

В эконометрических задачах взаимосвязи существуют объективно и как правило инткитиано ощущаются. Объективный характер взаимосвязи позволяет приблизительно вычислить эндогенные переменные (2)

Задача Кейнса

- Исходные данные в задаче Кейнса считается I - объём инвестиций в экономику страны на заданном отрезке времени;
- Искомые неизвестные:
  1. Y – уровень дохода в стране, в том же периоде (ВВП);
  2. C – велечина совокупного потребления;
- Взаимосвязи велечин (I, Y, C) отражены в следующих утверждения экономической теории:

1. Доход Y образует потребительские (государственные и индивидуальные) расходы C и инвестиционные расходы I;
2. Уровень потребления C объясняется величиной дохода Y;
3. Каждая дополнительная еденица дохода, $\triangle Y = 1$ потребляется, как правило, не полностью: часть её идёт на инсветиции;

В любой математической задачае можно выделить три принципа спецификации. Взаимосвязи записанные математическим языком образуют математическую модель данной задачи.

Эконометрика её задачи и методы

- Эконометрика – прикладная математическая дисциплина, в которой изучаются конкретные количесвенные взаимосвязи объектов и процессов;
- Задача эконметрики заключается в объяснении(прогнозе или приближённом вычислении) искомых количественных характеристик (2) по известным значениям (1) каких-то других количественных характеристик этого объекта (задачи или процесса). Приближённые значения $(\widetilde{y_1}, \widetilde{y_2}, \cdots, \widetilde{y_m})$;

Эконметрика её задача и метод эконометрики

Метод решения задачи эконометрики состоит в предварительном построении упрощённой схемы изучаемого объекта (задачи или процесса), $F(\vec{y}, \vec{x}) = 0$, составленной математическим языком и именуемой эконометрической моделью, а затем в вычислении по этой модели приближённых значений неизвестных (2), $\vec{y} = f(\vec{x})$.

Прокомментируем $F(\vec{y}, \vec{x}) = 0$ и $\vec{y} = f(\vec{x})$. Символом F обозначены взаимосвязи (1) и (2). $\vec{y}$ – это весь набор величин (2). $\vec{x}$ – это весь набор исходных величин (1). Выражение

$$\vec{y} = f(\vec{x}) \tag{3}$$

, где каждая искомая велечина (2) выражена только через известные велечины (1).

- (1) – экзогенные переменные;
- (2) – эндогенные переменные;
- (3) – приведённая форма;

Задача эконометрики состоит в поиске приближённых значений на основании известных характеристик. Метод решения этой задачи заключается в предварительном построении модели и вычислении (2).

Первый принцип спецификации модели и экономическая теория

Приступаем к изучению принципов/приёмов, которыми пользуются экономисты в процессе построения.

Модель $F(\vec{y}, \vec{x}) = 0$ возникает в итоге трансляции на математический язык экономических утверждений о взаимосвязях исходных данных (1) и искомых неизвестных (2) объекта (процесса или задачи). Результат трансляции неоднозначен (возможны варианты!). Стараются привлекать линейные функции, так как они простые.

Пример.(Задача Кейнса)

$$\begin{cases} Y = C + I; \\ C = a_0 + a_1 Y; \\ 0 < a_1 < 1 \end{cases} \tag{4}$$

$\boxed{\text{Д/з}}$ 2. и 3. записать математическим языком.

На следующем рисунке показан график:

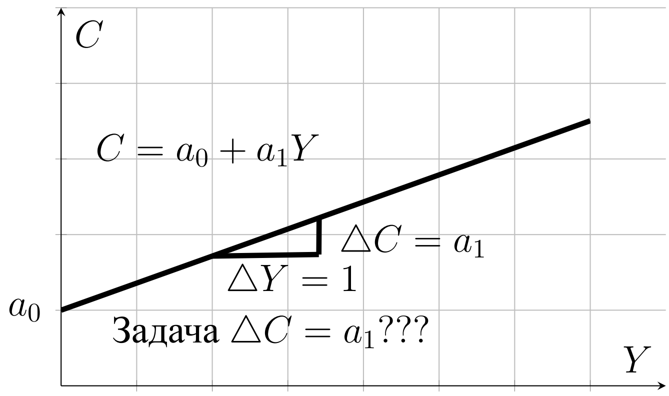


Рис. 1: Изменение потребления

Лекция №2
Завершение темы. Эконометрика, её задачи и методы

План

1. Второй принцип спецификации эконометрических моделей и приведённая форма простейшей макромодели Кейнса.
2. Отражения в спецификации эконометрической модели фактора времени: спецификация динамический моделей или 3-ий принцип спецификации эконометрических моделей, приведённая форма модели, как инструмент анализа изучаемого объекта.
3. Предельные величины в экономике.

На лекции 1 мы получили структурную форму макромодели Кейнса. Для расчётов в данной модели (по моделе Кейнса) необходимо трансформировать его к такому виду в котором каждая эндогенная переменная оказывается выраженной только через объясняющие переменные.

$$\begin{cases} Y = C + I; \\ C = a_0 + a_1 \cdot Y; \\ 0 < a_1 < 1; \end{cases} \tag{1}$$

Чтобы такая трансформация оказалась возможной необходим второй принцип: количество уравнений в структурной форме (1) совпадало с числом эндогенных переменных $\vec{y}$.

Проиллюстрируем трансформацию к модели Кейнса:

1. Правую часть первого уравнения структурной формы модели Кейнса, (1) подставим во второе уравнение и выразим из него искомую переменную C через экзогенную переменную, I. Получим приведенную форму C.
2. Приведенную форму C подставим в первое уравнение (1) и выразим из него Y. В итоге получим приведенную форму.

Приведённая форма модели Кейнса

$$\begin{cases} C = \dfrac{a_0}{1 - a_1} + \dfrac{a_1}{1 - a_1} \cdot I; \\ Y = \dfrac{a_0}{1 - a_1} + \dfrac{1}{1 - a_1} \cdot I; \end{cases} \tag{2}$$

Чтобы модель можно было трансформировать в приведённую форму число уравнений модели обязано должно совпадать с количеством объясняемых переменных.

Третий принцип спецификации модели: отражение фактора времени.

Фактор времени часто присутствует в условиях экономических задач. Для отражения в модели фактора времени переменные модели датируются. В итоге возникает динамическая модель.

Пример. Задача Линтнера о прогнозе уровня дивидендов: Ч. Ли, Дж. Финнерти Финансы корпораций "ИНФРА-М 2000, стр. 333)

- Исходные данные - чистая прибыль на акцию в текущем периоде, EPS_t
- Искомые неизвестные - уровень дивидендов на акцию в том же периоде, DPS_t

Взаимосвязи между величиной EPS и DPS сформулированы в следующих двух утверждениях (*Задача Линтнера о прогнозе уровня дивидендов*):

1. Фирма обладает долгосрочной целевой долей текущей прибыли (γ), которую она желает выплачивать в качестве дивидендов своим акционерам в текущем периоде.
2. Реальный уровень дивидендов в текущем периоде, DPS, определяется:
   (a) желаемым уровнем дивидендов в текущем периоде (DPS^w)
   (b) реальным уровнем дивидендов в предшествующем периоде (DPS_{t-1})

Запишем эти утверждения математическим языком, обращая внимание, что во втором утверждении мы обязаны сделать различия между дивидендами в текущем периоде, а это означает, что во втором утверждении содержится *фактор времени*.

$$\begin{cases} DPS_t^w = \gamma \cdot EPS_t; \\ DPS_t = \lambda DPS_t^w + (1-\lambda)DPS_{t-1} \\ 0 \leqslant \gamma \leqslant 1, \quad 0 \leqslant \lambda \leqslant 1 \end{cases} \tag{3}$$

t – время; λ – коэффициент корректировки.

Комментарий. *В первой строчке в левой части присутствует желаемый DPS^w это не наблюдаемая переменная, но она появилась в модели согласно этому утверждению и эта переменная рассматривается, как эндогенная. Во второй строчке в качестве линейной функции двух переменных DPS^w и DPS принято линейная однородная функция с положительными коэффициентами сумма которых равна 1. Добавим к сказанному, что уровень дивидендов это нечто среднее между желаемым уровнем дивидендов (DPS^w) и реальными дивидендами (DPS_t) в предшествующем периоде. Данная запись это структурная форма задачи Линтнера и в этой форме все переменные датированы (привязаны ко времени).*

Типы переменных в динамических моделях.

Объясняемые переменные в динамических моделях принято принято называть текущими эндогенными переменными; в модели Линтнера их две: DPS^w_t, DPS_t. Объясняющие могут включать в себя:

1. Лаговые эндогенные переменные, DPS_{t-1}
2. Лаговые экзогенные переменные, (смотри Семинар №2)
3. Текущие экзогенные переменные, EPS_t

Д/з Трансформировать модель Линтнера к приведённой форме. При отражении фактора времени возникает динамическая модель в которой объясняемые переменные. Приведённая форма, как инструмент анализа экономического объекта, задачи.

Вернёмся к модели Кейнса. И построим график функции C от объёма инвестиций I:

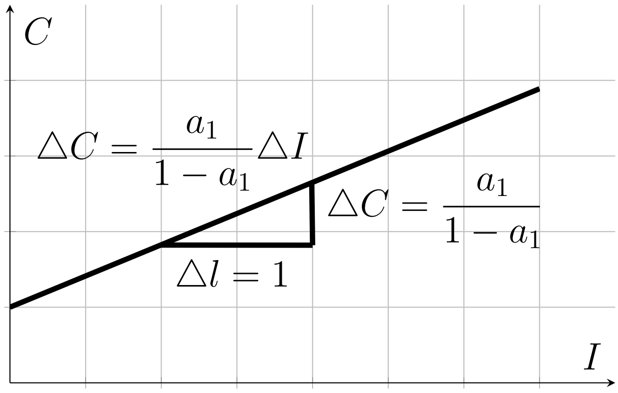


Рис. 1: Измение C от изменения инвестиций

Из первого уравнения приведённой формы можно найти взаимосвязь дополнительных инвестиций в экономику $\triangle C$. Получается, что коэффициент возникает в ответ на дополнительную инвестицию $\triangle I$. Построим график. Экономисты называют этот коэффициент придельным уровнем потребления по объёму инвестиций.

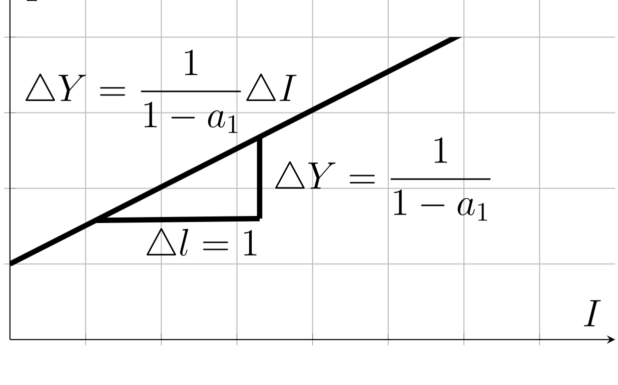


Рис. 2: Изменение дохода в ответ на изменение инвестиций

Взаимосвязь представлена уравнением:

$$\triangle Y = \frac{1}{1-a_1}\triangle I \tag{4}$$

$\boxed{\text{Д/з}}$. Проанализировать знак $\frac{1}{1-a_1}(+)$ вычислить значение при $a_1 = 0,6$, которые имеют название *мультипликатора инвестиций Кейнса* и дать экономическую трактовку этого коэффициента.

Обратим ещё раз внимание, что именно формулы $\triangle C = \frac{a_1}{1-a_1}\triangle I$ и $\triangle Y = \frac{1}{1-a_1}\triangle I$ действительно позволяют оценить отражения на уровне дохода и инвестиций в стране.

Вывод. Приведённая форма модели служит инструментом, как прогнозирования, так и анализа объекта. В линейных моделях коэффициенты в приведённой форме имеют смысл придельных величин в экономике.

Лекция №3
Отражение спецификации эконометрики в модели влияния неучтенных факторов. IV принцип спецификации

План

1. Функция потребления Кейнса и реальные данные;
2. Общий вид эконометрической модели с отражённым влиянием на эндогенные переменные неучтённых факторов;
3. Временной ряд и структура его уровней;

На прошлой лекции обсудили отражение в моделе фактора времени и использование модели, как инструмента анализа изучаемого объекта.

На сегодняшней лекции мы исследуем соответсвие математических моделей реальным данным и научимся отражать в моделе воздействие на искомые характреристик объекта (на текущие эндогенные переменные) неучтённых факторов. Наши исследования мы проведём на простейшей макромодели Кейнса:

$$\begin{cases} Y = C + I; \\ C = a_0 + a_1 \cdot Y; \quad 0 < a_1 < 1; \end{cases} \tag{1}$$

Согласуется ли эта функция с реальной статистикой?

Исследование проведём по следующей схеме:

На плоскости зададим декартову систему координат и по оси общисс отложим содержащиеся в табл.1 уровни ВВП РФ, на оси ординат отложим соотсветсвующие значения уравнений потребления; Если модель Кейнса соответсвует реальным данным, то точки графика расположатся на восходящей прямой.

Год	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
Y	6410	7288	8196	8915	10002	10767
C	4911	5554	6290	6739	7305	7773
I	1499	1734	1906	2175	2995	2994

Таблица 1: Статистические данные

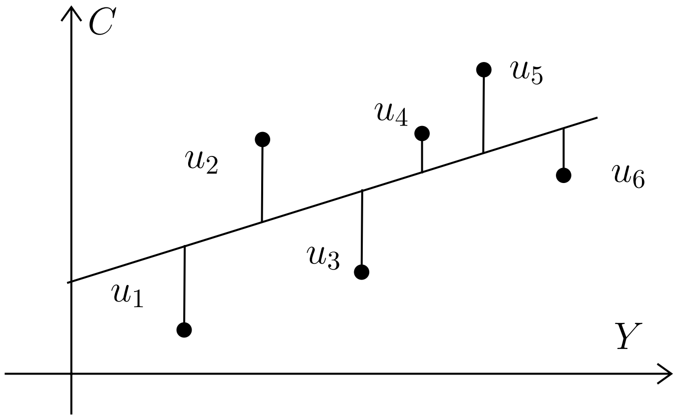


Рис. 1: Уравнение потребления

Рассмотрев построенный график, делаем следующие выводы:

1. Точки реальных данных (вот эти ромбики) не расположены на восходящей прямой, и это значит, что модель Кейнса в полной мере не соответствует реальным данным (не соответствует изучаемому объекту). Причина несоответствия – воздействие на совокупное потребление в стране неучтенных факторов. $\boxed{\text{ДЗ}}$ Сформулировать факторы, которые оказывают воздействие на совокуное потребление в стране и отсутсвует в модели Кейнса.

2. Точки реальных данных расположены вдоль ощущаемой восходящей прямой. Это значит, что модель Кейнса правильно отражает тенденцию, согласно которой изменяется совокупное потребление в стране в ответ на визменение дохода. Модель Кейнса не улавливает всех изменений совокупного потребления в стране, вызванныех неуточнёнными факторами, но правильно отражен главный фактор потребления - доход.

3. Точки реальных данных хаотично разбросаны вдоль восходящей прямой.

На основании п.1-3 можем предположить аналитическое описание диаграммы:

$$C \ = \ a_0 \ + \ a_1 \ \cdot Y \ + \ u(uncertain) \tag{2}$$

,где u - переменная велечина, которая принимает то положительное, то отрицательное значение рассеянное вокруг нуля. В силу хаотичности появления её значений экономисты называют случайным возмущением. Физики и в технических приложениях такие величины называются невязками или ошибками модели.

Основные характеристики случайного возмущения:

1. $E(u) = 0$ - среднее значение u, равное 0;

2. $E\left(u^2\right) \ = \sigma_u^2$, где σ_u - мера влияния неучтённых факторов; σ_u^2 средний квадрат разброса значений случайных возмущений вокруг мат. ожидания;

Отсюда следует спецификация эконометрической модели Кейнса в которой отражено влияние на C неуточнённых факторов:

$$\begin{cases} Y \ = \ C + I; \\ C \ = \ a_0 \ + \ a_1 \ \cdot Y \ + \ u; \\ 0 \ < \ a_1 \ < \ 1; \\ E(u) \ = \ 0, \ E\left(u^2\right) \ = \ \sigma_u^2; \end{cases} \tag{3}$$

Эконометричеными или регрессионными моделями называются дескриптивные ЭММ со случайными возмущениями в поведенческих уравнениях.

Приведём спецификацию эконометрической модели интерна

$$\begin{cases} DPS_t^e \ = \gamma \ \cdot EPS_t; \\ 0 \ \le \gamma \ \le 1; \ 0 \ \le \lambda \ \le 1; \\ DPS_t \ = \ \lambda \ \cdot \ DPS_t^e \ + \ (1-\lambda) \ \cdot DPS_{t-1} \ + \ v_t; \\ E(v_t) \ = \ 0, \ E\left(v_t^2\right) \ = \sigma_v^2; \end{cases} \tag{4}$$

Общий вид эконометрической модели в структурной форме:

$$F\left(\overrightarrow{y_t}, \overrightarrow{x_t}\right) = \overrightarrow{u_t} \tag{5}$$

Структурная форма эконометрической модели из линейной алгебры уравнений:

$$A \cdot \ \overrightarrow{y_t} \ + \ B \cdot \overrightarrow{x_t} \ = \ \overrightarrow{u_t} \tag{6}$$

$\overrightarrow{u_t}$ - вектор случайных возмущений, некоторые компоненты могут равняться 0.

В ситуации Линтерна $\overrightarrow{u_t}$ состоит из двух компонент: $\overrightarrow{u_t} \ = \ (0, \ v_t)$.

Задача: Найти $\overrightarrow{y_t}$ в моделе Кейнса. $\overrightarrow{x_t}$, $\overrightarrow{u_t}$?

Для отражения в деструктивной модели влияния на объясняемые переменные неучтённых факторов в правых частях поведенческих уранений включаются случайные возмущения; случайные возмущения - та часть эндогенной переменной, которая порождена неуточнёнными факторами.

Приведенная форма эконометрической модели:

$$\vec{y}_t \ = f\left(\vec{x}_t, \overrightarrow{u_t}\right) \tag{7}$$

Приведённая форма линейной эконометрической модели:

$$\overrightarrow{y_t} \ = M \ \cdot \ \overrightarrow{x_t} \ + \ \overrightarrow{\varepsilon_t} \tag{8}$$

$$E\left(\triangle \ \overrightarrow{y_t}\right) \ = \ M \cdot \triangle \ \overrightarrow{x_t} \tag{9}$$

Задача: Трансфорсировать (3) к приведённой форме.

Лекция №5.
Линейная модель множественной регрессии (базовая модель эконометрики)
План

1. Завершение обсуждения схемы построения эконометрических моделей;
2. **Линейная модель множественной и парной регрессии (базовая модель эконометрики);**
3. Уравнение наблюдения объекта (схема Гаусса-Маркова), компактная запись схемы Гаусса-Маркова и понятие статичсктической процедуры оценивания параметров модели.

На прошлой лекции присупили к обсуждению схемы построения эконометрических моделей:

1. Спецификая модели;
2. Сбор и проверка статистической информации;
3. Оценивание (найстройка) модели;
4. Проверка адекватности (верификации) модели

Мы разобрали **1 пункт.** Там же подчеркнули, что спецификация эконометрической модели непременно включает в себя неизвестные константы $a_0,\ a_1$, которые называют параметрами модели. В самом общем виде спецификация модели записывается так: $F\left(\vec{x}_t,\ \vec{y}_t;\ \vec{p}\right) \ = \ \vec{u}_t$. Здесь мы начинаем выявлять константы.

$$\begin{cases} Y \ = \ C \ + \ I \\ \ = \ a_0 \ + a_1 \ \cdot Y \ + \ u \\ 0 \ < \ a_1 \ < \ 1 \\ E(u) \ = \ 0, \ E\left(u^2\right) \ = \ \sigma_u^2 \end{cases}$$

Добавим, что спецификация временного ряда квартальных уровней ВВП России включает в себя 8 констант:

$$(a_0,\ a_1,\ a_2,\ a_3,\ b_1,\ b_2,\ b_3,\ \sigma_u)$$

Второй этап состоит в сборе и проверке статистической информации в виде конкретных реальных значений переменных входящих в модель. Примером второго этапа служат данные из таблицы № 1 из лекции № 1.

Замечание. Собранная статистическая информация разделяется на две части причём большая часть именуется обучаещей выборкой и предназначена для определения параметров модели; остальная информация именуется тестовой или контролирующей выборкой и используется для проверки адекватности модели.

На третьем этапе схемы методами математической статистики оцениваются параметры модели по обучающей выборке. 4 и 5 практическое занятие служит иллюстрация 3 этапа. Подчеркнём, что всегда удаётся вычислить только приближённые значения параметров $\widetilde{\vec{p}}$ (оценки). Причина приближённого значения оценок параметров состоит в наличии случайных возмущений, порождённых неучтёнными факторами.

Четвёртый этап состоит из проверки адекватности оценённой модели $F\left(\vec{x}_t,\ \vec{y}_t;\ \widetilde{\vec{p}}\right) = \vec{u}_t$ путём сопоставления прогнозов значений эндогенных переменных из контролирующей выборки $\tilde{y}_t = f\left(\vec{x}_t,\ \widetilde{\vec{p}}\right)$ (3.1.11) с реальными значениями. Модели признаётся адекватной, если ошибки прогнозов не превышают критические уровни $|\tilde{y}_t - y_t|$ (ошибка прогноза) $\leq e_{\text{крит}}\left(\tilde{y}_t\right)$ (3.1.12) (15% или $2 - 3\ \sigma$).

Вывод: схема построения эконометрических моделей состоит из 4 этапов и если модель признаётся не адекватной, то экономист возращается на первый этап и выявляет ошибки спецификации модели.

Линейная модель множественной регрессии (базовая модель эконометрики)

Модель со следующей спецификацией

$$\begin{cases} y = a_0 + a_1\,x_1 + a_2\,x_2 + \ldots + a_k\,x_k + u \\ E(u) = 0;\ E\left(u^2\right) = \sigma_u^2; \end{cases} \tag{3.2.1}$$

является базовой моделью эконометрики и называется *линеной моделью множественной регресии*. Символом y обозначена единственная объясняемая переменная; символом $(x_1,\ x_2, \ldots,\ x_k)$(3.2.3) обозначены предопределённые (объясняющие переменные); символами $(a_0,\ a_1, \ldots,\ a_k)$(3.2.4) обозначены константы и носят название *коэффициенты модели* (более полно, коэффициентов *функции регрессии*).

$$\tilde{y} = a_0 + a_1\,x_1 + a_2\,x_2 + \ldots + a_k\,x_k$$

Символом $\tilde{y}$ обозначена функция объясняющий перменных, имеющая смысл той части эндогенной переменной y, которая объясняется предопределёнными перменными модели; величина $\tilde{y}$ носит название *функции регрессии* и с точки зрения теории вероятностей является условным математическим ожидание велечины y; u обозначено случайное возмущение.

Смысл коэффициента a_g при переменной x_g

a_j имеет смысл ожидаемого предельного значения переменной y по переменной x_j.

$$E(\triangle\ y)\ =\ a_j \cdot \triangle\ x_j \tag{3.2.4}$$

Пример (линейной модели множественной модели). Вернёмся к нашей предшествующей лекции и рассмотрим модель квартальных уровней ВВП России с кубическим трендом. Объясняемые переменные - это квартальные уровни датированные кварталами или t, связанный с календарём следующим правилом $t\ =\ 1,$ для первого квартала 2011 года.

$$\begin{cases} Y_t\ =\ a_0\ +\ a_1\ \cdot t\ +\ b_1\ \cdot\ d_1(t)\ +\ b_2\ \cdot d_2(t)\ +\ \ b_3\ \cdot d_3(t)\ + u_t \\ E(u)\ =\ 0;\ \ E\left(u^2\right)\ =\ \sigma_u^2; \\ t\ = 1,\ 2,\ 3, \ldots \\ t\ =\ 1\ \Rightarrow 1 \text{ квартал 2011 года} \end{cases} \tag{3.2.5}$$

Объясняющими переменными служит 6 переменных: $(t,\ t^2,\ t^3,\ d_1(t),\ d_2(t),\ d_3(t))$.
Замечание. Обратим внимание, что в линейной модели множесвтенной регрессии среди объясящих переменных $d_1(t),\ d_2(t),\ d_3(t)$, только одная является независимой - переменная t. Это значит, что в ЛММР объясняющие переменные могут быть как независимыми друг от друга, так и являться известными функциями каких-то других переменных велечин. Запомним, что аргумент - независим, а экзогенная переменная может быть зависимой.

Линейная модель парной регрессии

Простейшим случаем ЛММР служит модель парной регрессии с одной объясняющей переменной.

$$\begin{cases} y\ =\ a_0\ +\ a_1\ \cdot x\ +\ u \\ E(u)\ =\ 0;\ \ E\left(u^2\right)\ =\ \sigma_u^2; \end{cases} \tag{3.2.6}$$

Приведём важный для инвестиционного анализа пример инвестиционной модели. Является рыночная модель ценной бумаги.

$$\begin{cases} r\ =\ \alpha\ +\ \beta\ \cdot r_I\ +\ u \\ E(u)\ =\ 0;\ \ E\left(u^2\right)\ =\ \sigma_u^2; \end{cases} \tag{3.2.7}$$

$\boxed{\text{ДЗ}}$ Рыночная модель ценной бумаги У. Шарп, Г. Александер, Д. Бэйли. Объясняющие - это доходность на акцию за 1 месяц r. Доходность на рыночный актив r_I.

Добавим в следующей таблице приведены значения переменных r b r_I рыночной модели компании Лукойл.

Вывод: линейная модель имеет спецификацию (3.2.1), которая имеет следующие параметры $(a_0,\ a_1,\ \ldots,\ a_k,\ \sigma_u)$. Приступаем к статистической процедуре оценивания этих параметров.

Разместим обучающую выборку при построении ЛММР в следующей таблице: Подставляем каждую строчку в уравнение модели (2.1)

№ наблюдений	y	x_1	x_2	$\ldots$	x_k
1	y_1	$x_{1,1}$	$x_{2,1}$	$\ldots$	$x_{k,1}$
2	y_2	$x_{1,2}$	$x_{2,2}$	$\ldots$	$x_{k,2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$
n	y_n	$x_{1,n}$	$x_{2,n}$	$\ldots$	$x_{k,n}$

В итоге получим следующую систему уравнений наблюдений в рамках (2.1):

$$\begin{cases} y_1 = a_0 + a_1\, x_{1,1} + a_2\, x_{2,1} + \ldots + a_k\, x_{k,1} + u \\ y_2 = a_0 + a_1\, x_{1,2} + a_2\, x_{2,2} + \ldots + a_k\, x_{k,2} + u \\ \ldots \\ y_n = a_0 + a_1\, x_{1,n} + a_2\, x_{2,n} + \ldots + a_k\, x_{k,n} + u \end{cases}$$

Её принято называть схемой Гаусса-Маркова.

Компактная запись:

$$\vec{y} = X \cdot \vec{a} + \vec{u} \tag{3.3.4}$$

$\vec{y}$ это схема левых частей. X – это матрица значений объяняющих перменных, расширенная столбцом 1 (если есть свободный член a_0), $\vec{a}$ –вектор коэффициентов модели $k+1$. $\vec{u}$ – веутор случайных возмущейний.

Лекция №6

Матрица X непременно должная быть вертикальной строк должно быть больше чем стобцов.

Символом $\vec{a} \;=\; (a_0,\ a_1,\ a_2,\ a_3)$ —в компактной записи обозначен вектор коэффициентов модели. $\vec{u}$ – вектор значения случайного возмущения, присутствующего в модели.

Уравнения наблюдений необходимы для оценивания параметров модели и не редко под моделью понимают уравнение наблюдений:

$$\vec{y} = \begin{pmatrix} C_{2003} \\ \dots \\ C_{2017} \end{pmatrix};\ \vec{a} \;=\; \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix}; X \;=\; \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,m} \\ \dots & \dots & \dots & \dots \\ x_{n,1} & x_{n,2} & \dots & x_{n,m} \end{pmatrix}; \vec{u} \;=\; \begin{pmatrix} u_{2003} \\ \dots \\ u_{2017} \end{pmatrix} \quad (1)$$

Подчёрнем, что экономист обладает информацией ввиде двух массивов $(\vec{y},\ X)$ связанные между собой уравнениями наблюдений. Упомянутые массивы мы будем называть *выборкой.*

Понятие статистической процедуры

Рассмторим лаканичную запись эконометрической модели:

$$F\left(y_t,\ \vec{x}_t;\ \vec{p}\right) \;= u_t$$

Пусть известна обучающая выборка. Статистической процедурой оценивания параметров модели принято называть некоторую функцию $\phi\left(\vec{y}, X\right)$ выборки, значение этой функции являются оценки параметров модели:

$$\widetilde{\vec{p}} \;=\; \begin{pmatrix} \widetilde{\vec{a}} \\ \widetilde{\vec{\sigma}}_n^2 \end{pmatrix} \;=\; \phi\left(\vec{y}, X\right) \qquad (2)$$

Процедура $\phi\left(\vec{y}, X\right)$ называется оптимальной в заданном классе функций, если доставляемые ею оценки параметров обладают следующими свойствами:

$$\begin{cases} E\left(\widetilde{\vec{p}}\right) \;=\; \vec{p}; \\ Var\left(\tilde{p}_j\right) \to \min; \end{cases} \qquad (3)$$

Математическое ожидание оценок парметров совпадает с истинными значениями. В математической статистике оценки с такими свойствами называют *несмещёнными.*

Второе свойство означает, что средний квадратический разброс оценок параметров относительно истинных значений параметров является минимально возможным. Символом Var мы обозначаем дисперсию оценок параметров.

Итог: статистическая процедура оценивания модели - это некоторая функция выборки, значениям этой функции служит оценки параметров. Процедура оптимальна в заданном классе функций, если её значения удовлетворяют требованиям (3).

Случайный вектор и его основые количественные характеристики

Случайный вектор - это упорядоченный набор случайных перменных принято называть случайным вектором:

$$\vec{x}^T \;=\; (x_1,\; x_2\; \ldots,\; x_n) \tag{4}$$

Для практики важны следующие две случайные характеристики:

1. Математической ожидание случайного вектора

$$E\left(\vec{x}^T\right) \;=\; \vec{m}_{\vec{x}}^T \;=\; (m_1, m_2,\; \ldots,\; m_n) \tag{5}$$

- это веткор из математический ожиданий случайных компонент. Математическое ожидание - это среднее значение. Математическое ожидание - это *константа*.

2. Коварицонная матрица:

$$Cov\left(\vec{u}, \vec{u}\right) \;=\; \mathbf{\Omega}_{\vec{x}} \;=\; \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \ldots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \ldots & \sigma_{2,n} \\ & & \ddots & \\ \ldots & \ldots & & \ldots \\ \sigma_{n1} & \sigma_{n2} & \ldots & \sigma_n^2 \end{pmatrix} \tag{6}$$

так принято называть квадратную симметричную матрицу, на главной диагонали которой располагаются дисперсии компонент случайного вектора, а недиагональные элементы - это ковариации компонент. Ковариация, например σ_{1n}, это *константа* характризующая взаимосвязь компоненты x_1 и x_n. Если x_1 и x_n независимые, то $\sigma_{1n} \;=\; 0$.

Основые количественные характеристики афинного преобразования случайного вектора

Афинное преобразование - это линейное неоднородное преобразование. Пусть символом $\vec{x}$ обозначен случайный вектор. Аффинным преобразованием этого вектора принято называть вектор $\vec{y}$, который вычисляется по следующему правилу:

$$\vec{y} \;=\; A\; \cdot \vec{x} \;+\; \vec{b}; \tag{7}$$

Здесь символом A обозначена матрица коэффициентов (констант), символом $\vec{b}$ обозначен вектор констант.

Отметим правила расчёта основных характеристик аффинного преобразования:

$$E\,(\vec{y})\ =\ A\ \cdot E\,(\vec{x})\ +\ \vec{b}; \tag{8}$$

$$Cov\,(\vec{y},\ \vec{y})\ =\ A\ \cdot Cov\,(\vec{x},\ \vec{x})\ +\ \vec{b}; \tag{9}$$

Веса компонент случайного вектора и факторизация его коварицонной матрицы

Пусть $\vec{x}^T\ =\ (x_1,\ x_2\ \ldots,\ x_n)$ случайный вектор, пусть x_i какая-то компонента вектора; вес компонент x_i – это константа, которая вычисляется по следующему правилу:

$$p_i\ = \frac{\sigma_0^2}{\sigma_i^2} \tag{10}$$

где σ_0^2 обозначена произвольная, но фиксированная положительная коестанта. Понятие веса позваляет представить коварицонную матрицу $\vec{x}$ в следующем виде:

$$Cov\,(\vec{x},\ \vec{x})\ =\ \sigma_0^2\cdot\begin{pmatrix}\sigma_1^2/\sigma_0^2 & \sigma_{12}^2/\sigma_0^2 & \ldots & \sigma_{1n}^2/\sigma_0^2\\ \sigma_{21}^2/\sigma_0^2 & \sigma_2^2/\sigma_0^2 & \ldots & \sigma_{2n}^2/\sigma_0^2\\ \ldots & \ldots & \ldots & \ldots\\ \sigma_{1n}^2/\sigma_0^2 & \sigma_{2n}^2/\sigma_0^2 & \ldots & \sigma_n^2/\sigma_0^2\end{pmatrix} = \sigma_0^2\cdot Q \tag{11}$$

Матрицу Q принято называть матрицей *весовых коэффициентов*. По главной диагонали этой матрицы размещаются обратные веса компонент.

Приступаем у обсуждению оптимальной процедуры оцениванию параметров линейной модели множественной регрессии.

$$\tilde{y}\ =\ a_0\ +\ a_1\,x_1\ +\ a_2\,x_2\ +\ \ldots\ +\ a_k\,x_k$$

Теорема Гаусса-Маркова. Пусть в уравнениях наблюдений,

$$\vec{y}\ =\ X\ \cdot\vec{a}\ +\ \vec{u}$$

0. Столбцы X линейно независимые,

1. $E(u_1)\ =\ E(u_2)\ =\ \ldots\ = E(u_n)\ =\ 0$, заложенно в спецификации;

2. $Var(u_1)\ =\ Var(u_2)\ =\ \ldots\ =\ Var(u_n)\ =\ \sigma_u^2$,заложенно в спецификации;

3. $Cov(u_i,\ u_j)\ =\ 0;$ при $i\ \neq j$; (в частности незваисимы друг от друга)

4. $Cov(u_i,\ x_{li})\ =\ 0$. Случайные возмущения некоррелированны с компонентами матрицы x.

Если все утверждения верны, тогда справедливы следующие утверждения:

$A) \ \widetilde{\vec{a}} = \left(X^T \cdot X\right)^{-1} \cdot X^T \cdot \vec{y} = Q \cdot X^T \cdot \vec{y};$

$B) \ \tilde{\sigma}_u^2 = \dfrac{\sum\limits_{i=1}^{n} \tilde{u}_i^2}{n - (k+1)}$, наилучшая оценка с минимальной дисперсией, где символом $\tilde{u}_i$ обозначена оценка случайного возмущения u_i. В знаменателе стоит число равная разности объёма обучающей выборки и количесвтва определяемых коэффициентов модели; это число называется количесвом степеней свободы. Оценки коэффициентов обладают замечательным свойством (C), которая служит общепринятое название A метод наименьших квадратов.

$C) \ \sum\limits_{i=1}^{n} \tilde{u}_i^2 \to \min.$

$Cov\left(\widetilde{\vec{a}}, \widetilde{\vec{a}}\right) = \tilde{\sigma}_u^2 \cdot \left(X^T \cdot X\right)^{-1} \ (= Q).$

Вывод: Метод наименьших квадратов при определённых условиях является наилучшей процедурой оценивания линейных эконометрических моделей.

$D) \ \begin{cases} S\tilde{a} = \tilde{\sigma}_u^2 \cdot \sqrt{q_{j+1\ /\ j+1}} \\ j = 0, 1, \ldots, k \end{cases}$

Лекция №7

Доказательство теоремы Гаусса-Маркова
План

1. Необходимые сведенья из теории вероятности;
2. Доказательство теоремы Гаусса-Маркова;
3. Консольтуция;

В первом пункте мы вспомним те сведенья из теории вероятностей, которые необходимы в эконометрике и конкретно в доказательстве теоремы Гаусса-Маркова. Начнём с понятия случайной переменной. **Случайной переменной** u – называется переменная велечина, возможные значения которой $(q_1,\ q_2,\ \ldots,\ q_n)$ появляются в результате некоторого эксперимента (опыта) с вероятностями этих значений $(p_1,\ p_2,\ \ldots,\ p_n)$; Вот полная запись определения случайной переменной, которая называется *законом распределения*:

$$u = \begin{Bmatrix} q_1,\ q_2,\ \ldots,\ q_n \\ p_1,\ p_2,\ \ldots,\ p_n \end{Bmatrix}.$$

Поясним примером понятие случайной переменной:

Опыт состоит в бросании монеты. Если монета выпала гербом, то мы будем считать, что наша переменная u приняла значение -14, а если выпадает решка, то мы предполагаем, что наша переменная приняла значение $+14$, мы можем записать в следующем виде:

$$u = \begin{cases} -n, & \text{если герб} \\ n, & \text{если решка} \end{cases}$$

Основные характеристики случайной переменной:

Первая характеристика - это *математическое ожидание*, так называют константу которая вычисляется по правилу:

$$m\ = E(u)\ =\ p_1 \cdot q_1\ + p_2\ \cdot q_2\ +\ \ldots p_n\ \cdot q_n;$$

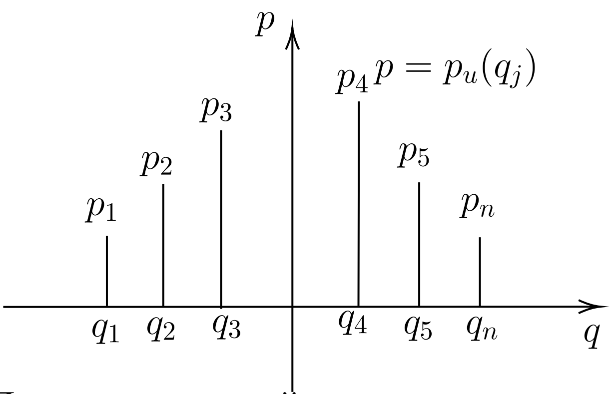
Вторая характеристика называется дисперсией Var и рассчитывается по правилу:

$$Var(u)\ =\ \ p_1 \cdot (q_1 - m)^2\ + p_2\ \cdot (q_2 - m)^2\ +\ \ldots p_n\ \cdot (q_n - m)^2$$

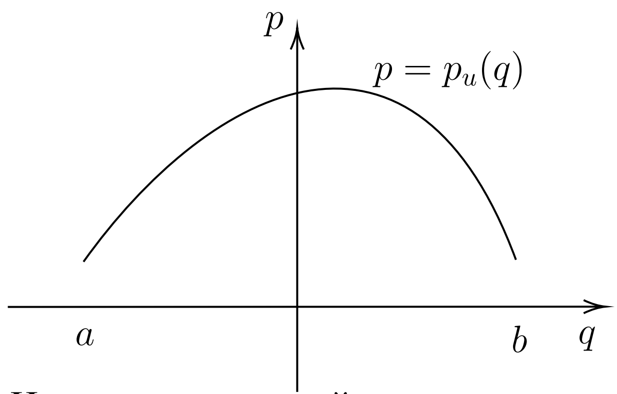
Дисперсия - это константа равная среднему квадрату разброса возможных значений случайной переменной относительно математического ожидания. Положительно квадратный корень из дисперсии называется средним квадратическим отклонением.

$\boxed{\text{ДЗ}}$ Доказать, что в приведённом выше примере $\sigma = 14$. Доказать, что самый точный прогноз случайной переменной - это её математическон ожидание, то есть доказать:

$$\min_c E(u - c)^2 = E(u - E(u))^2 = \sigma_u^2$$



Дискретная случайная переменная



Непрерывная случайная переменная

Закон распределения случайной переменной называют *диффиренциальным законом или вероятностной функцией, а в ситуации непрерывной случайно велечины - плотностью вероятности.*

Замечание. Случайная велечина u называется непрерывной, если множество её возможных значений есть некоторый интервал числовой прямой, а вероятность появления в опыте каждого конкретного значение равна 0.

Законы распределения используемые в эконометрике.

Первый закон в эконометрике (самый важный) - это нормальный закон (Муавра-Гаусса). Имеет уравнение (2.9)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

Второй закон называется законом распределения Стьюдента или $t-$распределение. Символом m обозначено кол-во степеней свободы (при $m > 30$ почти сопадает с нормальным).

$$f_t(y) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{y^2}{n}\right)^{-\frac{n+1}{2}}$$

Третий закон распределения Фишера. Это закон зависит от двух констант (m, n), которые называются степенями свободы.

$$P_F(x) = \frac{\sqrt{\frac{(d_1\ x)^{d_1}\ d_2^{d_2}}{(d_1\ x + d_2)^{d_1+d_2}}}}{x\,\mathrm{B}\left(\frac{d_1}{2}, \frac{d_2}{2}\right)}$$

Четвёртый закон распределения Хи-квадрат. В этом законе присутствует константа m, которая называется колическтвом степеней свободы.

$$P\chi^2(x) = \frac{(1/2)^{k/2}}{\Gamma(k/2)} x^{k/2-1} e^{-x/2}$$

Отметим функции Excel, которые обозначают эти законы:

1. НОРМРАСП();
2. СТЬЮДРАСП();
3. FРАСП();
4. ХИ2РАСП();

Характеристики вероятности взаимосвязи двух случайных переменных

Пусть x и y пара случайных переменных (пример: опыт состоит в бросании игральной кости $x-$это очнки которые выпадают на нижних гранях, а $y-$на верхней). Характеристика взаимосвязи расчитывается по формуле и называется ковариацией:

$$Cov(x, y) = \sigma_{x,\ y} = E(x \cdot y)\ - E(x) \cdot E(y)$$

Если ковариация положительная, то с ростом x возрастает y и наоборот. Если x и y независимые, то ковариация равна 0.

$\boxed{\text{ДЗ}}$ Можно показать, что ковариация очков на нижней и верхней грани равна $-\dfrac{35}{12}$. Нормированная ковариация вычисляется по формуле и носит название коэффициента корреляции:

$$Cor(x,\ y)\ = \rho_{x,y}\ = \frac{\sigma_{x,\ y}}{\sigma_x \cdot \sigma_y}$$

Вернёмся к основным характеристикам случайного вектора рассмотреным на лекции (6). Пусть случайным вектором является вектор случайных возмущений в уравнениях наблюдения объекта в схеме Гаусса-Матрока (смотри семинар (6)). Рассмотрим фактаризацию его коварицционной матрицы:

$$Cov\left(\vec{x},\ \vec{x}\right)\ =\ \sigma_0^2 \cdot \begin{pmatrix} \sigma_1^2/\sigma_0^2 & \sigma_{12}^2/\sigma_0^2 & \dots & \sigma_{1n}^2/\sigma_0^2 \\ \sigma_{21}^2/\sigma_0^2 & \sigma_2^2/\sigma_0^2 & \dots & \sigma_{2n}^2/\sigma_0^2 \\ \dots & \dots & \dots & \dots \\ \sigma_{1n}^2/\sigma_0^2 & \sigma_{2n}^2/\sigma_0^2 & \dots & \sigma_n^2/\sigma_0^2 \end{pmatrix}\ = \sigma_0^2 \cdot Q$$

Матрицу Q весовых коэффициентов мы обозначим P^{-1}.

Вывод: У случайной переменной есть две основные характеристки (две константы), взаимосвязь случайных переменных описывается их ковариацией и ковариации компонент случайного вектора заполняют его ковариационную матрицу.

Доказательство теоремы Гаусса-Маркова

Приступаем к доказательству утверждений теоремы Гаусса-Маркова. Мы расширим предпосылки с номерами 2 и 3 этой теоремы отказавшись от них и предполагая, что вектор случайных возмущений $\vec{u}$ имеет нулевое математическое ожидание и коварициoнную матрицу $Cov\ (\vec{u}, \vec{u}) = \sigma_0^2 \cdot P^{-1}$.

Докажем утверждение А:

А. 1) Мы будем разыскивать оценку вектора $\vec{a}$ в классе всех линейных функций определённых на векторе значений эндогенной переменной $\vec{y}$, так что определению подлежит матрица M этого линейного преобразования мы собираемся разыскать M.

$$\widetilde{\vec{a}} = M \cdot \vec{y}$$

2) Поиск матрицы M удобно осуществить создавая оптимальную статистическую процедуру оцениевания значения y_0 произвольной линейной функцией вектора коэффициентов модели

$$y_0 = \vec{x}_0^T \cdot \vec{a}$$

3) Процедуру оценивания числа y_0 мы будем отыскивать в классе линейных функций $\vec{y}$, где $\vec{m}-$это строка линейных коэффициентов.

$$\widetilde{y_0}\ = \overrightarrow{m_0}\ \cdot \vec{y}$$

И будем отыскивать опираясь на два требования оптимальности:

$$\begin{cases} E\left(\tilde{y}_0\right) = y_0 = \vec{x}_0^T \cdot \vec{a} \\ Var\left(\tilde{y}_0\right)\ \to \min \end{cases}$$

Вычислим мат ожидание символа $\tilde{y}_0$. Первое требование оптимальности приводит к следующим уравнению отностительно искомых коэффициентов $\vec{m}$:

$$E\left(\tilde{y}_0\right) = \vec{m} \cdot X \cdot \vec{a},\ \Rightarrow \vec{m}^T\ \cdot X\ =\ \overrightarrow{a'}\ = \vec{x}_0^T \cdot \vec{a}$$

Теперь найдём дисперсию опираясь на следующее выражение: $Cov\ (\vec{u}, \vec{u}) = \sigma_0^2 \cdot P^{-1}$.

Значит дисперсия: $Var\left(\tilde{y}_0\right) = \sigma_0^2\ \cdot \vec{m}^T \cdot P^{-1}\ \cdot \vec{m}$

Следовательно искомые оценки коэффициентов нужно найти в процессе решения следующей оптимизационной задачи:

$$\begin{cases} \vec{m}^T \cdot P^{-1}\ \cdot \vec{m}\ \to \min \\ X^T \cdot \vec{m}\ =\ \vec{x}_0 \end{cases}$$

Лекция №8
Обобщение метода наименьших квадратов
План

1. Завершение доказательства оптимальности оценок коэффициентов модели методом наименьших квадратов и его обобщение (обощённым методом наименьших квадратов)
2. Доказательство утверждения D уравнения Гаусса-Маркова и смысл символа (сигма) σ_0^2в выражении ковариационной матрицы вектора оценок коэффициентов.
3. Доказательства утверждения С теоремы Гаусса-Маркова, метод наименьших квадратов, как частный случай обобщённого метода наименьших квадратов. Взаимосвязь метода наименьших квадратов хи метода максимального правдоподобия при нормально расспределении случайного возмущения.

Приступаем к первому вопросу. На прошлой лекции мы приступили к доказательству утверждения А в предположении, что $Cov(\vec{u},\vec{u})$имеет общую структуру $\sigma_0^2 \cdot P^{-1}$.
Значит дисперсия: $Var\left(\widetilde{y}_0\right) = \sigma_0^2 \cdot \vec{m}^T \cdot P^{-1} \cdot \vec{m}$ (отказ от предпосылки 2) и недиагональные элементы (ковариации компонент) могут быть не нулевыми. При справедливых предпосылках 2 и 3 ковариационная матрица имеет структуру $Cov(\vec{u},\vec{u}) = \sigma_u^2 \cdot I$.
Наше доказательство А мы осуществим в процессе поиска <u>оптимальной линейной процедуры</u> оценивания значения y_0 производьной линейной функции веркторa истинных значений вектора $\vec{a}$ коэффициентов. $\left(y_0 = \vec{x}_0^T \cdot \vec{a}\right)$ Линейная процедура оценивания имеет вид: $\widetilde{y_0} = \vec{m}^T \cdot \vec{y}$, где строка $\vec{m}^T$ состоятельности. В первой строчке выражения

$$\begin{cases} E\left(\widetilde{y}_0\right) = y_0 = \vec{x}_0^T \cdot \vec{a} \\ Var\left(\widetilde{y}_0\right) \to \min \end{cases} \tag{5.11}$$

определим левую часть поэтому строка $\vec{m}^T$ обязана удовлетворять:
$E\left(\widetilde{y}_0\right) = \vec{m} \cdot X \cdot \vec{a}, \ \Rightarrow \vec{m}^T \cdot X = \vec{a} = \vec{x}_0^T \cdot \vec{a}$

Так как $\vec{a}$ произвольный вектор, то строка может быть найдена, как система уравнений 5.12

$$\vec{m}^T \cdot X = \vec{x}_0^T, \ \Leftrightarrow X^T \cdot \vec{m} = \vec{x}_0 \tag{5.12}$$

$\boxed{\text{ДЗ}}$ Сколько уравнений в системе 5.12 и является ли эта система недоопределённой?

Промежуточный итог: искомый вектор коэффициентов m согласно первому треованию оптимальности обязан быть решением 5.12.
Перейдём в левую часть второй строки 5.11 дисперсию находим по теореме

Фишера.

$$Var\left(\widetilde{y}_0\right) = \sigma_0^2 \cdot \vec{m}^T \cdot P^{-1} \cdot \vec{m} \tag{5.13}$$

Соеденим требования 5.13 и 5.12 в требование оптимальности 5.11. Следовательно, искомые коэффициенты $\vec{m}$ могут быть вычислены, как решение следующей задачи мат. программирования:

$$\begin{cases} \vec{m}^T \cdot P^{-1} \cdot \vec{m} \to \min \\ X^T \cdot \vec{m} = \vec{x}_0 \end{cases} \tag{5.11'}$$

Задачу 5.11' решаем методом Лагранжа:

Шаг 1. Составим функция Лагранжа:

$$L(\vec{m}, \vec{l}) = \vec{m}^T \cdot P^{-1} \cdot \vec{m} + \vec{l}^T \cdot \left(\vec{x}_0 - X^T \cdot \vec{m}\right)$$

символом l обозначен множитель Лагранжа.

$\boxed{\text{ДЗ}}$ Сколько этих l

Шаг 2. Составим необходимые условия экстремума функции Лагранжа:

$$\begin{cases} \dfrac{\partial L}{\partial \vec{m}} = 2 \cdot P^{-1} \cdot \vec{m} - X \cdot \vec{l} = 0, \\ \dfrac{\partial L}{\partial l} = \vec{x}_0 - X^T \cdot \vec{m} = 0. \end{cases} \tag{5.14}$$

Шаг 3. Система 5.14 решается аналитически или методом подстановки:

$$\vec{m} = P \cdot X \cdot \left(X^T \cdot P \cdot X\right)^{-1} \cdot \vec{x}_0 \tag{5.15}$$

Следовательно, наилучшая оценка расчитавается по правилу 5.16:

$$\widetilde{y}_0 = \vec{m}^T \cdot \vec{y} = \vec{x}_0^T \cdot \left(X^T \cdot P \cdot X\right)^{-1} \cdot \vec{y} \tag{5.16}$$

И последнее дейстие

$$y_0 = \vec{x}_0^T \cdot \vec{a} \Rightarrow \widetilde{\vec{a}} = \left(\left(X^T \cdot P \cdot X\right)^{-1} \cdot X^T\right)(= M) \cdot P \cdot \vec{y} \blacksquare$$

Какие размеры имеет матрица $\vec{m}$. Показать, что мат. ожидание 5.17 совпадает с вектором $\vec{a}$. Каждый элемент имеет наименьшую дисперсию в классе линейных процедур.

Следствие. Если матрица P является еденичной, что соответсвует события ..., то формула 5.17 превращается в формулу.

...

Обратимся к 5.17 и увидим, что

Лекция №9

Эконометрические модели с гетероскедастичными случайными возмущениями

План

1. Оценка параметров линейной эконометрической модели методом максимального правдоподобия и взаимосвязь этого метода с методом наименьших квадратов;
2. Тест Голдфелда-Кванта гомоскедастичности случайного возмущения в линейной эконометрической модели;
3. Простейшая модель гетероскедастичности случайного возмущения.

На прошлой лекции обсудили обобщение метода наименьших квадратов, которое (обобщение) сформулировано при отказе от предпосылки (2) и (3) теоремы Гаусса-Маркова

Предпосылка №2

$$Var(u_1) = Var(u_2) = \dots = Var(u_n)$$

Предпосылка №2 означает независимость дисперсий случайных возмущений от значений объясняющих переменных. И в этом случае случайное возмущение в моделе и в этом случае случайные возущения называются гомоскедастичным.

Предпосылка №3

$$Cov(u_i, u_j) = 0,\ i \neq j$$

Предпосылка №3 обозначает некоррелируемость или предпосылка об отсутсвии автокорреляции у случайного возмущения.

При отказе от этих предпосылок ковариационная матрица является не диагональной (полной) причём на главной диагонали матрицы расположены, возможно неодинаковые дисперсии случайных возмущений.

Отказ от предпосылок №2, №3 генерирует общую структуру уравнений вектора наблюдений.

$$Cov\left(\vec{u}, \vec{u}\right) = \sigma_0^2 \cdot P^{-1}$$

В этой общей ситуации наилучшая оценка коэффициентов модели рассчитавается по правилу:

$$\tilde{\vec{a}} = \left(\left(X^T \cdot P \cdot X\right)^{-1} \cdot X^T \cdot P\right)(= M) \cdot \vec{y}$$

Замечание. Оптимальность оценки коэффициентов спрадлива при любом законе распределения случайного возмущения.

Предположим, что случайное возмущение в модели имеет **нормальный закон** распределения. И это значит, что плотность вероятности вектора случайных возмущений имеет следующее аналитическое выражение:

$$f_{\vec{u}}\left(\vec{q}\right) = \frac{1}{(2 \cdot \pi)^{\frac{n}{2}} \cdot |\Omega_{\vec{u}}|^{\frac{1}{2}}} \exp\left\{-\frac{\vec{q}^{\,T} \cdot \Omega_{\vec{u}}^{-1} \cdot \vec{q}}{2}\right\}$$

$\vec{q}$ – вектор аргументов плотности вероятности.

Оценим параметры линейной эконометрической методом максимального правдоподобия по обучающей выборке $\left(\vec{y}, X\right)$. Функция правдоподобия выборки имеет следующий вид:

$$L = (2\cdot\pi)^{-\frac{n}{2}}\sigma_0^{-n}\cdot|P|^{\frac{1}{2}}\cdot\exp\left\{-\frac{\left(\vec{y}-X\cdot\vec{a}\right)^T\cdot P\cdot\left(\vec{y}-X\cdot\vec{a}\right)}{2\cdot\sigma_0^2}\right\}$$

Максимизация этой функции по искомым параметрам $\vec{a}, \sigma_0^2$ приводит к следующим оценкам параметра модели методом максимального правдоподобия:

$$\hat{\vec{a}} = \left(X^T\cdot P\cdot X\right)^{-1}\cdot X^T\cdot P\cdot\vec{y}$$

$$\hat{\sigma}_0^2 = \frac{\tilde{\vec{u}}^T\cdot P\cdot\tilde{\vec{u}}}{n}$$

Оценки методом максимального правдоподобия совпадают с оценкой обобщённым методом наименьших квадратов.

Вывод: при нормальном законе распределения случайного возмущения оценки наименьших квадратов совпадают с оценками максимального правдоподобия. Оценки максимального правдоподобия обладают многими замечательными свойствами, которые переносятся, тем самым, на оценки наименьших квадратов.

Линейные эконометрические модели с гетероскедастичными случайными возмущениями

Вернёмся к спецификации базовой модели эконометрики и теперь будем предполагать, что дисперсия случайного возмущения зависит от значений объясняющих переменных, это значит, что в предположении Гаусса-Маркова оказывается нарушена предпосылка №2, остальные предпосылки мы предпологаем справедливыми и при этом мы постулируем нормальный закон распределения у случайного возмущения

$$\begin{cases} y = a_0 + a_1\cdot x_1 + a_2\cdot x_2 + \ldots a_k\cdot x_k + u \\ E(u) = 0; E\left(u^2\right) = \sigma_u^2 \end{cases}$$

Теорема. (О законе распределения суммы квадратов оценок случайных возмущений ESS в уравнения наблюдений)

Пусть справедливы все предпосылки и случайные возмущения имеют нормальный закон распределения:

1) $E(u_1) = E(u_2) = \ldots = E(u_n);$
2) $Var\left(u_1^2\right) = Var\left(u_2^2\right) = \ldots = Var\left(u_n^2\right) = \sigma_n^2;$
3) $Cov(u_i, u_j) = 0,\ i \neq j$
4) $Cov(u_i, x_{lj}) = 0,$ при всех i, j, l

Тогда $\frac{ESS}{\sigma_u^2}$ имеет хи-квадрат распределение $p_{\chi^2_m}(q)$ с числом степеней свободы $m = n-(k+1)$.

Тест Голдфелда-Кванта гипотезы о гомоскедастичности случайного возмущения

Шаги теста предпосылки №2 Голдфелда-Кванта

Шаг 1. Составляется система уравнений наблюдений объекта

$$\begin{cases} y_1 = a_0 + a_1 \cdot x_{1,1} + a_2 \cdot x_{2,1} + \ldots a_k \cdot x_{k,1} + u_1 \\ y_2 = a_0 + a_1 \cdot x_{1,2} + a_2 \cdot x_{2,2} + \ldots a_k \cdot x_{k,2} + u_2 \\ \ldots \\ y_n = a_0 + a_1 \cdot x_{1,n} + a_2 \cdot x_{2,n} + \ldots a_k \cdot x_{k,n} + u_n \end{cases}$$

Замечание. Если справедлива предпосылка №2 теоремы Гаусса-Маркова, то при любых перестановках уравнений наблюдений дисперсии случайных возмущений остаются неизменными. Если же предпосылка №2 нарушается, то, как правило, дисперсии случайных возмущений в уравнения 1 и 2 возрастают (или убывают) в ответ (по мере) на возрастание абсолютных значений объясняющих переменных.

Шаг 2. Уравения наблюдений упорядычеваются по возрастанию сумм абсолютных значений объясняющих перменных

$$\sum_{j=1}^{k} |x_{ji}|$$

Шаг 3. По первым n_1 упорядоченным уравениям оцениваются методом наименьших квадратов параметры модели и запоминается значения ESS_1. Количество n_1 выбирается согласно следующим двум условиям:

$$a)\ n_1 \approx \frac{1}{3}n,\quad b)\ n_1 > k + 1$$

Аналогично оценивается модель по последним n_1 уравениям и запоминается значение ESS_2.

Шаг 4. Вычисляется по следующему правилу дробь:

$$GQ = \frac{ESS_1}{ESS_2} \sim P_F(q) \tag{6.1.10}$$

Эта дробь является <u>статистикой</u> критерия проверяемой гипотезы о гомоскедастичности случайного возмущения. Величина GQ имеет распредение Фишера с кол-ом степеней свободы m, n.

Шаг 5. Гипотеза о гомоскедастичности принимается как не противоречащая реальным данным, если оказываеются справедливыми следующие два неравенства:

$$\begin{cases} GQ \overset{?}{\leq} F_{\text{крит}} \\ \dfrac{1}{GQ} \overset{?}{\leq} F_{\text{крит}} \end{cases}$$

Где символом $F_{\text{крит}}$ мы обозначаем квантиль распределения Фишера заданного уровня $1-\alpha$, например $1-\alpha\ = 0.95$.

Итог: экономисты тестируют все предпосылки в частности предпосылка №2 тестируется тестом Голдфелда-Кванта.

Лекция №10

Модель гетероскедастичности случайного возмущения и оценивания эконометрических моделей с гетероскедастичными возмущениями взвешанным методом наименьших квадратов

План

1. Модель гетероскедастичности случайного возмущения;
2. Трансформация модели с гетескедастичным возмущением к модели с гомоскедастичным возмущением;
3. Оценивания эконометрических моделей с гетероскедастичными возмущениями взвешанным МНК;

На прошлой лекции мы обсудили тест Голдфилда-Кванта гомоскедастичности случайного возмущения модели. На практике предпосылка гомоскедастичности часто нарушается. В такой ситуации процедура МНК утрачивает свойство оптимальности (теряет свойство эффекстивности) и, самое главное, точностные характеристики оценок становятся некорректными. На сегодняшнем занятии мы обсудим оптимальную процедуру оценивания параметров модели с гетескедастичным возмущением.

Нам потребуетсся моделт гетерскедостичности случайного возмущения. Вот уравнение этой модели:

$$Var(u) = \sigma_u^2 = \sigma_0^2 \cdot \left(\sum_{j=0}^{k} |x_j| \right)^{\lambda} \tag{6.2.5}$$

В этом уравнении присутсвует два параметра: $\sigma_0^2 > 0$ (дисперсия единицы веса) и λ – некоторое априорно заданное число (подбирается эксперементально):

$$\lambda = 0, \pm\frac{1}{2}, \pm 1, \pm 2$$

$$p = \frac{\sigma_0^2}{\sigma_u^2} = \frac{1}{\left(\sum_{j=0}^{k} |x_j| \right)^{\lambda}} \quad \text{– вес случайного возмущения} \tag{6.2.6}$$

Трансформация модели с гетескедастичным возмущением к модели с гомоскедастичным возмущением

Пусть в линейной модели множественной регрессии случайное возмущение гетероскедастично и известна модель веса этого возмущения.

$$p = \frac{\sigma_0^2}{\sigma_u^2}$$

Умножим уравнение этой модели на квадратный из веса:

$$\begin{cases} \sqrt{p} \cdot y = \sqrt{p} \cdot a_0 + \sqrt{p} \cdot a_1 x_1 + \sqrt{p} \cdot a_2 x_2 + \ldots + \sqrt{p} \cdot a_k x_k + \sqrt{p} \cdot u \\ E\left(\sqrt{p} \cdot u\right) = 0; \;\; E\left(\left(\sqrt{p} \cdot u\right)^2\right) = \sigma_0^2 \end{cases}$$

$\boxed{\text{ДЗ (необязательное)}}$ Можно проверить, что $E\left(\sqrt{p}\cdot u\right) = 0$, а дисперсия в точности совпадает с константой σ_0^2.

Оценивание ЛММР с гетероскедастичным случайным возмушением взвешанным МНК

Пусть случайные возмущения в ЛММР гетескедастично, но при этом все остальные предпосылки теоремы Гаусса-Маркова.

$$\vec{y} = X\,(\text{ранг совпадает с кол} - \text{ом столбцов}) \cdot \vec{a} + \vec{u}$$

0. X- линейно независим.

1) $E(u_1) = E(u_2) = \dots = E(u_n)$;
2) $Var\left(u_1^2\right) = Var\left(u_2^2\right) = \dots = Var\left(u_n^2\right) = \sigma_n^2$;
3) $Cov(u_i, u_j) = 0,\ i \neq j$
4) $Cov(u_i, x_{lj}) = 0,$ при всех i, j, l

Пусть коварицонная матрица имеет следующий вид:

$$\Omega_{\vec{u}} = \sigma_0^2 \cdot \begin{pmatrix} \frac{1}{\sqrt{p_1}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{p_2}} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \frac{1}{\sqrt{p_n}} \end{pmatrix}^2$$

Составим уравнения наблюдений в рамках трансформированной модели

$$\begin{cases} \sqrt{p}\cdot y = \sqrt{p}\cdot a_0 + \sqrt{p}\cdot a_1 x_1 + \sqrt{p}\cdot a_2 x_2 + \ldots + \sqrt{p}\cdot a_k x_k + \sqrt{p}\cdot u \\ E\left(\sqrt{p}\cdot u\right) = 0;\ \ E\left(\left(\sqrt{p}\cdot u\right)^2\right) = \sigma_0^2 \end{cases}$$

$$v = \sqrt{p}\cdot u$$

$$\begin{cases} \sqrt{p}\cdot y = \sqrt{p}\cdot a_0 + \sqrt{p}\cdot a_1 x_1 + \sqrt{p}\cdot a_2 x_2 + \ldots + \sqrt{p}\cdot a_k x_k + v \\ E(v) = 0;\ \ E\left(v^2\right) = \sigma_0^2 \end{cases}$$

$$\underbrace{\boxed{P^{\frac{1}{2}}\cdot\vec{y}}}_{\vec{y}'} = \underbrace{\boxed{P^{\frac{1}{2}}\cdot X}}_{X'} \cdot \vec{a} + \vec{v} \qquad (6.4.5)$$

Символом $P^{\frac{1}{2}}$ обозначена следующая квадратная матрица:

$$P^{\frac{1}{2}} = \begin{pmatrix} \frac{1}{\sqrt{p_1}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{p_2}} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \frac{1}{\sqrt{p_n}} \end{pmatrix}$$

$\Omega_{\vec{v}} = \sigma_0^2 \cdot E$ – скалярная

$$\Omega_{\vec{u}} = \sigma_0^2 \cdot P^{-1} \tag{6.4.6}$$

Получается, что уравнения наблюдений (6.4.5) удовлетворяют всем предпосылкам и это значенит, что по всем этим уравнеям мы можем оценить параметры с помощью МНК.

$\boxed{\text{ДЗ(необязательно)}}$ Проверить, что утверждение a), b), c), d) теоремы Гаусса-Маркова применительно к уравнениям наблюдения 6.4.5 превращаются в следующие утверждения:

$$a)\vec{\tilde{a}} = \left(X^T \cdot P \cdot X\right)^{-1} \cdot X^T \cdot P \cdot \vec{y} = Q \cdot X^T \cdot P \cdot \vec{y} \tag{6.4.7}$$

$$b)\tilde{\sigma}_0^2 = \frac{\sum_{i=1}^{n} \tilde{v}_i^2}{n-(k+1)} = \frac{\sum_{i=1}^{n} p_i \cdot \tilde{u}_i^2}{n-(k+1)} \tag{6.4.8}$$

$$\tilde{u}_i = y_i - \left(\tilde{a}_0 + \tilde{a}_1 \cdot x_{1,i} + \ldots + \tilde{a}_k \cdot x_{k,i}\right) \tag{6.4.9}$$

$$\tilde{v}_i = \sqrt{p_i} \cdot \tilde{u}_i \tag{6.4.10}$$

$$c)\ \sum_{i=1}^{n} \tilde{v}_i^2 = \sum_{i=1}^{n} p_i \cdot \tilde{u}_i^2 \to \min \tag{6.4.11}$$

$$d) \begin{cases} S\tilde{a}_j = \tilde{\sigma}_0 \cdot \sqrt{q_{j+1\ j+1}} \\ j\ = 0, 1,\ \dots, k \end{cases} \tag{6.4.12}$$

Свойство $c)$ оценок из утверждения $a)$ принято называть *взвешанными наименьшими квадратами,* что является причиной общепринятого названия формулы процедуры (6.4.7) ВМНК

Лекция № 11

Линейные эконометрические модели с автокоррелированными случайными возмущениями (нарушена предпосылка №3 теоремы Гаусса-Маркова)

План

1. Тест Дарбина-Уотсена об отсутствии автокорреляции у случайного возмущетния (причиной нарушения этой предпосылки является пропуск объясняющей переменной);
2. Фундаментальная модель автокорреляции случайного возмущения (модель авторегрессии первого порядка);
3. Трансформации модели с автокоррелированным случайным возмущением к модели, где справедлива предпосылка № 3 теоремы Гаусса-Маркова;
4. Оценивание трансформированной модели нелинейным итерационным методом наименьших квадратов.

Предпосылка №2

$$Var(u_1) = Var(u_2) = \ldots = Var(u_n)$$

Предпосылка № 2 означает независимость дисперсий случайных возмущений от значений объясняющих переменных. И в этом случае случайное возмущение в моделе и в этом случае случайные возущения называются <u>гомоскедастичным</u>.

Предпосылка №3

$$Cov(u_i, u_j) = 0,\ i \neq j$$

Предпосылка № 3 обозначает некоррелируемость или предпосылка об отсутсвии автокорреляции у случайного возмущения.

На предыдущей лекции приступили к посроению оптимальных процедур оценивания эконометрических моделей в ситуации, когда оказываются нарушенными предпосылки №2, № 3 теоремы Гаусса-Маркова. Обсудили оптимальную процедуру оценивания модели с гетескедастичными случайными возмущениями.

Сегодня мы обсудим тест предпосылки № 3, затем обсудим фундаментальную модель автокорреляции случайного возмущения и наконец изучим процедуру оценивания модели с автокоррелированным случайным возмущением.

Рассмотрим базовую модель эконометрики:

$$\begin{cases} y = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + \ldots a_k \cdot x_k + u_t \\ E(u_t) = 0; E\left(u_t^2\right) = \sigma_u^2 \end{cases} \tag{1.1}$$

В ситуации когда нарушена предпосылка № 3 теоремы Гаусса-Маркова, а все остальные предпосылки справедливы. Тест предпосылки № 3 базируется на следующей теореме.

Теорема. Пусть в модели (1.1) справеливы все предпосылки Гаусса-Маркова и составлены все уравнения наблюдений:

$$\begin{cases} y_1 = a_0 + a_1 \cdot x_{1,1} + a_2 \cdot x_{2,1} + \ldots a_k \cdot x_{k,1} + u_1 \\ y_2 = a_0 + a_1 \cdot x_{1,2} + a_2 \cdot x_{2,2} + \ldots a_k \cdot x_{k,2} + u_2 \\ \ldots \\ y_n = a_0 + a_1 \cdot x_{1,n} + a_2 \cdot x_{2,n} + \ldots a_k \cdot x_{k,n} + u_n \end{cases} \tag{1.2}$$

Тогда следующая случайная величина V вычесленная по правилу (1.5)

$$V = \frac{\sum_{i=1}^{n-1}\left(\widetilde{u}_{i+1} - u_i\right)^2}{\sum_{i=1}^{n}\left(\widetilde{u}_i\right)^2} \qquad (1.6)$$

имеет математическое ожидание примерное равное 2; математическое ожидание числителя и знаменателя рассчитываются по следующим правилам:

$$E\left(\sum_{i=1}^{n-1}\left(\widetilde{u}_{i+1} - u_i\right)^2\right) = 2\cdot(n-1)\cdot\sigma_u^2$$

$$E\left(\sum_{i=1}^{n}\left(\widetilde{u}_i\right)^2\right) = n\cdot\sigma_u^2;$$

$$\Longrightarrow E(V) \approx 2$$

Тест Дарбина-Уотсена гипотезы о некоррелированности случайных возмущений в уравнениях наблюдений (1.2)

Проверяемая гипотеза в этом тесте имеет вид:

$$H_0 : Cov(u_{t+1}, u_t) = 0 \qquad (1.7)$$

Альтернативная гипотеза заключается в положительном значении ковариации u_{t+1}, u_t:

$$H_1 : Cov(u_{t+1}, u_t) > 0 \qquad (1.8)$$

Замечание. Альтернатива (1.8) имеет наиболее важное для практики значение. Если справедлива данная альтернатива, то причина этого обстоятельства чаще всего заключается в ошибочной спецификации модели (1.1). Например, в отстутсвии в этой модели значящих объясняющих переменных.

Тест $DW(\text{Дарбина} - \text{Уотсена})$ проводится в итоге следующих шагов:

Шаг 1. По уравнениям наблюдений оценивается модель (1.1) и вычисляется по правилу (1.9)

$$DW = \frac{\sum_{i=1}^{n-1}\left(\widetilde{u}_{i+1} - u_i\right)^2}{\sum_{i=1}^{n}\left(\widetilde{u}_i\right)^2} \qquad (1.9)$$

статистика критерия гипотезы H_0.

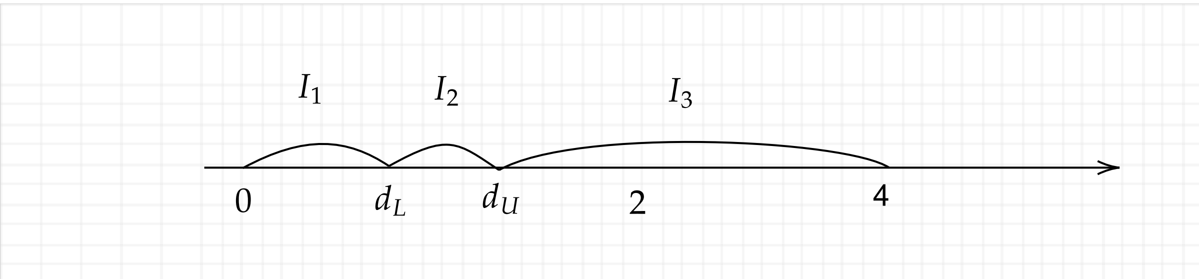
Шаг 2. По таблицам Дарбина-Уотсена

n	$k^1=1$		$k^1=2$		$k^1=3$		$k^1=4$		$k^1=5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
6	0,61	1,40	–	–	–	–				
7	0,70	1,36	0,47	1,90	–	–				
8	0,76	1,33	0,56	1,78	0,37	2,29				
9	0,82	1,32	0,63	1,70	0,46	2,13				
10	0,88	1,32	0,70	1,64	0,53	2,02				
11	0,93	1,32	0,66	1,60	0,60	1,93				
12	0,97	1,33	0,81	1,58	0,66	1,86				
13	1,01	1,34	0,86	1,56	0,72	1,82				
14	1.05	1.35	0.91	1.55	0.77	1.78				
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83

Figure 1: Значения статистики Дарбина - Уотсона

Выбираются две величины $d_L,\ d_U$ используя два входа $n,\ k$.

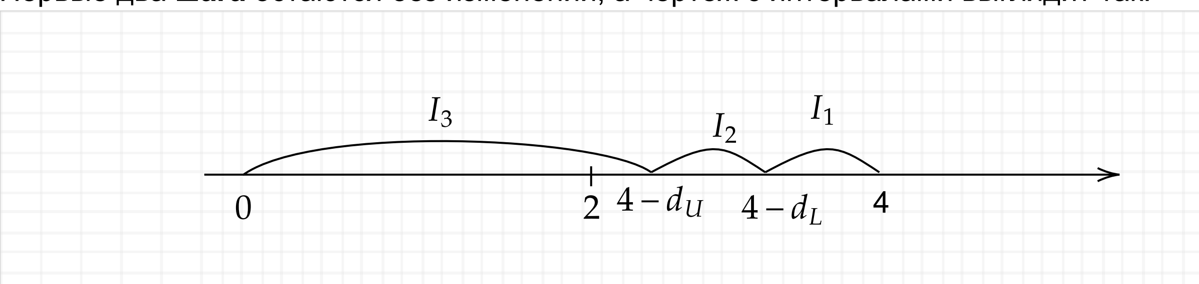
Шаг 3. Определяется один из трёх интервалов в который попадает статистика DW.



Если DW попало в I_3, то H_0 принимается, если в I_1, то гипотеза H_0 отвергается в пользу гипотезы H_1; если в интервал I_2, то ничего сказать нельзя - это *интервал неопределённости*.

Проверка гипотезы (1.7) при альтернативе $Cov(u_{t+1}, u_t) < 0$

Первые два **шага** остаются без изменений, а чертёж с интервалами выглядит так:



Если статистика DW попадает I_1, то гипотеза H_0 отклонятся в пользу гипотезы H_1 (очень редкий случай), если в I_2 то ничего сказать нельзя - это *интервал неопределённости;* Кесли DW попадает в I_3, то H_0 принимается.

$\boxed{\text{ДЗ}}$ Сформулировать тест Дарбина Уотсена на основании обсуждённого материала проверки гипотезы H_0 против альтернативы $Cov(u_{t+1}, u_t) \neq 0$.

$\boxed{\text{ДЗ}}$ В выражении (1.9) статистики Дарбина-Уотсена раскрыть в числителе квадрат разности двух чисел и после преобразования этой формулы показать справделивость следующего утверждения: обозначим символом ρ коэффициент корреляции случайных остатков модели в два соседние момента времени. Тогда:

1. Если $\rho \to +1$, то $DW \to 0^+$; если $\rho \to 0$, то $DW \to 2$. Если $\rho \to -1$, то $DW \to 0^-$. **Указание.** Коэффициент корреляции ρ (точнее оценка $\widetilde{\rho}$), может быть вычислено по правилу: $\widetilde{\rho} = \dfrac{\sum_{i=1}^{n-1} u_i \cdot u_{i+1}}{\sum_{i=1}^{n} \left(\widetilde{u}_i\right)^2}$.

Замечание. В тесте Дарбина Уотсена предполагается, что *во-первых:* в модели (1.1) присутсвует свободный член a_0, *второе:* среди объясняющих переменных нет *лаговых* значений эндогенной переменной y_t.

Выходим в соответсвующую модель этой предпосылки. Ниже потребуется модель автокорреляции случайного возмущения u_t в линейного модели множественной регресии. Эта модель имеет абривиатуру $AR(1)$ и её спецификация выглядит так:

$$\begin{cases} u_t = \rho \cdot u_{t-1} + \xi_t, \\ Var(u_t) \equiv \sigma_u^2, \\ |\rho| < 1, \\ Var(\xi_t) \equiv \sigma_\xi^2 \end{cases} \tag{2.2}$$

Первая строчка этой модели описывает процесс формирования значения случайного возмущения в период времени t. Значение u_t складывается из двух величин, а именно: из части случайного возмущения в предшествующий период времени $\rho \cdot u_{t-1}$ и независимой случайной величины ξ_t[кси тэ], которая имеет $E(\xi) = 0$, постоянную дисперсию $Var(\xi_t) \equiv \sigma_\varepsilon^2$ и некоррелированные уровни во все периоды времени. Величину ξ_t принято называть *белым шумом.* Позже проверим, что параметр ρ равен коэффициенту корреляции u_t, u_{t-1}.

Трансформации модели с автокоррелированным случайным возмущением к модели, где справедлива предпосылка № 3 теоремы Гаусса-Маркова

Вернёмся к спецификации текущей модели и лаконично обозначим функцию регрессии без свободного члена $\vec{a}^T \cdot \vec{x}_t$, то есть:

$$y = a_0 + \vec{a}^T \cdot \vec{x}_t + u_t$$

С учётом модели мы можем переписать спецификаю в следующем виде:

$$\begin{cases} y_t = a_0 + \vec{a}^T \cdot \vec{x}_t + u_t \\ E(u_t) = 0; E\left(u_t^2\right) = \sigma_u^2 = \dfrac{\sigma_\xi^2}{1-\rho^2}; \\ u_t = \rho \cdot u_{t-1} + \xi_t. \end{cases} \qquad (2.5)$$

Выпишем уравнение модели в период времени $t-1$:

$$y_{t-1} = a_0 + \vec{a}^T \cdot \vec{x}_{t-1} + u_{t-1}$$

Предполагая, что ρ известно умножим последнее уравнение на ρ в итоге получим:

$$\rho \cdot y_{t-1} = \rho \cdot a_0 + \rho \cdot \vec{a}^T \cdot \vec{x}_{t-1} + \rho \cdot u_{t-1} \qquad (2.6)$$

Наконец из первого уравнения в спецификации (2.5) вычтем (2.6) в итоге получим спецификаю трансформированной модели (2.7):

$$y_t - \rho \cdot y_{t-1} = a_0 \cdot (1-\rho) + \vec{a}^T \cdot \left(\vec{x}_t - \rho \cdot \vec{x}_{t-1}\right) + \xi_t$$

$$E(\xi) = 0,\ E\left(\xi^2\right) = \sigma_\xi^\alpha, Cov(\xi_t, \xi_{t-1}) = 0$$

Случайное возмущение ξ_t является белым шумом и удовлетворяется всем предпосылкам Гаусса-Маркова.

Лекции №12

Характеристики качества спецификации эконометрических моделей

План

1. Оценивание эконометрической модели с автокоррелированным случайным возмущением нелинейным итерационным методом наименьших квадратов;
2. Коэффицент детерминации модели, как количественная характеристика качества выбора экономиста объясняющих переменных модели;
3. F – тест качества спецификации эконометрической модели.
4. Скорректированный коэффициент детерминации, как инструмент модификации модели ();

Инструмент отбора в модель объясняющих переменных

На прошой леции обсудили тест Дарбина-Уотсена. Статистика выглядит следующим образом:

$$DW = \frac{\sum_{i=1}^{n-1}\left(\widetilde{u}_{i+1} - u_i\right)^2}{\sum_{i=1}^{n}\left(\widetilde{u}_i\right)^2}$$

Авторегрессивная модедь первого порадка:

$$\begin{cases} y_t = a_0 + \vec{a}^T \cdot \vec{x}_t + u_t \\ E(u_t) = 0; E\left(u_t^2\right) = \sigma_u^2 = \dfrac{\sigma_\xi^2}{1-\rho^2}; \\ u_t = \rho \cdot u_{t-1} + \xi_t. \end{cases}$$

Трансформировали к модели неавтокоррелированным остатком:

$$\begin{cases} y_t - \rho \cdot y_{t-1} = a_0 \cdot (1-\rho) + \vec{a}^T \cdot (\vec{x}_t - \rho \cdot \vec{x}_{t-1}) + \xi_t \\ E(\xi) = 0,\ E\left(\xi^2\right) = \sigma_\xi^\alpha, Cov(\xi_t, \xi_{t-1}) = 0 \end{cases} \qquad (2.7)$$

Обратимся к трансформированной модели (2.7)

Если бы параметр ρ в трансформированной модели был известен, то в моделе (2.7) были бы справедливы все предпосылки Гаусса-Маркова и можно было бы оценить параметры этой модели. В частности коэффициенты $a_0,\ \vec{a}^T$ методом наименьших квадратов (смотри теорему Гаусса-Маркова). Параметры (2.7) не относятся к линейным моделям так как $a_0 \cdot (1-\rho)$ противоречит линейности. А уравнения наблюдений в этой модели:

$$\begin{cases} y_2 - \rho \cdot y_1 = a_0 \cdot (1-\rho) + \vec{a}^T \cdot (\vec{x}_2 - \rho \cdot \vec{x}_1) + \xi_2 \\ y_3 - \rho \cdot y_2 = a_0 \cdot (1-\rho) + \vec{a}^T \cdot (\vec{x}_3 - \rho \cdot \vec{x}_2) + \xi_3 \\ \dots \\ y_n - \rho \cdot y_{n-1} = a_0 \cdot (1-\rho) + \vec{a}^T \cdot (\vec{x}_n - \rho \cdot \vec{x}_{n-1}) + \xi_n \end{cases} \qquad (2.8)$$

не образуют схему Гаусса-Маркова. Процедура оценивания параметров можели (2.7) осуществляется нелинейным итерационным методом наименьших квадратов в результате следующих шагов:

Шаг 1. Задаётся множество пробных значений параметра $\rho = 0.1;\ 0.2;\ \dots 0.9$;

Шаг 2. Выбирается первое пробное значение и это пробное значение подстваляется в (2.8) в итоге эти уравнения превращаются в схему Гаусса-Маркова. По уравнениям (2.8) оцениваются параметры можели коэффициенты модели $a_1,\ \dots, a_k$ и вычисляется сумма квадратов остатков:

$$ESS = \sum_{i=2}^{n} \tilde{\xi}_i^2 (p(= \text{допустим } 0.1)) \to \min \qquad (2.9)$$

Шаг 3. Повторяется при других значениях ρ и выбирается такое значение ρ при котором оказывается справедливым условие (2.9) (условие наименьших квадратов). Именно при этом ρ получаются условия оценки коэффициентов модели. Кроме того по оценки дисперсии белого шума рассчитавается оценка дисперсии случайного возмущения исходной модели $\sigma_u^2 = \dfrac{\sigma_\xi^2}{1-\rho^2}$.

Вывод: в основании процедуры оценивания модели с автокоррелированным случайным возмущением лежит фундаметальная модель автокорреляции:

$$\begin{cases} u_t = \rho \cdot u_{t-1} + \xi_t, \\ Var(u_t) \equiv \sigma_u^2, \\ |\rho| < 1, \\ Var(\xi_t) \equiv \sigma_\xi^2 \end{cases}$$

и оценивание осуществляется нелинейным методом наименьших квадратов.

На двух предыдущих лекциях мы обсудили процедуры оценивания параметров базовой модели эконометрики (2.7) в ситуации, когда оказываются нарушенными предпосылки №2 и №3 теоремы Гаусса-Маркова. Если оказываются одновременно нарушенными обе эти предпосылки, то наилучшие оценки параметров модели вычисляются обобщённым методом наименьших квадратов, который мы обсудили на лекции №8.

Качество спецификации эконометрической модели

Оценивание эконометрической модели осуществляется на 3-ем этапе схемы её построения. На этом же этапе появляется возможность исследовать качество выбора объясняющих переменных модели. Простешей характеристикой качества служит коэффициент детерминации моеди, который по традиции обозначается символом R^2. Отметим смысл R^2 – это доля эндогенной переменной модели (точнее доля её дисперсии), которая (доля) объясняется предопределёнными переменными модели.

Выдем формулу для величины R^2 приминительно модели с одной объясняющей переменной.

Шаг 1. По уравнениям наблюдений рассчитываем оценки случайных возмущений

$$\tilde{u}_i = y_i - \left(\tilde{a}_0 + \tilde{a}_1 \cdot x_i\right) = y_i - \tilde{y}_i$$

Перепишем следующим образом уравнения наблюдений

$$\begin{cases} y_1 = \widetilde{y}_1 + \widetilde{u}_1 \\ y_2 = \widetilde{y}_2 + \widetilde{u}_2 \\ \dots \\ y_n = \widetilde{y}_n + \widetilde{u}_n \end{cases} \tag{4.14}$$

В уравнении (4.14) первое слагаемое в правой части объясняются перменной x, а вторые слагаемые необъясняются x-ами. Справедливо, следующая теорема:

$$\sum (y_i - \overline{y})^2 = \sum \left(\widetilde{y}_i - \overline{\widetilde{y}}\right)^2 + \sum \widetilde{u}_i^2 \tag{4.15}$$

В левой части тождества размещается характеристика изменчивости эндогенной переменной - *волатильность.* Первое слагаемое в правой части обозначим его символом RSS объясняется изменчивостью предопределённых значений и полность объясняется x. А второе слагаемое пораждено неучтёнными факторами $ESS = \sum \widetilde{u}_i^2$, левую часть $TSS = \sum \left(\widetilde{y}_i - \overline{\widetilde{y}}\right)^2$. И разделим обе часть тождества (4.15) на велечину TSS в итоге придём у формуле (4.16).

$$R^2 = 1 - \frac{ESS}{TSS} \tag{4.16}$$

Рассматривая (4.16) мы констатируем, что R^2 – это доля эндогенной переменной модели, которая объясняется предопределёнными перменными R^2.

F – тест качества спецификации эконометрической модели

F – тест является формализированной процедурой проверки гипотезы о неудовлетворительной спецификации эконометрической модели:

$$H_0 : a_1 = a_2 = \dots = a_k = 0$$

То есть гипотеза о том, что ни одна объясняющая переменная не несёт в себе информацию об эндогенной переменной y. Альтернативой для H_0 служит гипотеза:

$$H_1 = \overline{H_0}$$

Означающая, что хотя бы один из коэффициентов отличны от нуля.

Порядок F – теста

Шаг 1. Создаваемая модель оценивается методом наименьших квадратов и рассчитывается статистика F критерия гипотезы H_0:

$$F = \frac{R^2 / k}{\left(1 - R^2\right) / (n - (k + 1))}$$

Если эта гипотеза верна, то случайная переменная F имеет закон распределения Фишера с кол-ами степеней свободы $k,\ n - k + 1$. Если велина F превосходит квантиль распределения Фишера уровня $1 - \alpha$, где $\alpha = 0.01\ - 0.05$, то гипотеза H_0 отвергается. Эта квантиль обозначена $F_{\text{крит}}$.

Вывод: F – тест позволяет объективно объяснить качество перменных модели.

Лекция №13

Прогнозирование по оценённой эконометрической модели и проверка её адекватности

План

1. Завершение темы "Характеристики качества спецификации эконометрических моделей";
2. Оптимальный точечны прогноз и характеристика точности прогноза (стандартная ошибка прогноза);
3. Интервальное прогнозирование по модели и проверка её адекватности;

В конце предыдущей лекции мы обсудили $F-$тест качества спецификации модели. Подчеркнем, что если в итоге $F-$теста принимается гипотеза $H_0: a_1 = a_2 = \ldots = a_k = 0$, то это означает, то ни одна из объясняющих переменных не содержит в себе информацию о эндогенной переменной модели. Если же H_0 отклоняется в пользу гипотезы $H_1 = \overline{H_0}$, это значит, что хотя бы одна экзогенная переменная содержит информацию об эндогенной переменной модели.

Скорректированнный коэффициент детерминации, как интструмент модификации модели

Скоррективонная коэффицент детерминации рассчитывается по следующему правилу:

$$\overline{R}^2 = 1 - \frac{\frac{ESS}{n-(k+1)}}{\frac{TSS}{n-1}}$$

$$\overline{R}^2 \leqslant R^2$$

Числитель и позволяет отбирать в модель объясняющие перменные. Если при включении в модель новой объясняющей переменной велечина $\overline{R}^2$ возрастает (лучше сказать не убывает), то включение этой переменной в модель полезно, если убывает, то бесмысленно.

Замечание. Первое: В числителе вычитаемого $\dfrac{ESS}{n-(k+1)}$ размещается оценка дисперсии случайного возмущения. В знаменателе находится дисперсия эндогенной переменной. При добавлении в модель новой объясняющей переменной меняется только числитель (знаменатель остаётся всегда неизменным) и поэтому увеличение $\overline{R}^2$ равносильно снижению дисперсии случайного возмущения и значит уменьшению экзогенных переменных. Добавим, что всегда имеет место следующее неравенство, при чём $\overline{R}^2$ может быть меньше 0.

Эконометрические модели создаются в частности для прогнозирования неизвестных объясняемых переменных (эндогенных переменных). Поясним на примере базовой модели эконометрики (ЛММР) существо задачи прогнозирования:

$$\begin{cases} y = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + \ldots a_k \cdot x_k + u_t \\ E(u_t) = 0; E\left(u_t^2\right) = \sigma_u^2 \end{cases} \tag{1}$$

Пусть при заданных значениях $(1, x_{1,0}, x_{2,0}, \dots, x_{k,0}) = \vec{x}_0^{\,T}$ нужно знать неизвестное значение y_0 эндогенной переменной y. Величина y_0, по предположению, задана экзогенными переменными уравнениям модели (1):

$$y_0 = a_0 + a_1 \cdot x_{1,0} + a_2 \cdot x_{2,0} + \dots a_k \cdot x_{k,0} + u_0 \qquad (2)$$

Обозначим $\tilde{y}_0$ прогноз значения y_0. Прогноз y_0 принято называть оптимальным, если оказываются справедливыми следующие два свойства этого прогноза:

$$\begin{cases} E\left(\tilde{y}_0 - y_0\right) = E\left(\triangle \tilde{y}_0\right) = 0, \\ E\left(\tilde{y}_0 - y_0\right)^2 = \sigma^2\left(\triangle \tilde{y}_0\right) \rightarrow \min \end{cases} \qquad (3)$$

Прокомментируем оба требования: первое требование означает равенство 0 математичкого ожидания истинной ошибки прогноза $\left(\tilde{y}_0 - y_0\right)$, несмящённая оценка прогноза, второе требование означает, что среднее растояние между прогнозом и истинной минимально. Подчеркнём, что эти два требования являются аналогом требованиям оптимальности к статистической процедуре оценивания моделей.
Справедлива следующая теорема:
Пусть модель (1) оценивается методом наименьших квадратов по уравнениям наблюдений, где справедливы все предпосылки теоремы Гаусса-Маркова. Тогда оптимальный прогноз рассчитывается по формуле (4):

$$A)\ \tilde{y}_0 = \tilde{a}_0 + \tilde{a}_1 \cdot x_{1,0} + \dots + \tilde{a}_k \cdot x_{k,0} \qquad (4)$$

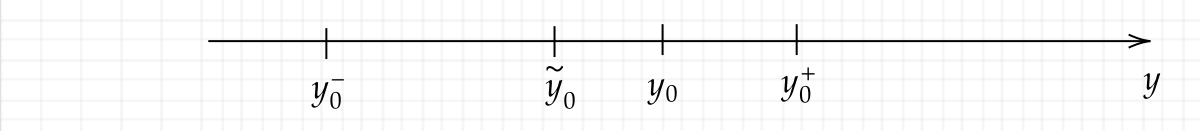
Характеристика точности прогноз:

$$\sigma^2\left(\triangle \tilde{y}_0\right) = \sigma_u^2 + \sigma_u^2 \cdot \vec{x}_0^{\,T} \cdot Q \cdot \vec{x}_0 = \sigma_u^2(1 + q_0)$$

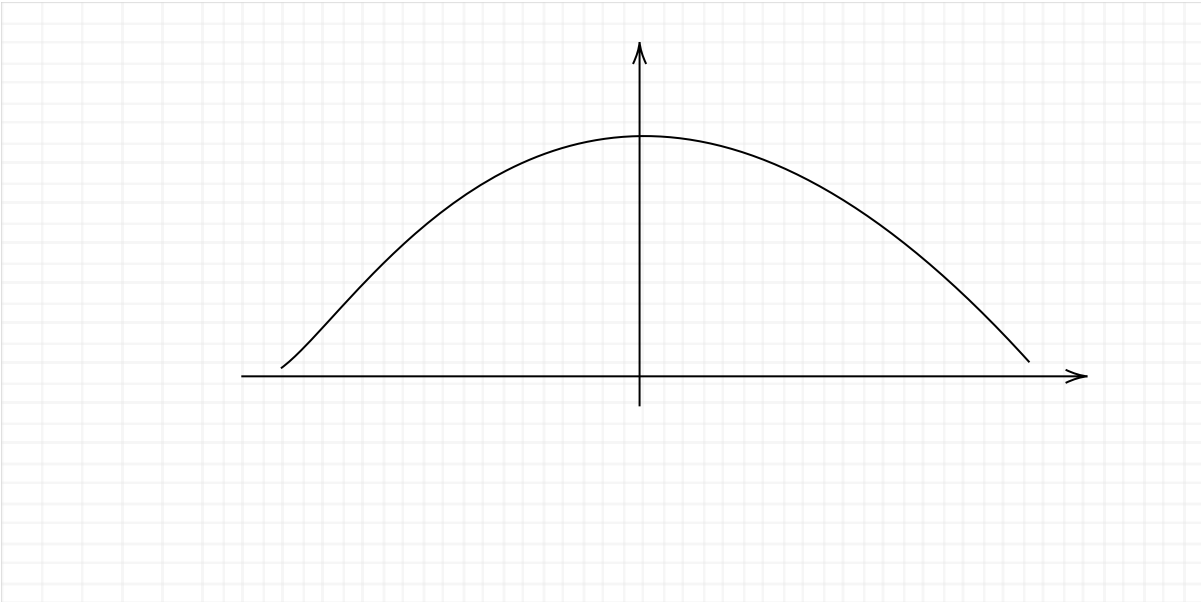
Замечание. Присутствие в характеристике точности прогноза слагаемого $\vec{x}_0$ обусловлен случайными ошибками в оценке коэффициентов модели. Первое слагаемое в характеристике обусловлено отсутвием случайного возмущения u_0.
Подводим итог, оптимальный прогноз по модели вычисляется в итоге подстановки в оценку функции регресси заданых значений случайных переменных.

Интервальное прогнозирование по модели и проверка её адекватности

В пукте выше мы обсудили оптимальный прогноз $\tilde{y}_0$.



По мимо точечного прогноза в финансово-экономической сфере прогноз искомой величины y_0 часто строится ввиде интервала с левой границей y_0^- и правой границей y_0^+. Этот интервал накрывает неизвестное значение y_0 с заданной доверительной вероятностью. В основании лежит следующая теорема:
Пусть в моделе (1) выполнены все предпосылки теоремы Гаусса-Маркова и случайные возмущения имеют нормальный закон распределения



Тогда следующая дробь:

$$t = \frac{\widetilde{y}_0 - y_0}{\widetilde{\sigma}\left(\triangle \widetilde{y}_0\right)} \sim t(m), \text{ где } m = m - k + 1$$

имеет закон распределения Стьюдента с количеством степеней свододы $n - k + 1$.
Из теоремы выше вытекает равенство (3):

$$P\left(\left|\frac{\widetilde{y}_0 - y_0}{\widetilde{\sigma}\left(\triangle \widetilde{y}_0\right)}\right| \leqslant t_{\text{крит}}\right) = 1 - \alpha \qquad (3)$$

То есть существует такое $t_{\text{крит}}$ при котором справедливо равенство (3). $t_{\text{крит}}$ имеет значение двухсторонней квантили с кол-ом степеней свободы $n - k + 1$.
Освобождаясь от модуля мы перепишем формулу в следующем виде:

$$P\left(\widetilde{y}_0 - t_{\text{крит}} \cdot \widetilde{\sigma}\left(\triangle \widetilde{y}_0\right) \leqslant y_0 \leqslant \widetilde{y}_0 + t_{\text{крит}} \cdot \widetilde{\sigma}\left(\triangle \widetilde{y}_0\right)\right) \qquad (3')$$

Из которого и следуют формулы границ доверительно интервала.

Лекция №14

Прогнозирование по оценённой эконометрической модели и проверка её адекватности

План

1. Завершение темы "Характеристики качества спецификации эконометрических моделей";
2. Оптимальный точечны прогноз и характеристика точности прогноза (стандартная ошибка прогноза);
3. Интервальное прогнозирование по модели и проверка её адекватности;

В конце предыдущей лекции мы обсудили $F-$тест качества спецификации модели. Подчеркнем, что если в итоге $F-$теста принимается гипотеза $H_0: a_1 = a_2 = \ldots = a_k = 0$, то это означает, то ни одна из объясняющих переменных не содержит в себе информацию о эндогенной переменной модели. Если же H_0 отклоняется в пользу гипотезы $H_1 = \overline{H_0}$, это значит, что хотя бы одна экзогенная переменная содержит информацию об эндогенной переменной модели.

Скорректированнный коэффициент детерминации, как интструмент модификации модели

Скоррективонная коэффицент детерминации рассчитывается по следующему правилу:

$$\overline{R}^2 = 1 - \frac{\frac{ESS}{n-(k+1)}}{\frac{TSS}{n-1}}$$

$$\overline{R}^2 \leqslant R^2$$

Числитель и позволяет отбирать в модель объясняющие перменные. Если при включении в модель новой объясняющей переменной велечина $\overline{R}^2$ возрастает (лучше сказать не убывает), то включение этой переменной в модель полезно, если убывает, то бесмысленно.

Замечание. Первое: В числителе вычитаемого $\dfrac{ESS}{n-(k+1)}$ размещается оценка дисперсии случайного возмущения. В знаменателе находится дисперсия эндогенной переменной. При добавлении в модель новой объясняющей переменной меняется только числитель (знаменатель остаётся всегда неизменным) и поэтому увеличение $\overline{R}^2$ равносильно снижению дисперсии случайного возмущения и значит уменьшению экзогенных переменных. Добавим, что всегда имеет место следующее неравенство, при чём $\overline{R}^2$ может быть меньше 0.

Эконометрические модели создаются в частности для прогнозирования неизвестных объясняемых переменных (эндогенных переменных). Поясним на примере базовой модели эконометрики (ЛММР) существо задачи прогнозирования:

$$\begin{cases} y = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + \ldots a_k \cdot x_k + u_t \\ E(u_t) = 0; E\left(u_t^2\right) = \sigma_u^2 \end{cases} \qquad (1)$$

Пусть при заданных значениях $(1, x_{1,0}, x_{2,0}, \ldots, x_{k,0}) = \vec{x}_0^T$ нужно знать неизвестное значение y_0 эндогенной переменной y. Величина y_0, по предположению, задана экзогенными переменными уравнениям модели (1):

$$y_0 = a_0 + a_1 \cdot x_{1,0} + a_2 \cdot x_{2,0} + \ldots a_k \cdot x_{k,0} + u_0 \qquad (2)$$

Обозначим $\tilde{y}_0$ прогноз значения y_0. Прогноз y_0 принято называть оптимальным, если оказываются справедливыми следующие два свойства этого прогноза:

$$\begin{cases} E\left(\tilde{y}_0 - y_0\right) = E\left(\triangle \tilde{y}_0\right) = 0, \\ E\left(\tilde{y}_0 - y_0\right)^2 = \sigma^2\left(\triangle \tilde{y}_0\right) \to \min \end{cases} \qquad (3)$$

Прокомментируем оба требования: первое требование означает равенство 0 математичкого ожидания истинной ошибки прогноза $\left(\tilde{y}_0 - y_0\right)$, несмящённая оценка прогноза, второе требование означает, что среднее растояние между прогнозом и истинной минимально. Подчеркнём, что эти два требования являются аналогом требованиям оптимальности к статистической процедуре оценивания моделей.
Справедлива следующая теорема:
Пусть модель (1) оценивается методом наименьших квадратов по уравнениям наблюдений, где справедливы все предпосылки теоремы Гаусса-Маркова. Тогда оптимальный прогноз рассчитывается по формуле (4):

$$A)\ \tilde{y}_0 = \tilde{a}_0 + \tilde{a}_1 \cdot x_{1,0} + \ldots + \tilde{a}_k \cdot x_{k,0} \qquad (4)$$

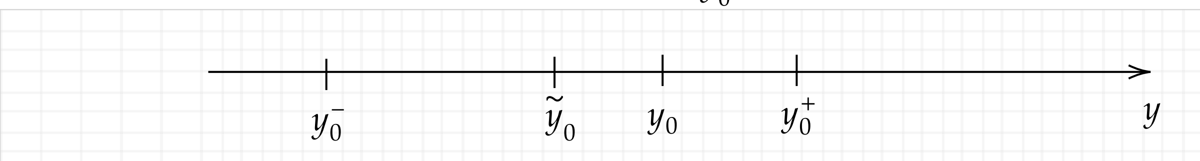
Характеристика точности прогноз:

$$\sigma^2\left(\triangle \tilde{y}_0\right) = \sigma_u^2 + \sigma_u^2 \cdot \vec{x}_0^T \cdot Q \cdot \vec{x}_0 = \sigma_u^2(1 + q_0)$$

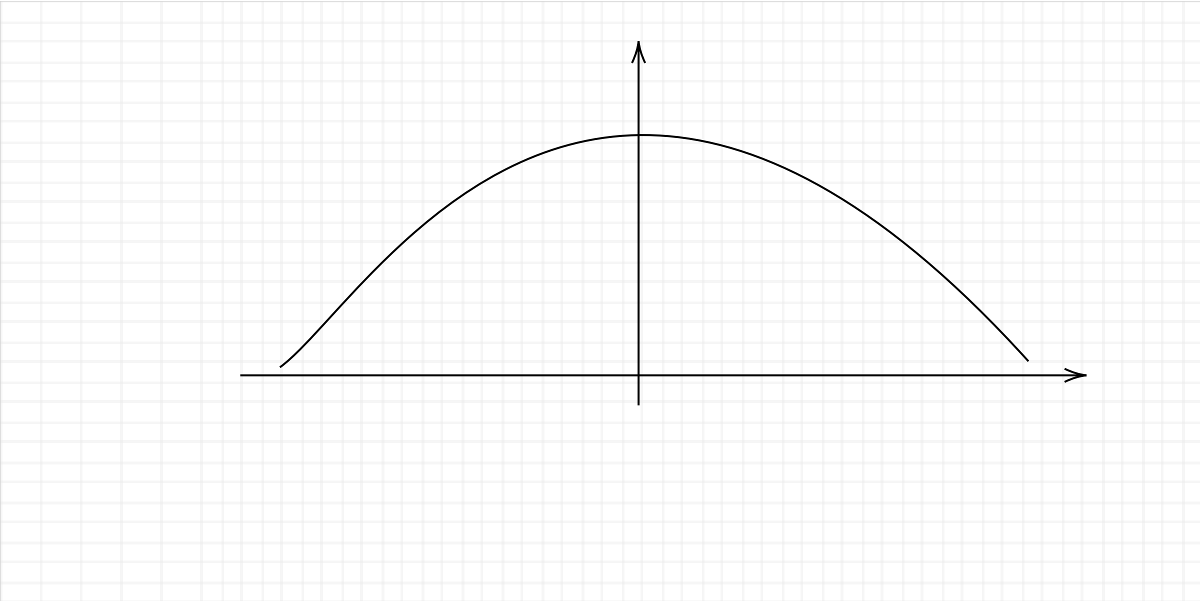
Замечание. Присутствие в характеристике точности прогноза слагаемого $\vec{x}_0$ обусловлен случайными ошибками в оценке коэффициентов модели. Первое слагаемое в характеристике обусловлено отсутвием случайного возмущения u_0.
Подводим итог, оптимальный прогноз по модели вычисляется в итоге подстановки в оценку функции регресси заданых значений случайных переменных.

Интервальное прогнозирование по модели и проверка её адекватности

В пукте выше мы обсудили оптимальный прогноз $\tilde{y}_0$.



По мимо точечного прогноза в финансово-экономической сфере прогноз искомой величины y_0 часто строится ввиде интервала с левой границей y_0^- и правой границей y_0^+. Этот интервал накрывает неизвестное значение y_0 с заданной доверительной вероятностью. В основании лежит следующая теорема:
Пусть в моделе (1) выполнены все предпосылки теоремы Гаусса-Маркова и случайные возмущения имеют нормальный закон распределения



Тогда следующая дробь:

$$t = \frac{\widetilde{y}_0 - y_0}{\widetilde{\sigma}\left(\triangle \widetilde{y}_0\right)} \sim t(m), \text{ где } m = m - k + 1$$

имеет закон распределения Стьюдента с количеством степеней свододы $n - k + 1$.
Из теоремы выше вытекает равенство (3):

$$P\left(\left|\frac{\widetilde{y}_0 - y_0}{S\widetilde{y}_0}\right| \leqslant t_{\text{крит}}\right) = 1 - \alpha \qquad (3)$$

То есть существует такое $t_{\text{крит}}$ при котором справедливо равенство (3). $t_{\text{крит}}$ имеет значение двухсторонней квантили с кол-ом степеней свободы $n - k + 1$. Освобождаясь от модуля мы перепишем формулу в следующем виде:

$$P\left(\widetilde{y}_0 - t_{\text{крит}} \cdot \widetilde{\sigma}\left(\triangle \widetilde{y}_0\right) \leqslant y_0 \leqslant \widetilde{y}_0 + t_{\text{крит}} \cdot \widetilde{\sigma}\left(\triangle \widetilde{y}_0\right)\right) = 1 - \alpha \qquad (3')$$

$\left[y_0^-,\ y_0^+\right]$ – доверительный интервал

Из которого и следуют формулы границ доверительно интервала.

Лекция №15

$t-$ тест значимости объясняющих перменных модели и построениен моделей с нелинейными по коэффициентам функциями регрессии

План

1. $t-$ тест значимости объясняющих перменных модели;
2. Построение эконометрических моделей с произвольными нелинейными по коэффициентам функциями регрессии;
3. Спецификация и трансформация к базовой модели эконометрики нелинейных моделей со стандартными функциями регресии;

На прошлой лекции обсудили ошибки спецификации моделей и последствия этих ошибок. Одна из ошибок спецификации состоит в пресутствие в моделе незначащих объясняющих переменных.

Обсудим тест именуемый $t-$ тестом позваляющий идентифицировать незначащие объясняемые переменные в оценённой модели. Запись оценённой модели:

$$\begin{cases} y_t = \underset{Sa_0}{\tilde{a}_0} + \underset{Sa_1}{\tilde{a}_1} x_1 + \ldots + \underset{Sa_k}{\tilde{a}_k} \cdot x_k + \underset{\tilde{\sigma}_u}{u_t}; \\ R^2 = \ \ldots \end{cases} \tag{1}$$

Вспомним определение незначащей объясняющей переменной, так например x_1, является незначащим, если справедлива гипотеза:

$$H_0 : a_1 = 0 \tag{2}$$

Если же она не справедлива, то есть справедлива альтернативная гипотеза:

$$H_1 : a_1 \neq 0$$

то переменная x_1 является значащей и её необходимо сохранить в моделе.

$t-$ тест позируется на следующей **теореме:**

Пусть выполнены все предпосылки теоремы Гаусса-Маркова, а случайные возмущения имеют нормальный закон распределения. Пусть модель оценена методом наименьших квадратов, то гогда следуюзщая дробь:

$$t = \frac{\tilde{a}_1}{S\tilde{a}_1}$$

является случайной переменной расапределённой по закону Стьюдента с количеством степеней свободы $m = (n - (k+1))$.

Порядок $t-$ теста о незначимости объясняющей переменной в оценённой модели

Шаг 1. Визуальный поиск в оценённой моделе таких объясняющих переменных в которых спраделиво следующее неравенство:

$$|\tilde{a}_j| \leqslant S\tilde{a}_j |t| \overset{2}{\leqslant} t_{\text{крит}} \tag{3}$$

Если находится такая объясняющая переменная x_j для которого справедливо данное неравенство, то это означает, что значение оценки коэффициента $\tilde{a}_j$ скорее всего вызвана ошибкой оценивания коэффициента $a_j = 0$.

Именно с таких объясняющих переменных нужно приступать к $t-$ тесту.
Шаг 2. Расчёт статистики t:

$$t = \frac{\tilde{a}_1}{S\tilde{a}_1}$$

Что гипотеза $H_0\ a_j = 0$.
Шаг 3. Задаться значением $\alpha \in [0, 0.05]$ и при количестве степеней свободы $m = (n - (k + 1))$ найти при помощи функции "СТЬЮДЕНТ.ОБР.2Х" найти двустороннюю квантиль уровня $1 - \alpha$ распределения Стьюдента. Пример, выбираем уровень значимости 0.05 с кол-ом степеней свободы $m = 11$, тогда упомянутая выше квантиль равна ≈ 2.2. Часто такую квантиль обозначают $t_{\text{крит}}$.
Шаг 4. Проверить справедливость следующего неравенства:

$$t| \overset{2}{\leqslant} t_{\text{крит}}$$

Если оно справедли, то гипоза $H_0 : a_j = 0$ может быть принята, как не противоречащая реальным данным и переменная x_j удалена из модели, в противном случае приниматся гипотеза H_1 переменная x_j интерпритируется, как значащая и сохраняется в моделе.
Замечание. Из курса математической статистики известно, что процедура проверк статистической гипотезы, может приводить к ошибкам I или II рода. Ошибка I рода - отвергнуть гипотезу H_0, когда она верна. Ошибка II рода - когда мы приняли гипотезу, когда она не верна. В ситуации $t-$ теста гораздо опаснее принять гипотезу H_0, когда она не верна и следовательно исключить значащую переменную из модели (лучше сохранить незначащую чем сохранить значащую). После удаления из модели незначащих переменных необходимо повторить все тесты в моделе.
Взаимосвязь двусторонней и обычной квантили распределения Стьюдента
Двусторонную квантиль уровня $1 - \alpha$ обозначаем символом $t_{\text{крит}}$; обычную квантиль уровня $1 - \frac{a}{2}$ обозначаем символом $t_{1-\frac{a}{2}}$ справедливо следующее равенство $t_{\text{крит}} = t_{1-\frac{a}{2}}$.

Построение эконометрических моделей с произвольными нелинейными по коэффициентам функциями регрессии

Начнём с примера. Приведём пример эконометрической модели с нелинейной по коэффициентам уровня регрессии.

$$\begin{cases} Y = A \cdot K^{\alpha} \cdot L^{\beta} + u \\ A > 0, 0 < \alpha < 1, 0 < \beta < 1 \\ E(u) = 0, E\left(u^2\right) = \sigma^2 \end{cases}$$

До настоящего времени, мы обсуждали только линейные по коэффициентам модели. В этой теме мы обсудим построение таких моделей. Начнём с самого общего случая:

$$\begin{cases} y = f(\vec{x};\vec{a}) + u \\ E(u) = 0, E\left(u^2\right) = \sigma^2 \\ \vec{p} = \left(\vec{a}, \sigma^2\right) \end{cases} \quad (4.3)$$

Символом $\vec{x}$ – набор объясняющих переменных (K, L), $\vec{a}$ – неизвестных коэффициентов (A, α, β).

Спецификация и трансформация к базовой модели эконометрики нелинейных моделей со стандартными функциями регресии

Обозначим символом $\vec{x}$ – обозначим какие-то известные приближённые значения коэффициентов и разложим функцию регрессии в ряд Тейлора в окресности точки $(\vec{a}_0)$:

$$f(\vec{x};\vec{a}) \approx f(\vec{x};\vec{a}_0) + f'_0 \cdot \triangle a_1 + \ldots + f'_k \cdot \triangle a_k \quad (4.4)$$

$$\triangle a_0 = a_0 - a_0^0;\ \ \triangle a_1 = a_1 - a_1^0;\ \ \ldots;\ \triangle a_k = a_k - a_k^0 \quad (4.5)$$

, где символами $f'_0,\ \ldots,\ f'_k$ обозначены частные производные.

Спецификация (4.3) примет вид (4.7), где приняты обозначения (4.6):

$$\triangle y = y - f\left(\vec{x};\vec{a}^0\right);\ z_0 = f'_0;\ z_1 = f'_1,\ \ldots, z_k = f'_k \quad (4.6)$$

$$\begin{cases} \triangle y = z_0 \cdot \triangle a_0 + z_1 \cdot \triangle a_1 + \ldots + z_k \cdot \triangle a_k + u \\ E(u) = 0, E\left(u^2\right) = \sigma^2 \\ \vec{p} = \left(\vec{a}, \sigma^2\right) \end{cases} \quad (4.7)$$

Лекция №16

Тестирование 0 и 1 предпосылок теормы Гаусса-Маркова

План

1. Построение модели со стандартными нелинейными функциями регрессии;
2. Применение метода дополнительной регрессии для тестирования нулевой предпосылки теоремы Гаусса-Маркова;
3. Тест Ремзи первой предпосылки теоремы Гаусса-Маркова - рисет тест;

На прошлой лекции мы обсудили:

$$\begin{cases} y = f(\vec{x};\vec{a}) + u \\ E(u) = 0, E\left(u^2\right) = \sigma^2 \\ \vec{p} = \left(\vec{a}, \sigma^2\right) \end{cases} \tag{4.3}$$

В эконометрике часто встречаются нелинейные по коэффициентам модели у которых функция регрессии имеют специальную структуру. Примером такой модели служит производственная модель с функцией Кобба-Дугласса. Заметим, что для такой производственной модели:

$$\begin{cases} Y = A \cdot K^{\alpha} \cdot L^{\beta} + u; \\ A > 0; 0 < \alpha; 0 < \beta; \\ E(u) = 0, E\left(u^2\right) = \sigma^2 \end{cases} \tag{4.8}$$

процедура логорифмирования функции Кобба-Дугласа позволяет трансформировать эту функцию к линейной функции:

$$\begin{cases} \underbrace{\ln Y}_{y} = \underbrace{\ln A}_{a_0} + \alpha \cdot \underbrace{\ln K}_{x_1} + \beta \cdot \underbrace{\ln L}_{x_2} + u; \\ E(u) = 0; Var(u) = \sigma_u^2; \\ A > 0; 0 < \alpha; 0 < \beta; \end{cases} \tag{4.9}$$

Функция Кобба-Дугласса является примером специальных функций, вот определение этого понятия: функцию $x_1, \dots, x_k$ называется специальной функцией, если она приводится к линейной функции, либо при помощи замены переменных, либо при помощи операции логорифмирования. $\boxed{\text{ДЗ}}$ Является ли специальной CES функция.

Спецификация модели на примере производственной модели

$$\begin{cases} Y = A \cdot K^{\alpha} \cdot L^{\beta} \cdot e^{u}; \\ A > 0; 0 < \alpha; 0 < \beta; \\ E(u) = 0; Var(u) = \sigma^2; \end{cases} \tag{4.11}$$

Вернёмся к моделе товаров и услуг функции Кобба-Дугласа и изменим её спецификацию.

После логорифмирования поведенческого уравнения данная спецификации преваращается в базовой модели эконометрики

$$\begin{cases} \ln Y = \ln A + \alpha \cdot \ln K + \beta \cdot \ln L + u \\ E(u) = 0; Var(u) = \sigma^2; \end{cases} \tag{4.12}$$

Рассмотрим несколько нюансов модели. Обратим внимание на смысл параметра σ в спецификациях (11) и (12). Величина σ (при гомоскедастичном возмущении σ – константа) имеет смысл меры относительного влияния неучённых факторов эндогенную переменную y. То есть σ – это безразмерная величина. Добавим, что в моделе (4.8) σ – это мера абсолютного влияния неучтённых факторов на переменную y (размерность σ совпадает с размерностью y). Добавим наконец, что после оценивания модели (4.12) проводят тестрирование всех предпосылое теоремы Гаусса-Маркова (смотри ниже) и проверку адекватности этой модели. От построенной модели (4.12) можно вернуться к построенной модели (4.11).

Итог: При построении моделей с нелинейными по коэффициентам функции регрессии в ситуации специальных функций регрессии, спецификация модели создаётся с учётом её трансформации к базовой модели эконометрики.

Тестирование нулевой предпосылки методом дополнительной регресси

Вернёмся к предпосылкам теоремы Гаусса-Маркова и сейчас объектом нашего внимания является нулевая предпосылка ($\vec{y} = X \cdot \vec{a} + \vec{u}$ Столбцы X линейно независимы). Сначала мы отметим критерий этой предпосылки справедлива следующая теорема: пусть нулевая предпосылка теоремы Гаусса-Маркова нарушена, то есть ранг матрицы X меньше числа столбцов этой матртицы. В этом и только в этом случае определитель корреляционной матрицы объясняющих переменных равен 0.

Спецификация ЛММР и нулевая предпосылка теоремы Гаусса-Маркова:

$$\begin{cases} y = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + \ldots + a_k \cdot x_k + u; \\ E(u) = 0; Var(u) = \sigma^2; \end{cases} \qquad (4.22)$$

$$rk(X) < (k+1) \Leftrightarrow |Cor(\vec{x}, \vec{x})| = 0$$

Символом $\vec{x}$ обозначен вектор объясняющих переменных $x_1, x_2, \ldots, x_k$ линейной модели множественной регрессии. Рассмотрим элемент этой матрицы который стоит в первой строчке во втором столбце, чтобы найти этот элемент нужно посчитать коэффициент корреляции по элементам второго и третьего столбцов матрицы X, используя (например) функцию КОРРЕЛ в Excel.

Замечание. Если нарушена 0-ая предпосылка теоремы Гаусса-Маркова, то говорят, что в линейной модели множественной регрессии присутствует совершенная мультиколлениарность.

Следующий метод может удалить линейно незвисимые случайные переменные. Метод состоит из следующих шагов:

Шаг 1. Обозначаем номером j номер объясняющей переменной в модели (22) и принимаем $j = 1$.

Шаг 2. Создаём спецификаю модели (дополнительной регресии) с объясняемой переменной x_j

$$\begin{cases} x_j = b_0 + b_1 \cdot x_1 + \ldots + b_{j-1} x_{j-1} + b_{j+1} x_{j+1} + b_k x_k + v_j \\ R_j^2 = \ ; \ j = 1, 2, \ldots, k \end{cases}$$

и оцениваем эту модель МНК.

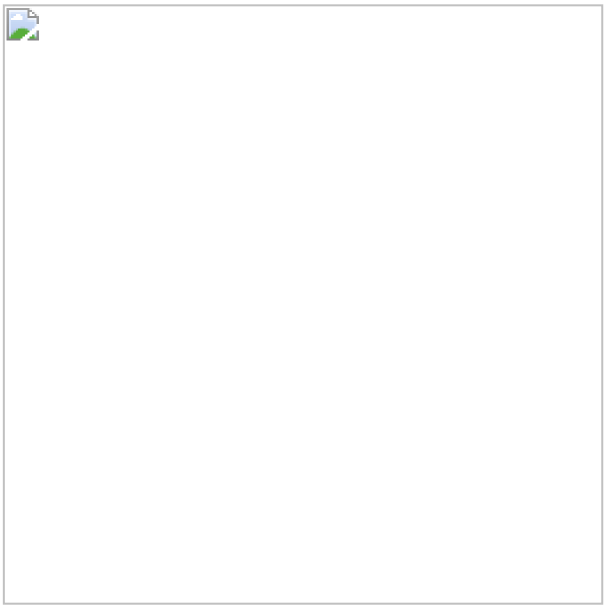
Шаг 2. $j + 1 = 2$ и вернуться к предыдущему шагу. Действовать до k.

При помощи метода дополнительной регрессии можно проверить справедливость нулевой предпосылки теоремы Гаусса-Маркова.

Тест Ремзи первой предпосылки теоремы Гаусса-Маркова - RESET тест

Нарушение предпосылки теоремы Гаусса-Маркова влечёт смещение оценок коэффициентов модели и обычной причиной нарушения этой предпосылки является пропуск значащих переменных модели или неправильный выбор функции регрессии. Мы обсудим тест первой предпосылки. Тест состоит из следующих шагов:

Шаг 1. Методом наименьших квадратов оценивается линейная модель множественной регрессии и отмечается



Лекция №17

Завершение теста Рамзи и дополнительные сюжеты

План

1. Завершение обсуждения теста Рэмзи первой предпосылки теоремы Гаусса-Маркова;
2. Тест Джарки-Бера нормального распределения случайного возмущения в модели;
3. Проблема несовершенной мультиколениарности в линейной модели и отбор объясняющих перменных методом дополнительной регрессии при помощи скорректированного коэффициента детерминации;
4. Вложенные эконометрические модели и тест Вальда ограничений во вложенной модели;

На прошлой лекции мы приступили к обсуждения теста первой предпосылки теоремы Гаусса-Маркова. Этот тест состоит из следующих шагов:

Шаг 1. Методом наименьших квадратов оценивается базовая модель эконометрики и вычисляются прогнозные значения:

$$\tilde{u}_i = y_i - \left(\tilde{a}_0 + \tilde{a}_1 \cdot x_{1,i} + \ldots + \tilde{a}_k \cdot x_{k,i}\right) \tag{3.5.8}$$

Прогнозные значения эндогенной переменной в модели; эти прогнозные значения $\left(\tilde{a}_0 + \tilde{a}_1 \cdot x_{1,i} + \ldots + \tilde{a}_k \cdot x_{k,i}\right)$ мы обозначим $\tilde{y}_i$.

Шаг 2. Задаётся целочисленное значение $(p = 2, 3)$ и оценивается вспомогательная модель (3.2):

$$\begin{cases} y_i = a_1 \cdot x_{1t} + a_2 \cdot x_{2t} + \ldots + a_k \cdot x_{kt} + b_1 \cdot \tilde{y}_t^2 + \ldots + b_p \cdot \tilde{y}_t^p + \varepsilon_t, \\ \text{где } \tilde{y}_i = \tilde{a}_1 \cdot x_{1t} + \tilde{a}_2 \cdot x_{2t} + \ldots + \tilde{a}_k \cdot x_{kt} \end{cases}$$

По правилу (3.1) вычисляется статистика критерия предпосылки 1 теоремы Гаусса-Маркова.

$$RESET = RAM\left(\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_n\right) =$$
$$= (ESS - ESS_{3.2}) : p - 1 / ESS_{3.2} : (n - (k + p - 1))$$

Если гипотеза H_0 справедлива и случайные возмущения в создоваемой модели нормально распределено, то статистика имеет F распределение Фишера с кол-ом степеней свободы $[(p-1),\ (n + (k + p - 1))]$. Если величина $RESET$ меньше чем квантиль распределения $1 - \alpha$, то принимается, если больше, то отвергается.

$$P_{RESET}(q|H_0) = P_{F_{(p-1),(n-(k+p-1))}}(q). \tag{3.3}$$

$$RESET_{\text{крит}} = F_{(p-1),(n-(k+p-1))}(1 - \alpha) \tag{3.4}$$

Тест Джарки-Бера нормального распределения случайного возмущения в модели

В этом тесте исследуется гипотеза о том, что в уравнениях наблюдений объекта случайные возмущения имеют нормальный закон распределения с 0 математическим ожиданием с одной и тойже дисперсией и являются независимыми

случайными переменными.
Тест состоит из следующих шагов:
Шаг 1. Модель оценивается методом наименьших квадратов и вычисляются оценки случайных возмущений (остатки).
Шаг 2. По правилу (2.1) расчитывается статистика критерия гипотезы H_0

$$JB = JB\left(\widetilde{u}_1, \widetilde{u}_2, \dots, \widetilde{u}_n\right) = n \cdot \left\{\frac{sk^2}{6} + \frac{(kur-3)^2}{24}\right\}; \quad (2.1)$$

$$sk = m_3 / m_2^{\frac{3}{2}},\ kur = \frac{m_4}{(m_2)^2}, m_k = \frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{(u_i)}\right)^k$$

m_k - оценка начального момента случайного остатка. Где sk обозначена оценка скошенности (ассиметрии). Для нормального закона скошенность равна 0. Символом kur обозначена оценка островершинности распределения (оценка куртозиса или эксцесса). Для нормального распределения истинное значение $kur = 3$. При справедливой гипотезе H_0 величина JB имеет распределение χ^2 с двумя степенями свободы (2.3):

$$P_{JB}(q|H_0) = P_{\chi_2^2}(q)$$

$$P(JB < JB_{\text{крит}}|\text{в}H_0) = \int_0^{JB_{\text{крит}} = \chi_2^2(1-\alpha)} p_{\chi_2^2}(q)dq = 1 - \alpha$$

Гипотеза H_0, когда велина $JB < JB_{\text{крит}}$, то принимается, а если больше то отвергается

Проблема несовершенной мультиколениарности в линейной модели и отбор объясняющих перменных методом дополнительной регрессии при помощи скорректированного коэффициента детерминации

Отметили критерий мультиколениарности:

$$rk(X) < (k+1) \Leftrightarrow |Cor(\vec{x}, \vec{x})| = 0$$

На практике часто встречаются ситуации, когда упомянутый выше определитель не равен 0, но очень мал. Тогда говорят, что в исходной модели присутствует несовершенная мультиколениарность. Отметим симптомы несовершенной мультиколениарности.
Симптомы:
Симптом №1. F – тест отвергает гипотезу:

$$H_0: a_1 = a_2 = \dots = a_k = 0,$$

но t – тест **не отвергает** гипотезу $H_0: a_j = 0$ о незначимости многих x_j.
Симптом № 2. При добавлении в обучающую выборку одного уравнения (или удаления одного из уравнений), оценки коэффициентов модели резко меняют свои значения, то есть оценки неустойчивые $\widetilde{a}_j$ и значит они не надёжные.
При несовершенной мультиколлениарности часто используется процедура пошаговой регресии отбора в модель объясняющих переменных. Процедура состоит из следующих шагов:

Шаг № 1. Модель оценивается по всем объясняющим переменным и отмечается скорректированный коэффициент детерминации (обозначим символом $\overline{R}_0^2$).

Шаг № 2. Из исходной оценённой модели выбирается коэффициент $\widetilde{a}_j$ с максимальным значением $|t_j| = \max x_j$. Удаляется из модели, переоценивается и запоминается значение скорректированной детерминации. Если $\overline{R}_0^2 \geqslant \overline{R}_1^2$, то удаление x_j нецелесообразно.

Шаг №3. Шаг №1 повторяется пока условие $\overline{R}_j^2 \geqslant \overline{R}_{j+1}^2$ перестанет повторяться

Лекция №18

Понятие вложенных моделей и тест Вальда вложенностей моделей

План

1. Понятие вложенной модели. Тест Вальда;

В процессе проверки адекватности модели нередко возникает задача по удалению из модели незначащих переменных. Удаление из модели таких переменных (незначащих) модет быть осуществленно с помощь либо t-теста (лекция от 9 декабря), либо при помощи теста Вальда. И в том и в другом случае в основании этих тестов лежит понятие вложенной модели. Вот определение этого понятия: модель М1 называется вложенной в модель М2:

$$M1 \quad y = a_0 + a_1 x_1 + u \tag{5}$$

$$M2 \quad y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + v \tag{6}$$

если для модели М2 справедлива гипотеза:

$$H_0 : a_2 = 0; a_3 = 0; \tag{7}$$

Модель М1 иногда также называется *моделью М2 в которой справедливо ограничение (7).*

Модель М2 следует предпочесть, если справедлива гипотеза 8:

$$H_1 : a_2 \neq 0 \cup a_3 \neq 0.(\text{хотя бы одна является значащей})$$

Гипотеза H_1 является отрицанием гипотезы H_0. Тест Вальда исследует гипотезу (7) против альтернативы (8). Он состоит из следующих шагов:

Шаг 1. МНК оценивается модель (5) и отмечается значение суммы квадратов оценок случайных возмущений.

Шаг 2. Оценивается модель (6) МНК и отмечается ESS_2.

И шаг 1 и шаг 2 осуществляется по одной и той же выборке.

Шаг 3. По правилу (9) вычисляется статистика критерия гипотезы H_0:

$$W = (ESS_1 - ESS_2) / (ESS_2 / (n - k_2)) \tag{9}$$

$$W \sim \chi_r^2 \text{ при } n \to \infty; r - \text{число}$$

Пусть справеливы все предпосылки теоремы Гаусса-Маркова при произовальном законе распределения случайных возмущений. Если справедлива гипотеза H_0, то статистика асимптотически (с ростом объёма выборки) имеет распределение хи-квадрат с числом степеней свободы r, где r число ограничений (7) в модели (6).

Шаг 4. Гипотеза H_0 отвергается в пользу гипотезы H_1 (т.е принимается модель (6)), если значение статистики превосходит квантиль распределения χ_r^2 уровня $1-\alpha$:

$$W > \chi_r^2(1-\alpha)$$

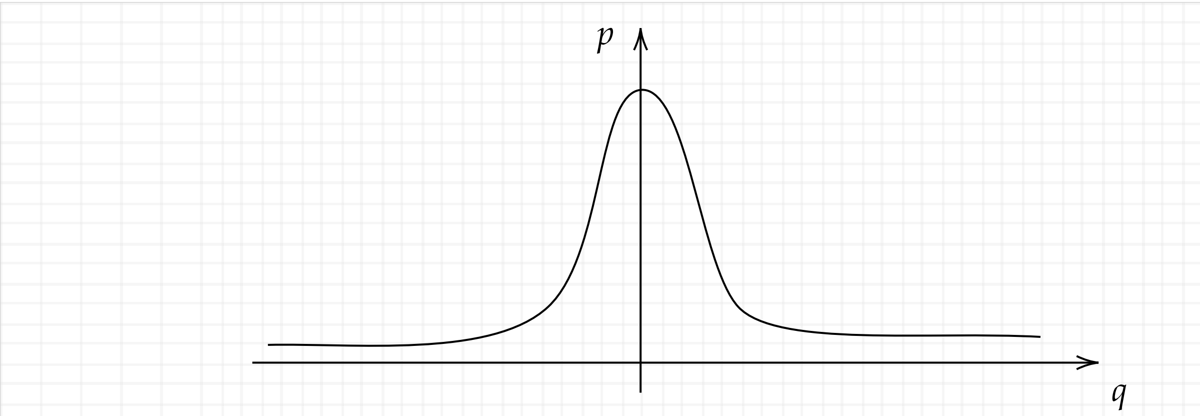
Следствие. Если случайное возмущение в модели М2 имеет нормальный закон распределения, то:

$$\frac{W}{r} \sim F_{r,n-k_2} \tag{10}$$

В такой ситуации гипотеза H_0 отвергается, если справедливо следующее неравенство:

$$\frac{W}{r} > F_{r,n-k_2}(1-\alpha)$$

Квантиль:



Итог: тест Вальда является расшириным вариантом $t-\text{теста}$, причём последний $t-\text{тест}$ совпадает с тестом Вальда при $r = 1$.

Схема построения эконометрической модели состоит из 4 этапов главным из которых является этап спецификации модели, то есть этап записи математическим языком взаимосвязей известных характеристик объекта и искомых характеристик объекта. В итоге такой записи появляется структурная форма модели. После использования всех принципов спецификации модели, структурная форма модели имеет следующий вид:

$$F(\vec{y}_t, \vec{x}_t) = \vec{u}_t;$$

Самым важным на практике случаем является линейные эконометрические модели.

Базовая модель экометрики имеет спецификацию

$$\begin{cases} y = a_0 \ + a_1 \cdot x_1 + a_2 \cdot x_2 + \dots a_k \cdot x_k + u_t \\ E(u_t) = 0; E\left(u_t^2\right) \ = \sigma_u^2 \end{cases}$$

и называется линейной моделью множественной регрессии.

МНК является оптимальной статистической процедурой оценивания параметров базовой модели экономктрики.

$A) \ \tilde{\vec{a}} \ = \ \left(X^T \ \cdot \ X\right)^{-1} \ \cdot X^T \ \cdot \ \vec{y} \ = \ Q \ \cdot X^T \ \cdot \vec{y};$

$B) \ \tilde{\sigma}_u^2 \ = \ \dfrac{\sum\limits_{i=1}^{n} \tilde{u}_i^2}{n \ - \ (k \ +1)},$ наилучшая оценка с минимальной дисперсией, где символом $\tilde{u}_i$.

Все предпосылки теоремы Гаусса-Маркова исследуются следующими тестами: 0 предпосылка методом дополнительной регрессии, 1 предпосылка - тестом Ремзи, 2 предпосылка - тест Голдфелда-Кванта, 3 предпосыка - тест Дарбина-Уотсона.

Если функция регрессии (точнее модель функции регрессии) оказывается нелинейной по коэффициентам

$$\begin{cases} Y = A \cdot K^{\alpha} \cdot L^{\beta} \cdot e^{u}; \\ A > 0; 0 < \alpha; 0 < \beta; \\ E(u) = 0; Var(u) = \sigma_u^2; \end{cases}$$

и после логорифмирования она трансформируется к базовой модели эконометрики.

$$\begin{cases} \underbrace{\ln Y}_{y} = \underbrace{\ln A}_{a_0} + \alpha \cdot \underbrace{\ln K}_{x_1} + \beta \cdot \underbrace{\ln L}_{x_2} + u; \\ E(u) = 0; Var(u) = \sigma_u^2; \end{cases}$$

Подчеркнём, что в экзамеционные билеты входят задачи со следующей структурой:

1. Составить спецификаю модели;
2. Оценить параметры модели по обучающей выборки;
3. Выполнить тот или иной тест с оценённой моделью (например, тест Дарбина-Уотсона);

Семинар №1
Два принципа спецификации эконометрических моделей и две их формы

План

1. Первый принцип спецификации экономической модели и ее структурная форма.

2. Второй принцип спецификации эконометрических моделей и приведенная форма моделей. Расчётная схема задач.

У любого изучаемого экономического объекта имеются известные количественные характеристики и искомые количественные характеристики. Математическая модель объекта — запись математическом языком взаимосвязей известных и искомых характеристик объектов. Модель нужна для искомых характеристик и .. Для расчётных. В процессе записи математическом языком взаимосвязей известных и искомых характеристик стараются привлекать, прежде всего, линейные алгебраические функции. Известные характеристики объекта называют экзогенными переменными, искомые — эндогенными.

Задача 1. Изучаемым объектом называется конкурентный рынок некоторого блага. Искомыми характеристиками данного объекта являются:

- Уровень спроса данного блага (demand) - y^d

- Уровень предложения блага (supply) - y^s

- Цена блага (price) - p

  В обсуждаемой задаче величины являются эндогенными. Известной характеристико в данной задаче будет служить душевой доход потребителя. Обозначим символом . Между спросом и предложением существует имеются объективно существующие взаимосвязи, которые можно сформулировать следующим образом:

  1. Уровень спроса объясняется его ценой и душевым доходом. С ростом цены спрос снижается для нормальных товаров. С ростом дохода потребителя спрос возрастает. Такое благо называется ценным.

  2. Закон предложения. Уровень предложения блага объясняется его ценой и с ростом цены предложение увеличивается.

  3. Рыночная цена блага формируется при балансе спроса и предложения.

Теперь необходимо математическом языком записать данные утверждения. Таким способом мы получим структурную форму простейшей модели спроса-предложения. (1) означает что переменная y^d - функция переменных P и X. $y^d = y^d(P, X)$

Воспользовавшись первым принципом спецификации мы выберем подходящую линейную функцию аргументов P и X.

$$\begin{cases} y^d = a_0 + a_1 p + a_2 x; \\ a_1 < 0; \\ a_2 > 0; \end{cases} \tag{1}$$

(3) - простейшая функция спроса $y^s = y^s(p)$

$$\begin{cases} y^s = a_3 + a_4 p; \\ a_2 > 0; \end{cases} \tag{2}$$

Структурная форма простейшей модели спроса-предложения нормального ценного блага на конкурентном рынке. В структурной форме данной модели искомые величины и известная величина тесно переплетены. Для расчета по модели ее нужно трансформировать к такому виду, где каждая искомая величина будет выражена только через известную величину .

$$\begin{cases} y^d = y^d(x); \\ y^s = y^s(x) \\ P = P(x); \end{cases} \tag{3}$$

Второй принцип спецификации эконометрических моделей и приведенная форма модели. Принцип требует чтобы количество уравнений совпадало с количеством искомых переменных. Является необходимым условием для представления каждой искомой величины. Это служит и является необходимым условием для представления искомой величины в виде явной функции известных характеристик (экзогенных переменных)

$\boxed{\text{Д/з}}$. Проверить, что наши шаги корректные (нет деления на ноль). Найти экономическое представление.

Третий принцип спецификации экономических моделей: отражение в модели, факторы времени(Семинар №2)

План

1. Спецификация динамической модели спроса-предложения на конкурентом рынке. Типы переменных в динамических моделях.

2. Трансформация динамической модели к приведенной форме. Предельные величины в экономике.

3. ДЗ, защита ДЗ

Мы обсудили два принципа спецификации эконометрических моделей и две формы; обсуждение провели на примере простейшей модели спроса-предложения на конкурентом рынке.

3 эндогенные переменные: спрос, предложение и цена; 1 эндогенная: x.
На Семинаре 1 было обсуждено 2 принципа спецификации модели и 2 формы $(y^d, y^s, p), x$.

В этой модели (о взаимосвязи эндогенных и экзогенных переменных) по существу заложено предположение, что эндогенные переменные реагируют на уровень душевого дохода (x) и уровень предложения мгновенно реагирует на цену блага (p).

Между тем уровень предложения блага (y^d) в текущем периоде обладает определённой инерцией по отношению к изменению цены блага (p). Точнее уровень предложения в текущем периоде лучше объясняется ценой блага в предшествую периоде (p_{t-1}), так как производителю необходимо время для перестройки производства. Подчеркнём что в этом утверждении содержится фактор времени и мы обязаны в процессе записи математическим языком данного утверждения различать цену блага в текущем периоде (p_t) и в предшествующем (p_{t-1}).

Обозначим цену блага текущем периоде p_t, обозначим цену блага в предшествующем p_{t-1} (лагавой ценой). Таким образом мы можем сформулировать закон: Уровень предложения растёт ($y_t^s = y_t^s(p_{t-1}) \uparrow$) с ростом лаговой цены $p_{t-1} \uparrow$. Напротив уровень блага <u>мгновенно</u> реагирует на изменение цены (p_t) и душевного дохода x: $y_t^d = y_t^d(p_t, x) \downarrow\uparrow$ (в таком случае благо нормальное и ценное).

Третий закон формирования рыночной цены в текущем периоде сохраняется и в данном случае:p_t *(цена в текущем периоде) формируется при балансе текущего спроса и текущего предложения*. Требутся составить модель которая позволяет объяснять уровень спроса (y_t^s), уровень предложения (y_t^d) душевым доход в текущем периоде (p_t).

$$\begin{cases} p_t, p_{t-1} \\ y_t^s = y_t^s(p_{t-1}) \uparrow \\ y_t^d = y_t^d(p_t, x) \downarrow\uparrow \end{cases}$$

Таким образом в данной задаче с уточнённым законом предложения будут присутствовать

- Текущей эндогенной перменной

$$(y_t^s, y_t^d, p_t) \tag{1}$$

- Текущая экзогенная и лагавая эндогенная переменная

$$(x_t, p_{t-1}) \tag{2}$$

$$\begin{cases} \begin{cases} y_t^d = a_0 + a_1 p_t + a_2 x_t - \text{(прос. лин. модель спроса)}, \\ a_1 < 0, a_2 > 0 \end{cases} \\ \begin{cases} y_t^s = b_0 + b_1 p_{t-1} - \text{(прос. лин. модель предложения)}, \\ b_0 > 0 \end{cases} \\ y_t^s = y_t^d \end{cases} \tag{3}$$

Три уравнения образуют структурную форму простейшей экономической модели нормально ценного блага на конкурентном рынке.

Итог. Для отражения в модели фактора времени все переменные модели датируются, т.е. привязываются ко времени и в итоге возникает спецификация динамической модели. Подчеркнём, что в набор объясняющих переменных (2) могут входить лаговые эндогенные переменные.

$\boxed{\text{Д/з}}$ p_t^m или p_{t-1}^m

Задача. Трасформировать модель (3) к приведённой

$$a_0 + a_1 p_t + a_2 x_1 = b_0 + b_1 p_{t-1}$$

$$a_1 p_t = b_0 + b_1 p_{t-1} - a_2 x_t - a_0$$

$$p_t = \frac{b_0 + b_1 p_{t-1} - a_2 x_t - a_0}{a_1}$$

$$p_t = \frac{b_0 - a_0}{a_1} + \frac{b_1}{a_1} p_{t-1} - \frac{a_2}{a_1} x_t$$

Приведённая форма предложения уже содержится в структурной форме (3):

$$y_t^s = b_0 + b_1 p_{t-1}$$

В силу $y_t^s = y_t^d$, $y_t^s = b_0 + b_1 p_{t-1} = y_t^d$

Получаем систему уравнений:

$$\begin{cases} y_t^s = b_0 + b_1 p_{t-1} = y_t^d; & (4) \\ p_t = \dfrac{b_0 - a_0}{a_1} + \dfrac{b_1}{a_1} p_{t-1} - \dfrac{a_2}{a_1} x_t; & (5) \end{cases}$$

Это система называется простейшей динамической моделью спроса и предложения.

Сопоставляя приведённые формы статической модели спроса и предложения (Семинар 1) и динамической модели мы видим, что это совершенно различные модели.

Предельные величины в экономике

Вернёмся к приведённой форме (5) и обозначим: $\alpha_0 = \dfrac{b_0 - a_0}{a_1}, \alpha_1 = \dfrac{b_1}{a_1}, \alpha_2 = \dfrac{a_2}{a_1}$ получим:

$$p = \alpha_0 + \alpha_1 p_{t-1} - \alpha_2 x_t \tag{6}$$

Наша цель выяснить экономический смысл α_1, α_2. Предположим, что $p_{t-1}, x_t + \triangle x_t$, тогда:

$$p_t + \triangle p_t = \alpha_0 + \alpha_1 p_{t-1} + \alpha_2 (x_t + \triangle x_t) \tag{7}$$

Вычитая из уравнения (7) - (6) получим:

$$\triangle p_t = \alpha_2 \cdot \triangle x_t \tag{8}$$

Предположим, что $\triangle x_t = 1$, тогда $\triangle p_t = \alpha_2$

Таким образом α_2 изменение эндогенной переменной p_t в ответ на дополнительную еденицу, объясняющую x_t.

α_2 – предельным значением p_t по объясняющей переменной x_t.

Добавим, что α_2 можно расчитать по правилу:

$$\alpha_2 = \frac{\partial p_t}{\partial x_t} \tag{9}$$

Задача. Вычислить α_2 и дать экономическую интерпретацию.

Рассматривая знаки коэффициентов в структурной форме (3) и выражение коэффициента (5), мы убеждаемся, что $\alpha_2 > 0$.

$\boxed{\text{Д/з}}$ Уточнить динамический закон предложения, согласно уточнённому закону $y_t^s = y_t^s(p_{t-1}) \uparrow \Rightarrow y_t^s = y_t^s(p_{t-1}, p_{t-1}^m) \uparrow\downarrow$. Лагаваю цену сырья ($p_{t-1}^m$) интерпретировать, как лаговую экзогенную переменную. Трансформировать такую динамическую модель к приведённой форме. И выяснисть знак у текущего спроса по лаговой цене сырья $\dfrac{\partial y_t^d}{\partial p_{t-1}^{(m)}}$

Семинар №3
Отражение спецификации эконометрической модели влияния на текущие эндогенные переменные неучтенных в модели факторов

1. Аналитическое описание диаграммы рассеянья (Лаговый доход – текущее потребление домохозяйств). Четвёртый принцип спецификации эконометрических моделей;

2. Обсуждение домашних заданий;

На двух предшествующих занятиях обсудили три принципа спецификации эконометрических моделей:

1. Словестный (вербальный) модель возникает в результате записи мат. языком взаимосвязей

2. Количество уравнений моделей равно количеству текущих эндогенных переменных

3. Отражения фактрора време, датирование всех переменных моделей

На сегодняшнем занятии мы рассмотрим отражение её не полного соответсветсвия изучаемому объекту, т.е. отражение в спецификации модели влияние на текущие эндогенные переменные неучтённых факторов. В качестве изучаемого объекта мы выберем экономику россии, изучать этот объект мы будем превлекая модель Самуэльсона-Хикса делового цикла экономики. Состояние экономики в текущем периоде t описывается основными макроэкономическими переменными: Y_t – уровень ВВП страны C_t – уровень расхода домохозяйств на потребление, I_t – объём инвестиций в экономику (накопление капитала), G_t – уровень государственных расходов.

Сейчас требуется составить спецификацию модели позволяющей объяснять переменные $Y_t,\ C_t,\ I_t,\ G_t$ их лаговыми значениями. При составлении спецификации нужно учесть следующие экономические утверждения (стр. 38 задача 2.5):

1. Текущий расходы домохозяйств на конечное потребление C_t объясняются уровнем ВВП в предыдущем периоде, возрастая вместе с ним;

2. Величина инвестиций в текущем периоде (валовое накопление капитала) I_t объясняется приростом ВВП за предыдущий период (прирост ВВП за предшествующий период - это разность $\triangle Y_{t-1} = Y_{t-1} - Y_{t-2}$); с ростом $\triangle Y_{t-1}$ увеличивается и I_t;

3. Государственные расходы G_t возрастают с постоянным темпом роста;

4. Текущее значение ВВП Y_t есть сумма текущих уровней потребления домохозяйств, инвестиций и государственных расходов (основное тождество системы национальны счетов)

$$\begin{cases} C_t \;=\; a_0 \;+\; a_1 \;\cdot Y_{t-1}; & a_1 \;>\; 0 \\ I_t \;=\; b\,(Y_{t-1} \;-\; Y_{t-2}); & b \;> 0 \\ G_t \;=\; g\,G_{t-1}; & g > 1 \\ Y_t \;=\; C_t \;+\; I_t \;+\; G_t; & \end{cases}$$

$$\text{Объсняющие переменные : } \overrightarrow{x_t} = (1, Y_{t-1}, Y_{t-2}, G_{t-1})$$
$$\text{Объясняемые переменные : } \overrightarrow{y_t} \;=\; (Y_t,\ C_t,\ I_t,\ G_t)$$

Зададимся следующим вопросом. Согласуется ли данная модель с изучаемым объектом (экономикой россии)? Другими словами соответсвует ли эта модель реальной статистической информации из реальных счетов россии? Чтобы ответить на этот вопрос проведём следующее иследование:

1. Выберем в моделе Самуэльсона-Хикса уравнение текущих расходов домохозяйств $C_t \;=\; a_0 \;+\; a_1 \;\cdot Y_t$ и проведем соответсвие этой модели статистической информации из системы национальных счетов России; Это информация содержится в файле (на почте) элементы использования ВВП России; Эту информации можно собрать на сайте www.gks.ru/;

2. В Excel построим график, откладывая вдоль горизонтальной оси лаговые уровни ВВП, в доль вертикальной оси соответсвующие уровни домохозяйств C_t. Если модель Самуэльсона-Хикса в полной мере соответсвует реальным данным, то построенный график (построенное множество точек) будет представлять собой восходящую прямую. Если же график окажется иным, то будет означать, что по крайней мере первое уравнение не соответсвует реально статистике.

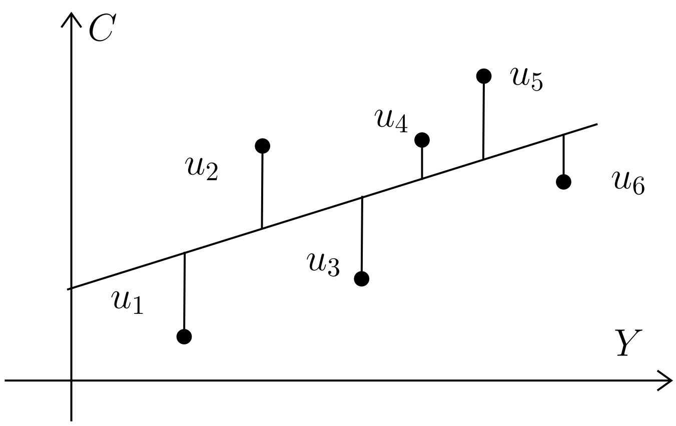


Рис. 1: Уравнение потребления

3. Рассматривая данный график мы констатируем, что точки не распределились на воходящей прямой и это значит, что модель Самуэльсона-Хикса не в полной мере соответсвует изчуемому объекту; Основная причина неполного соответствие влияние на уровень потребления домохозяйств неучтённых факторов.

$\boxed{\text{ДЗ}}$ осуществить аналогичное исследование модели Самуэльсона-Хикса для уравнения инвестиций и государственных расходов.

Аналитическое описание диаграммы рассеивания. Построенный график принято называть диаграммой рассеивания. Нам предстоит аналитически описать эту диаграмму. Рассматривая эту диаграмму мы можем сделать следующие выводы:

1. Точки реальных данных разметились вдоль восходящей прямой;

2. Точки реальных данных хаотично разместились вдоль восходящей прямой, то выше, то ниже, то ближе, то дальше.

Вот аналитическая запись данной диаграммы:

$$\begin{cases} C_t \ = a_0 \ + a_1 Y_{t-1} \ + \ U_t; \\ E(U_t) \ = \ 0; Var(U_t) = \sigma_u^2 \end{cases} \tag{1}$$

Символом U_t мы обозначили переменную величину, которая хаотично принимает то положительные, то отрицательные значения рассеяные вокруг 0. Символов $E(U_t)$ мы обозначили среднее значение, символом $Var(U_t)$ диспресию. В силу хаотичности появления её значений экономисты называют *случайным возмущением*. Физики и в технических приложениях такие величины называются невязками или ошибками модели.

Итог: Выражение (1) более адекватно описывает лаговое потребление в стране и текущего потреблени домохозяйств.

Общий вывод: для отражения на текущие эндогенные переменные неучтённых факторов в правых частях поведенческих моделей дискриптивных моделей включаются случайные возмущения. В этом и состоит четвёртый принцип спецификации экнометрических моделей.

$\boxed{\text{ДЗ}}$ построить графики в Excel.

Схема построения эконометрической модели

План

1. Этапы схемы построения эконометрической модели
2. Построение модели государственных расходов РФ
3. $\boxed{\text{ДЗ}}$

На прошлых занятиях мы обсудили приципы спецификации эконометрических моделей.

Сегодня обсудим схему построения эконометрический моделей. Оно состоит из четырёх **шагов**:

Шаг 1. Спецификация модели. В частности, фрагмент модели сельско-хозяйственных государственных расходов имеет следующую спецификацию:

$$\begin{cases} G_t = g\,G_{t-1} + w_t & g > 1 \\ E(w_t) = 0;\ Var(w_t) = \sigma_w^2 \end{cases} \tag{1}$$

Спецификация эконометрической модели обязательно содержит неизвестные константы. Они называются *параметрами модели.* В (1) параметры модели $g,\ \sigma_w$, где g – темп роста государственных расходов, σ_w – среднеквадратичное отклонение случайного возмущения или мера влияния неучтённых факторов.

Шаг 2. Сбор и проверка статистической информации в конкретных значений переменных, входящих в модель. (Примером такой информации служит файл "Элементы использования ВВП")

Собранную статистичестическую информацию разделяют на две части: большую $\approx 80\%$ часть имеет *обучающая* вырока и используется для определения параметров модели. Остальную часть отправляют на проверку инфляции и именуют *тестовой или контрольрующей выборкой.*

Примем обучающейся информации C (2002-2017 годов). Данные за 2018 год отнесём к контролирующей выборке.

Шаг 3. Оценивание по обучающей выборке неизвестных параметров модели методами математической статистики. На этом этапе по обучающейся выборке вычислим оценку $(\widetilde{g}, \widetilde{\sigma}_w)$ (3). Оценки (3) вычислим методом наименьших квадратов.

Шаг 4. Оценённая модель проходит проверку адекватности:

$$\left\{ G_t = \widetilde{g}\,(S_{\widetilde{g}})G_{t-1} + w_t\left(\widetilde{\sigma}_w\right) \right. \tag{4}$$

$$\widetilde{G}_{t(2018)} = \widetilde{g} \cdot G_{t-1(2017)}$$

$$\delta = |\widetilde{G}_t - G_t|\ (G_t) \cdot 100 \le 15\% \tag{5}$$

Модель признаётся адекватной, если отностительная ошибка прогноза не превышает 15%.

Оценивание пар-ов $g,\ \sigma_w$ модели государственных расходов РФ. МНК

1. Сотставление системы уравнений наблюдения. Подставим значения из обучающейся выборки в уравнения данной модели:

$$\begin{cases} G_{2003} \ = \ g\, G_{2002} \ + \ w_{2003} \\ G_{2004} \ = \ g\, G_{2003} \ + \ w_{2004} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ G_t \ = \ g\, G_{t-1} \ + \ w_t \end{cases}$$

$$\begin{cases} 6540.2 \ = \ g\, 6390 \ + \ w_{2003} \\ 6679 \ = \ g\, 6540.2 \ + \ w_{2004} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ 7264.3 \ = \ g\, 7238,3 \ + \ w_{2017} \end{cases}$$

В системе уравнений наблюдений неизвестными являются консанта g, случайные возмущения и их среднекрадратичное отклонение.

2. В итоге решения следующей задачи на безусловный экстремум:

$$\begin{cases} ESS\left(\tilde{g}\right) \ = \ \sum\limits_{t=2003}^{t=T} \left(G_t \ - \ \tilde{g} \ \cdot G_{t-1}\right)^2 \ \to \min \\ \tilde{w}_t \ = \ G_t \ - \ \tilde{g} \ \cdot G_{t-1} \end{cases} \qquad (2)$$

Вычисляется сначала оценка $\tilde{g}$ параметра g, а затем вычисляются оценки случайных возмущений в уравнениях наблюдений.

ДЗ №1 Доказать (по правилами дифференцирования сложной функции), что величина $\tilde{g}$, обеспечивающая эксремум $ESS\left(\tilde{g}\right)$ может быть вычислена в процессе решения линейного алгербраического уравнения $R \ \cdot \tilde{g} = S$, где $R \ = \ \sum\limits_{t=2003}^{T} G_{t-1}^2$,

$$S \ = \ \sum_{t=2003}^{T} G_{t-1} \cdot G_t.$$

Рассчитаем R и S, а найдём $\tilde{g}$:

$$R \ = \ 757881587,6$$
$$S = 763691322,5$$
$$\tilde{g} = \frac{S}{R} = R^{-1} \cdot S = 1,00767$$

$\tilde{g}$ – оценка темпов роста государственных расходов России.

Величина $\tilde{g} \ = 1,00767$ говорит о том, что государсвенные расходы в среднем возрастают на $1,008$.

Рассчитаем оценки случайных возмущений:

$$\tilde{w}_t \ = \ G_t \ - \ \tilde{g} \ \cdot G_{t-1}$$

Мы видим, что эти оценки принимают то большие, то меньшие оценки разных знаков.

$\boxed{\text{ДЗ №2}}$ Завершить обсуждение **2 этапа** рассчётом: $\widetilde{\sigma}_w = \sqrt{\dfrac{\sum\left(\widetilde{w}_t^2\right)}{n(=15)-1}}$

Семинар №5
Линейная модель множественной регрессии и оценивание её параметров при помощи функции line

1. Проверка адекватности модели Самуэльсона-Хигса для гос. Расходов России

2. Модифицированная модель Самуэльсона-Хигса расходов домохозяйств в России и её оценивание при помощи функции line excel

3. Проверка ДЗ

На занятии вычислили значения случайных возмущений по правилу:

$$\vec{w}_t \;=\; G_t \;-\tilde{g}\, G_{t-1} \tag{4}$$

$$\tilde{\sigma}_w \;=\; \sqrt{\frac{\sum \tilde{w}_t^2}{n\;-\;k}} \approx 140 \tag{5}$$

Таким образом **третий этап** схемы завершается записью оценнённой модели:

$$\begin{cases} G_t \;=\; 1.008 \cdot \tilde{g} \cdot G_{t-1} \;+\; w_t \\ \tilde{\sigma}_w = 140 \end{cases} \tag{6}$$

На четвёртом этапе осуществляется прогноз по оценённой модели значений эндогенных переменных из контролирующей выборки. В нашем примере рассчитывается расход на 2018 год. После рассчёта прогноза вычисляется относительная ошибка прогноза.

$$\tilde{G}_{2018} = \tilde{g}\;(=\;1.008) \cdot G_{2017} \tag{7}$$

Модель является адекватной, если относительная ошибка прогноза не превосходит 15% от прогнозируемых велечин.

$$\delta \;=\; 100 \;\cdot\; |\tilde{G}_{2018} \;-\; G_{2018}| \;\leq\; 15\% \tag{8}$$

Задача.

Вычислить по правилу (7) прогноз и проверить адекватность модели:

$$\tilde{G}_{2018} \;=\; \tilde{g} \;\cdot G_{2017} \;=\; 1.008 \;\cdot 7264.2719268 \;= 7322.386$$

$$\delta \;=\; 100 \;\cdot\; |7320 \;-\; 7322.386| \;= 0.09\%$$

Ещё один вариант суждения об адекватности модели базируется на правиле 2-3 сигм (σ): модель признаётся адекватной, если абсолютные ошибки прогноща не превосходя 2-3 σ в нашем примере мы бы признали модель (6) адекватной, если абсолютная ошибка прогноза:

$$e\left(\tilde{G}_{2018}\right) = |\tilde{G}_{2018} - G_{2018}| \le 2 \cdot \tilde{\sigma}_w = 280 \text{ млрд. руб.}$$

Модифицируем модель Самуэльсона-Хикса при помощи более глубоко обсуждения диаграммы рассеивания. Вернёмся к диаграмме рассеивания "лаговое ВВП России - текущее потребление домохозяйтсв"и внимательно изучим эту диаграмму, обращая внимание на наличие явныъ выбросов. Рассматривая диаграмму можем сделать следующий вывод: первый очевидный выброс датируется 2009 годом C это значит, что его причиной является мировой финансвый кризис. Остальные выбросы датируется следующими годами: 2015, 2016, 2017. И их причиной являются санкции западных стран.

Итог: модель Самуэльсона-Хикса нужно модифицировать отразив в ней воздействие мирового крищиса и санкции западных стран. Вот модифиццированных фрагметн модели Самуэльсона-Хикса рассхода домохозяйств России.

Обозначим фиктивную переменную свзязанную с мировым финансовым кризисом Gr_t она равна 0, если в период t кризис отсутсвует и 1, если существует

$$Gr_t = \begin{cases} 0, \text{ если в период t кризис отсутсвует} \\ 1, \text{ если существует} \end{cases} \tag{9}$$

Gr_t –это *индикатор кризиса*. Аналогично индикатор крищиса:

$$San_t = \begin{cases} 1 \text{ , если существует} \\ 0, \text{ если отсутвует} \end{cases} \tag{9}$$

С помощью этих величин модифицируем модель Самуэльсона-Хикса:

$$\begin{cases} C_t = a_0 + a_1 \cdot Y_{t-1} + a_2 Gr_t + a_3 \cdot San_t + u_t \\ E(u_t) = 0; Var(u_t) = \sigma^2 \end{cases} \tag{10}$$

Спецификация (10) включает в себя 5 параметров: $(a_0, a_1, a_2, a_3, \sigma_u)(11)$. Обратим внимание, что спецификация (10) служит конкретным примером базовой модели эконометрики, которая носит называние *линейной модели множественной регресии*. Добавим, что при определённых свойствах случайного возмущения u_t параметры модели (11) оптимально оцениваются методом наименьших квадратов и на сегодняшнем занятии мы познакомимся с функцией ЛИНЕЙН в которой запрограммирована процедура наименьших квадратов с которой мы познакомились на прошлом занятии при оценивании модели (1).

Оценивание параметров (11) при помощи функции ЛИНЕЙН

1. На листе Excel занесём сивмол $date$ времени t(2003 -2017). Символом C_t - ВВП, Y_{t-1} - лаговый доход. $Cr_t -$ кризис. $San_t -$ санкции. $n = 15$. Ввели

заголовки и заполнили значениями переменных из обучающей модели. *Замечание.* Результат первого шага можно интерпретировать, как запись уравнений наблюдений в рамках модели (10)

2. Размещаем курсор со значениями эндогенных переменных (C_t) и кликаем по символу формул. В стоблце категория выбираем статистические далее выбираем ЛИНЕЙН и кликаем ОК. В первую строчку заносим адрес массива эндогенной переменной. Во вторую строчку заносим объясняющие. В третьей и четрвертой строчке следует вывеости 1.

3. Запишем модель точно также как модель (1).

$$\begin{cases} C_t \ = \ -12107 \ + \ 0.84\, Y_{t-1} \ - \ 3606\, Gr_t \ - \ 2009 \ \cdot San_t \ + \ u_t \\ \tilde{\sigma}_w \ = 406.5 \ .. \end{cases} \tag{12}$$

Комментарий. В первой строчке выделенного массива (протокола функции линейн) расположенны в обратном порядке оценки коэффициентов. Мера точности во второй строке. Величина $\tilde{\sigma}_w$ всегда содердится в 3 строчке 2 столбца протокола. Остальное содержимое мы обсудим позже.

$\boxed{\text{ДЗ}}$ Проанализировать диаграмму рассеивания остальных элементов диаграммы Самуэльсона-Хикса и если есть основания модифицировать отразив на низ влияние кризиса и санкций. Воспользоваться функцией линейн и оценить параматеры двух остальных параметров модели Самуэльсона-Хикса.

Семинар №6

$$\begin{cases} C_{2003} \;=\; a_0 \;+\; a_1 \;\cdot Y_{2002} \;+\; a_2 \;\cdot Cr_{2003} \;+\; a_3 \;\cdot San_{2003} \;+\; u_{2003}; \\ C_{2017} \;\; a_0 \;+\; a_1 \;\cdot Y_{2016} \;+\; a_2 \;\cdot Cr_{2017} \;+\; a_3 \;\cdot San_{2017} \;+\; u_{2017}; \end{cases} \tag{2}$$

$$P \;=\; \left(a_0,\; a_1,\; a_2,\; a_3;\; \sigma_u^2\right) \tag{3}$$

Компактная запись:

$$\vec{y} \;=\; X \;\cdot \overrightarrow{a} \;+\; \vec{u}; \tag{4}$$

Ситуации уравнений (2) наблюдений составить формулу в компактную запись (4) этих уравнений.

$$\vec{y} = \begin{pmatrix} C_{2003} \\ \dots \\ C_{2017} \end{pmatrix};\; \vec{a} \;=\; \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix};X \;=\; \begin{pmatrix} 1 & Y_{2002} & Gr_{2003} & San_{2003} \\ \dots & \dots & \dots & \dots \\ 1 & Y_{2016} & Gr_{2017} & San_{2017} \end{pmatrix};\vec{u} \;=\; \begin{pmatrix} u_{2003} \\ \dots \\ u_{2017} \end{pmatrix} \tag{5}$$

Обратим внимание, что первый столбец матрицы X, состоит из 1 тогда и только тогда, когда есть сводный член a_0. Вспоминая действия с матрицами мы проверим, что элементы компактной записи (5), модели (4), идентичны системе (2).

Тогда в итоге получится:

$$\begin{pmatrix} C_{2003} \\ \dots \\ C_{2017} \end{pmatrix} \;=\; \begin{pmatrix} 1 & Y_{2002} & Gr_{2003} & San_{2003} \\ \dots & \dots & \dots & \dots \\ 1 & Y_{2016} & Gr_{2017} & San_{2017} \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} \;+\; \begin{pmatrix} u_{2003} \\ \dots \\ u_{2017} \end{pmatrix}$$

$\boxed{\text{ДЗ}}$ составить элементы компактной запись, остальных двух элементов модифицированной модели Самуэльсона-Хикса.

Случайный вектор и его основные характеристики

Обратимся к компактной записи (4) и подчеркнём, что вектор случайных возмущений $\vec{u}$ предсавляет собой набор величин случайного вектора. У случайного вектора есть две важнейшие для практики количественные характеристики:

1. Математическое ожидание

$$E\left(\vec{u}\right) \;=\; \begin{pmatrix} E(u_1) \\ \dots \\ E(u_n) \end{pmatrix} \;=\; \begin{pmatrix} 0 \\ \dots \\ 0 \end{pmatrix} \;=\; \vec{0}; \tag{6}$$

2. Ковариация:

$$Cov\left(\vec{u}, \vec{u}\right) \;=\; V\left(\vec{u}\right) \;=\; \begin{pmatrix} \sigma_{u_1}^2 & \sigma_{1,2} & \dots & \sigma_{1,n} \\ \sigma_{2,1} & \sigma_{u_2}^2 & \dots & \sigma_{2,n} \\ \dots & \dots & \ddots & \dots \\ \sigma_{n,1} & \sigma_{n,2} & \dots & \sigma_{u_n}^2 \end{pmatrix} \tag{7}$$

Обычно постурируется некоррелированность элементов случайного вектора (вектора случайных возмущений) и поэтому коварicionная матрица в векторе случайных возмущений (2) будет выглядить так:

$$Cov\left(\vec{u}, \vec{u}\right) \;=\; \sigma_u^2 \;\cdot\; I \;\; \text{(еденичная матрица)} \tag{7$'$}$$

Основные количественные характеристики аффинного преобразования случайного вектора

$$\vec{v} \;=\; A \;\cdot \vec{u} \;+\; \vec{b} \tag{8}$$

Аффинным преобразование $\vec{u}$ является вектор $\vec{v}$, который рассчитывается по формуле (8).

Вот важное для практики правила рассчёта основных характеристик вектора $\vec{b}$.

$$\begin{aligned} E\left(\vec{v}\right) \;&=\; A \;\cdot\; E\left(\vec{u}\right) \;+\; b; \\ Cov\left(\vec{u}, \vec{u}\right) \;&=\; A \;\cdot\; Cov\left(\vec{u}, \vec{u}\right) \;\cdot\; A^T; \end{aligned} \tag{9}$$

Пусть в схеме Гаусса-Маркова (4) вектор $\vec{u}$ является случайным с остальными характеристиками и коварicionная матрица имеет вид (7'), $\vec{a}$ и X является не случайными, показать:

1. Что вектор $\vec{y}$ будет случайным;
2. Определить его характеристики;

В правой части (4) первое слагаемое $X \cdot \vec{a} \;=\; \alpha$ –это вектор констант, его смысл смотри ниже, а второй вектор случайный $\vec{u}$ и мы можем трактовать вектор $\vec{y}$ выражения (4), как афинное преобразование ветора $\vec{u}$. Рассматривая уравнение (4) мы убеждаемся, что матрица A является единичной и мы можем использовать вот эти важные для практики формулы:

$$E\left(\vec{y}\right) \;= \alpha \;+\; I \cdot E\left(\vec{u}\right)(=0) \;=\; \alpha$$

Первое слагаемое в правой части (4) имеет смысл ожидаемого значения вектора $\vec{y}$, т.е. первое слагаемое состоит из компонент равных ожидаемых значений $(E\left(\vec{y}\right))$ эндогенной переменной модели.

Тогда:

$$Cov\left(\vec{y},\; \vec{y}\right) = I \;\cdot\; Cov\left(\vec{u}, \vec{u}\right) \;\cdot\; I^T \;=\; Cov\left(\vec{u}, \vec{u}\right); \;=\; \sigma_u^2 \;\cdot\; I$$

Свойство оценок параметров методом наименьших квадратов

На сегодняшней лекции мы сформулируем важный для практики результат, состоящий в том, что оптимальные оценки коэффициентов вектора $\vec{a}$ из уравнения коэффициентов (4) вычисляются по правилу (11):

$$\widetilde{\vec{a}} = \left(X^T \cdot X\right)^{-1} \cdot X^T \cdot \vec{y};$$

$\boxed{\text{ДЗ}}$ Показать, что вектор $\widetilde{\vec{a}}$ является случайным и опираясь на форммулы (9) найти ожидаемое значение вектора ($E\left(\widetilde{\vec{a}}\right) - ?$) и его коварриционную матрицу ($Cov\left(\widetilde{\vec{a}},\ \widetilde{\vec{a}}\right) -?$).

Семинар №8
Тестирование предпосылки Гаусса-Маркова
о постоянстве дисперсии случаных возмущений
План

1. Тест Голдфелта-Кванта гомоскедастичность случайных возмущений;
2. тестирование гомоскедастичности случайных возмущений в модели Самуэльсона-Хикса расходов домохозяйств России;
3. ДЗ

Оценки параметров линейной эконометрич модели, вычисленные МНК является оптимальными при справедливых предпосылках **Гаусса-Маркова:**

$$H_0: \ Var(u_1) = \ \dots \ = Var(u_n) = \sigma_u^2 \tag{1}$$

Если (1) верно, то говорят, что случайные возмущения в модели *гомоскедастичность* (дисперсия возмущений не зависит от значений случайных переменных, в противном случае, гетероскедастичность.)

$$\begin{cases} C_t \ = a_0 \ + a_1 Y_{t-1} \ + \ u_t; \\ E(u_t) \ = \ 0; Var(U_t) = \sigma_u^2 \end{cases} \tag{2}$$

Тест предпосылки (1) называется тестом Гаусса-Маркова и состоит из следующих шагов:

Шаг 1. Составление уравнений наблюдений в рамках тестируемой модели:

$$\begin{cases} C_{2003} = a_0 + a_1 Y_{2002} + u_{2003} \\ \dots \\ C_{2018} = a_0 + a_1 Y_{2007} + u_{2018} \end{cases} \tag{3}$$

Если предпосылка (1) верна, то дисперсии случайных возмущений в уравнении (3) все одинаковые и следовательно не зависят от значений лагового ВВП(Y). Если предпосылка (1) нарушается, то это нарушение, как правило означает, следующее: *с ростом абсолютных значнений объясняющей переменной увеличивается и дисперсия случайного возмущения*.

Шаг 2. Вычисляются абсолютные значения объясняющей перменной и уравнения наблюдений упорядычеваются по возрастанию велечины

$Z_t = |Y_{t-1}|, \ t \ = 2003, \ \dots, 2018$.

Шаг3. В упорядоченной системе уравнений наблюдений отмечаются n_1 первых уравненений и n_1 последних уравнений. Оптимальное значение:

$n_1 \ = \dfrac{1}{3}n$ (от общего кол-ва уравнений),

$$n_1; \ n_1 \ \approx \frac{1}{3}n, \ n_1 \ > \ k \ + \ 1 \tag{4}$$

Шаг 4. По первым n_1 уравнений вычисляется:

$$\begin{cases} ESS_1 = \sum \widetilde{u}_i^{(1)^2} \\ ESS_2 = \sum \widetilde{u}_i^{(2)^2} \end{cases} \qquad (5)$$

По велечинам вычисляется величина:

$$GQ = \frac{ESS_1}{ESS_2} - \text{статистика Гольфилта-Кванта} \qquad (6)$$

Если гипотеза H_0 справедлива, то велечина GQ имеет *распределений Фишера*, где $m_1 = n_1 - (k+1)(=2)$:

$$\begin{cases} GQ = \dfrac{ESS_1}{ESS_2} \sim F_{m_1, m_1} \\ m_1 = n_1 - (k+1) \end{cases} \qquad (7)$$

В Excel вычисляется критическое значение $F_{\text{крит}}$ при заданном значении уровня значимости α, где обычно $\alpha = 0.05$

$$F_{\text{крит}_{1-\alpha}}; \alpha = 0.05$$

Шаг5. Проверяется справедливость двух неравенств:

$$\begin{cases} GQ \le F_{\text{крит}} \\ \dfrac{1}{GQ} \le F_{\text{крит}} \end{cases}$$

Если обе гипотезы справедливы, то гипотеза или предпосылка (1) принимается, как не противоречащая реальным данным. И случайные возмущения формулируются, как *гомоскедастичные*.

Если одно неравнств не верно, то предпосылка 1 является невыполненой и случайные возмущения называются *гетероскедастичность.*

Тестирование предпосылки один для модели 2

EXCEL

$\boxed{\text{ДЗ}}$ Дома протестировать для двух остальных оригинальных фрагментов модели Самуэльсона-Хикса

Семинар №9

Тестирование предпосылки об отсутсвии автокорреляции у случайного возмущения в эконометрической модели. Тест Дарбина-Уотсена

План

1. Тестирование отсутстивия автокорреляции у случайного возмущения в модифицированной модели Самуэльсона-Хикса расходов домохозяйств россии;
2. $\boxed{\text{ДЗ}}$

Вернёмся к предпосылкам теоремы Гаусса-Маркова и конкретно к предпосылке №3:

$$Cov(u_i, u_j) = 0,\ i \neq j$$

Её принято называть *предпослылкой об отсутсвии автокорреляции у случайного возмущения*. Подчеркнём, что если эта предпосылка нарушается, то процедура наименьших квадратов теряет свойство оптимальности, а во-вторых, причиной нарушения этой предпосылки чаще всего служит ошибка спецификации модели, например пропуск значащих объясняющих переменных модели. По этим причинам экономисты всегда тестируют эту предпослыку.

Порядок теста Дарбина-Уотсена предпослки №3 теоремы Гаусса-Маркова:

Проверяемая гипотеза

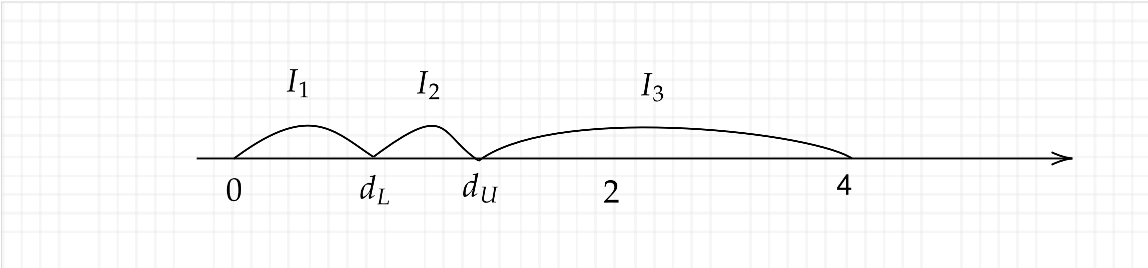
$$H_0 : Cov(u_{t+1}, u_t) = 0$$

Альтернативная гипотеза состоит в положительной корреляции состоит в положительной корреляции в два соседних момента времени.

$$H_1 : Cov(u_{t+1}, u_t) > 0$$

Тест осуществляется в виде следующих шагов:

1. Создаваемая модель оценивается МНК и вычисляются: оценки остатков и величины $\sum_{i=1}^{n} \tilde{u}_i^2$
2. Вычисляется статистика по правилу: $DW = \dfrac{\sum_{i=1}^{n-1}\left(\tilde{u}_{i+1} - \tilde{u}_i\right)^2}{\sum_{i=1}^{n}\left(\tilde{u}_i\right)^2}$
3. Из таблиц Дарбина-Уотсена по аргументам (n, k) определяются две величины d_L, d_U.
4. Определяется интервал куда попала статистика DW, если DW попало в I_3, то H_0 принимается, если в I_1, то гипотеза H_0 отвергается в пользу гипотезы H_1; если в интервал I_2, то ничего сказать нельзя - это *интервал неопределённости*.



Задача №1. Исследовать отсутсвие автокорреляции у случайного возмущения в модифицированной модели Самуэльсона-Хикса расходов домохозяйств России.

Решение.

Открываем файл Excel созданный на занятии номер №5.

Там есть оценённая часть поэтому шаг 1 выполнен (Кризис и санкции). Рассмотрим протокол оценивания данной модели:

a3	a2	a1	a0
-2009.150407	**-3606.31029**	**0.839684**	**-12106.6**
306.1664607	437.9424085	0.023019	851.7167
0.992798771	**406.519065**	#Н/Д	#Н/Д
505.5056988	11	#Н/Д	#Н/Д
250616203.5	**1817835.253**	#Н/Д	#Н/Д

$\sum_{i=1}^{n} \tilde{u}_i^2$ находится в пятой строчке второго столбца протокола.

Приступаем к расчётам. Приступаем к пункту 2 (вычисляется статистику DW)

Рассчитаем оценки случайных возмущений:

=E4-($H$21+$G$21*F4+$F$21*G4+$E$21*H4)

t	Ct	Yt-1	Crt	Sant	ut
2003	11159.8	27312.3	0	0	*H4)
2004	12550.7	29304.9	0	0	50.40268
2005	14087.4	31407.8	0	0	-178.669
2006	15799.7	33410.5	0	0	-148.004
2007	18060.8	36134.6	0	0	-174.287
2008	19967	39218.7	0	0	-857.757
2009	18946.6	41276.8	1	0	0
2010	19993.8	38048.6	0	0	151.5576
2011	21356.2	39762.2	0	0	75.07511
2012	23053.8	41457.8	0	0	348.9069
2013	24263.15976	42973.5	0	0	285.5576
2014	24736.37084	43740.70411	0	0	114.5597
2015	22418.53533	44063.79745	0	1	-465.422
2016	21780.76321	42945.28104	0	1	-163.994
2017	22511.92051	42871.14425	0	1	629.4153

a3	a2	a1	a0
-2009.150407	-3606.31029	0.839684	-12106.6
306.1664607	437.9424085	0.023019	851.7167
0.992798771	406.519065	#Н/Д	#Н/Д
505.5056988	11	#Н/Д	#Н/Д
250616203.5	1817835.253	#Н/Д	#Н/Д

Д3

t	Ct	Yt-1	Crt	Sant	ut	$(ut+1-ut)^2$
2003	11159.8	27312.3	0	0	332.657	
2004	12550.7	29304.9	0	0	50.40268	79667.52
2005	14087.4	31407.8	0	0	-178.669	52473.75
2006	15799.7	33410.5	0	0	-148.004	940.3323
2007	18060.8	36134.6	0	0	-174.287	690.8068
2008	19967	39218.7	0	0	-857.757	467130.5
2009	18946.6	41276.8	1	0	0	735746.4
2010	19993.8	38048.6	0	0	151.5576	22969.71
2011	21356.2	39762.2	0	0	75.07511	5849.575
2012	23053.8	41457.8	0	0	348.9069	74983.85
2013	24263.15976	42973.5	0	0	285.5576	4013.132
2014	24736.37084	43740.70411	0	0	114.5597	29240.3
2015	22418.53533	44063.79745	0	1	-465.422	336378.4
2016	21780.76321	42945.28104	0	1	-163.994	90858.98
2017	22511.92051	42871.14425	0	1	629.4153	629497.5
					Сумма:	2530441

a3	a2	a1	a0		
-2009.150407	-3606.31029	0.839684	-12106.6	DW=	=J19/F25
306.1664607	437.9424085	0.023019	851.7167		
0.992798771	406.519065	#Н/Д	#Н/Д		
505.5056988	11	#Н/Д	#Н/Д		
250616203.5	1817835.253	#Н/Д	#Н/Д		

Получаем:

DW=	1.3920078

Число $k = 3,\ n = 15$ из таблицы выбираем $d_L = 0.82,\ d_U = 1.75$ В нашем случай статистика DW попала в интервал между d_L, d_U следовательно мы не можем не принять не отвергнуть

$\boxed{\text{Д3}}$ Исследовать гипотезу о отсутвия автокорреляции случайного возмущения в инвистиционном фрагменте в модифицированной модели Самуэльсона-Хикса экономики России.

Семинар № 10

Исследование качества спецификации эконометрической модели

План

1. Коэффициент детерминации модели и F-тест качества её спецификации
2. Скорректированный коэффициент детерминации как инструмент отбора в модель объясняющих переменных (как инструмент модификации)
3. $\boxed{\text{ДЗ}}$

В процессе оценивания эконометрической мод. модели Самуэльсона-Хиггса расходов России методом наименьших квадратов полезно исследовать в качестве

$$\begin{cases} C_t = a_0 + a_1 \cdot Y_{t-1} + a_2 \cdot Cr_t + a_3 \cdot San_t + u_t; \\ E(u_t) = 0; Var(u_t) = \sigma_u^2 \end{cases} \quad (1)$$

Коэффициент детерминации, который вычисляется по следующему правилу:

$$R^2 = 1 - \frac{\sum \widetilde{u}_i^2}{\sum\limits_{i=1} (y_i - y)^2} \quad (2)$$

имеет смысл *доли текущих энодогенных переменных*, которые объясняются предопределёнными переменными. Этот коэффициент автоматически вычисляется в статистических пакетах например в протоколе функции ЛИНЕЙН размещается в 3 строчке 1 стоблца этого протокола. Этот коэффициент также определяется функцией ml пакета R.

F текст

F тест формализованная процедура проверки гипотезы о неудовлетворительной спецификации эконометрической модели, то есть гипотезы, что ни одна объясняющая переменная модели не несёт в себе какую либо информации об эндогенной переменной модели, то есть этот тест проверяет гипотезу, что все коэффициенты при объясняющих коэффициентах равны 0. Например приминительно к моделе (1) этот тест проверяет следующую гипотезу

$$H_0 : a_1 = a_2 = a_3 = 0 \quad (3)$$

Если эта гипотеза отвергается, то качество выбора объясняющих переменных считается удовлетворительным.

Задача. Исследовать качество спецификации модели (1) анализируя коэффициент детерминации этой модели и осуществляя F-тест. Статистика в критерии гипотезы (3) в F-тесте автоматически рассчитывается пакетами и в протоколе функции ЛИНЕЙН находится в 4 строчке 1 столбца протокола.

Решение:

Коэффицент детерминации

	a3	a2	a1	a0
	-2009.150407	**-3606.31029**	**0.839684**	**-12106.6**
	306.1664607	437.9424085	0.023019	851.7167
R2=	0.992798771	**406.519065**	#Н/Д	#Н/Д
	505.5056988	11	#Н/Д	#Н/Д
	250616203.5	**1817835.253**	#Н/Д	#Н/Д

В нашем примере коэффициент детерминации равен 0.99 и это значит, что в модели (1) лаговый доход, индикаторы кризиса и санкции объясняют на 99% текущий уровень потребления домохозяйств. Добавим добавим, что статистика критерия (3) о неудовлетворительной спецификации модели находится в 4 строчке 1 стобца и в нашем примере равна 505.5

Отметим правило по которому рассчитано значение статистики:

$$F = \frac{R^2 / k}{\left(1 - R^2\right) / (n - (k+1))}$$

Гипотеза (3) о неудовлетворительной спецификации должна быть отвергнута, если величина F превышает критический уровень $F_{\text{крит}}$. Этот уровень имеет смысл квантили F распределения уровня $1-\alpha$ с количеством степеней свободы $k=3$ (экзогенные переменные) и n (кол-во ур-ний наблюдений) $-k+1$; величина $n-k+1$ автоматически расчитывается функцией ЛИНЕЙН и всегда расположено в 4 строчке 2 столбца протокола. И у нас она равна 11. Рассчитаем $F_{\text{крит}}$ с помощью функции F.ОБР

	a3	a2	a1	a0			
	-2009.150407	**-3606.31029**	**0.839684**	**-12106.6**		DW=	**1.3920078**
	306.1664607	437.9424085	0.023019	851.7167		Fкрит =	**3.5874337**
R2=	0.992798771	**406.519065**	#Н/Д	#Н/Д			
F=	**505.5056988**	**11**	#Н/Д	#Н/Д			
	250616203.5	**1817835.253**	#Н/Д	#Н/Д			

Если $F > F_{\text{крит}}$ то мы отвергаем гипотезы.

Теперь осуществим F тест в $R\ Studio$.

Шаг 1. Копируем базу данных копируя из $Excel$ в блокнот и сохранить в рабочей директории.

Шаг 2. Открываем в $RStudio$

```
# Оценивание множественной регресси
getwd()
file.show("dataRStudio.txt")
C <-read.table("dataRStudio.txt", sep="", dec=".", header = TRUE)
C
Cmodel <- lm(data = C, Ct~Yt+Crt+Sant)
summary(Cmodel)
```

Гипотеза (3) о неудовлетворительной спецификации отвергается, если в протоколе $RStudio$ величина $p-value$ меньше чем 0.05. В нашем случае $p-value: 0.5$. Если же $p-value$ больше чем 0.5, то гипотеза не может быть отвергнута.

Скорректированный коэффициент детерминации как инструмент модификации модели (как инструмент отбора в модели дополнительных объясняющих переменных)

В протоколе функции $RStudio$ скоректированный коэффициент детерминации находится в предпоследней строчке протокола справа и в нашем примере он имеет значение 0.9908 обсудим методику отбора объясняющих переменных в эконометрическую модель.

Шаг 1. Сначала включаются все объясняющие переменные и фиксируется значение $\bar{R}^2$.

Шаг 2. Удаляется сомнительная переменная например идикатор санкции и фиксируется $\bar{R}^2$.

После удаления из модели San_t скорректированный коэффициент детерминации уменьшился 0.9587 является значащей её нельзя удалять из модели

```
# Оценивание множественной регресси
getwd()
file.show("dataRStudio.txt")
C <-read.table("dataRStudio.txt", sep="", dec=".", header = TRUE)
C
# Модификация модели "От общего к частному"
# Шаг 1
Cmodel<-lm(data = C, Ct~Yt+Crt+Sant)
summary(Cmodel)

# Шаг 2
Cmodel2<-lm(data = C, Ct~Yt+Crt)
summary(Cmodel2)
```

Шаг 3. Удалим следующую сомнительную переменную идикатор кризиса.

```
# Оценивание множественной регресси
getwd()
file.show("dataRStudio.txt")
C <-read.table("dataRStudio.txt", sep="", dec=".", header = TRUE)
C
# Модификация модели "От общего к частному"
# Шаг 1
Cmodel<-lm(data = C, Ct~Yt+Crt+Sant)
summary(Cmodel)
```

```
# Шаг 2
Cmodel2<-lm(data = C, Ct~Yt+Crt)
summary(Cmodel2)

# Шаг 3
Cmodel3<-lm(data = C, Ct~Yt+Sant)
summary(Cmodel3)
```

После удаления из модели Cr_t скорректированный коэффициент детерминации уменьшился 0.9398 является значащей её нельзя удалять из модели.

$\boxed{\text{ДЗ}}$ Исследовать качество модификации в двух оставшихся моделей.

Семинар №11

Прогнозирование по оценённой эконометрической модели и проверка её адекватности

План

1. Точечный прогноз по модели и характеристика точности прогноза (стандартная ошибка прогноза);
2. Интервальное прогнозирование и проверка адекватности модели;
3. $\boxed{\text{ДЗ}}$

Пусть в результате 3-его этапа построения модели получена оценённая модель множественной регрессии:

$$\begin{cases} y_t = \underset{(S\tilde{a}_0)}{\tilde{a}_0} + \underset{(S\tilde{a}_1)}{\tilde{a}_1} \cdot x_{1,t} + \underset{(S\tilde{a}_2)}{\tilde{a}_2} \cdot x_{2,t} + \underset{(\tilde{\sigma}_u)}{u_t}\,; \\ R^2 = \ldots,\ GQ = \ldots,\ DW = \ldots; \end{cases} \quad (1)$$

В скобках под оценками коэффициентов размешаются их стандартные ошибки. Эконометрические модели создаются в частности для прогноза неизвестных значений эндогенных переменнных в примере (1) для прогноза неизвестных значений величины y. Обозначим y_0 значение переменной y_t, которое неизвестно и его нужно спрогнозировать по известным значениям объясняющих переменных. Оптимальный прогноз величины y_0 рассчитывается по правилу:

$$\tilde{y}_0 = \tilde{a}_0 + \tilde{a}_1 \cdot x_{1,0} + \ldots + \tilde{a}_k \cdot x_{k,0} = \vec{x}_0^{\,T} \cdot \vec{\tilde{a}}; \quad (2)$$

$$\vec{x}_0^{\,T} = \begin{pmatrix} 1 & x_{1,0} & \ldots & x_{k,0} \end{pmatrix}, \vec{\tilde{a}} = \begin{pmatrix} \tilde{a}_0 \\ \tilde{a}_1 \\ \ldots \\ \tilde{a}_k \end{pmatrix}$$

Точность прогноза рассчитывается по правилу (3):

$$\begin{cases} S\tilde{y}_0 = \tilde{\sigma}_u \cdot \sqrt{1 + q_0} \\ q_0 = \vec{x}_0^{\,T} \cdot Q \cdot \vec{x}_0 \\ Q = \left(X^T \cdot X\right)^{-1} \end{cases} \quad (3)$$

Задача №1. Доказать, справедливость правила (3) оценки точности прогноза (2).

Решение: Обозначим символом $\triangle \tilde{y}_0 = \tilde{y}_0 - y_0$. Величина $\triangle \tilde{y}_0$ – это *истинная ошибка прогноза*.

$$\triangle \tilde{y}_0 = \tilde{y}_0 - y_0 = \vec{x}_0^{\,T} \cdot \vec{\tilde{a}} - (\underbrace{a_0 + a_1 \cdot x_{1,0} + \ldots + a_k \cdot x_{k,0}}_{\vec{x}_0^{\,T} \cdot \vec{a}} + u_0) = \vec{x}_0^{\,T} \cdot \underbrace{\left(\vec{\tilde{a}} - \vec{a}\right)}_{\triangle \vec{a}} - u_0 \quad (4)$$

$$Cov\left(\vec{\tilde{a}}, \vec{\tilde{a}}\right) = \widetilde{Cov}\left(\triangle \vec{\tilde{a}}, \triangle \vec{\tilde{a}}\right) = \tilde{\sigma}_u^2 \cdot Q \quad (5)$$

Утверждение D теоремы Гаусса-Маркова, а сейчас возращаемся к правилу (4) и отыщем основные количественные характеристики истинной ошибки прогноза:

$$E\left(\triangle \widetilde{y}_0\right) = E\left(\vec{x}_0^{\,T} \cdot \triangle \vec{a}\right) - E(u_0) = E\left(\vec{x}_0^{\,T} \cdot \left(\vec{\widetilde{a}} - \vec{a}\right)\right)$$

$$E\left(\vec{\widetilde{a}}\right) = \vec{a} \tag{6}$$

$$E\left(\triangle \widetilde{y}_0\right) = 0 \tag{7}$$

$$Var\left(\triangle \widetilde{y}_0\right) = E\left(\triangle \widetilde{y}_0^{\,2}\right) = \sigma_u^2 \cdot \vec{x}_0^{\,T} \cdot Q \cdot \vec{x}_0 + \sigma_u^2 = \sigma_u^2(1 + q_0) \tag{8}$$

$\boxed{\text{ДЗ}}$ Вывести формулу (7) и (8).

(8) доказывает формулу (3). Добавим, что правило (2) называется *точечным прогнозом*.

Задача. Сделать прогноз расхода домохозяйств России на 2018 год и оценить точность.

Приступаем к расчёту точечного прогноза расхода домохозяйств на 2018 год. Нам потребуются значения объясняющих переменных.

		Y_{2017}	Cr2017	San2017
x_0^T=	1	43533.75	0	1
C^p_{2018}=	22438.88			
C_{2018}=	23010			

Шаг 1. Как мы видим наш прогноз отличается от реального значения, поэтому вычислим стандартную ошибку прогноза. Формируем матрицу X у которой 15 строк и 4 столбца

X=	1	27312.3	0	0
	1	29304.9	0	0
	1	31407.8	0	0
	1	33410.5	0	0
	1	36134.6	0	0
	1	39218.7	0	0
	1	41276.8	1	0
	1	38048.6	0	0
	1	39762.2	0	0
	1	41457.8	0	0
	1	42973.5	0	0
	1	43740.704	0	0
	1	44063.797	0	1
	1	42945.281	0	1
	1	42871.144	0	1

И транспанированную к ней X^T.

Шаг 2. Вычислим матрицу $X^T \cdot X$.

X^T X=	15	573928.6269	1	3
	573928.6	22386339280	41276.8	129880.2
	1	41276.8	1	0
	3	129880.2227	0	3

Шаг 3. Рассчитаем матицу Q.

Q=	4.389636	-0.0001174	0.456322804	0.693075587
	-0.00012	3.20632E-09	-1.49453E-05	-2.14112E-05
	0.456323	-1.4945E-05	1.160572215	0.190711062
	0.693076	-2.1411E-05	0.190711062	0.567222424

Шаг 4. Рассчитаем $q_0 = \vec{x}_0^T \cdot Q \cdot \vec{x}_0$

x_0=	1		x_0^T*Q=	-0.02822	7.70616E-07	-0.00359	0.328187309
	43533.74999						
	0		q_0=	0.333519			
	1						

Шаг 5. Рассчитаем значение стандартной ошибки.

SC^P_{2018}	469.440385

Замечание. Интервальный прогноз мы осуществим на следующем занятии.

$\boxed{\text{ДЗ}}$ Вычислить прогнозы гос. расходов и инвестиций по модели Самуэльсона-Хикса по контролирующей выборке.

Интервальное прогнозирование по оцененной эконометрической модели

1. Прогнозирование доверительными интервалами и проверка адекватности модели
2. Формулировка домашнего творческого задания на тему “построение модели делового цикла экономики России с учетом мирового финансового кризиса, санкций и стоимостью денег во времени”
3. $\boxed{\text{ДЗ}}$

Таким образом интервал нижняя граница $Y_0^- = 21407{,}07$, верхняя граница $Y_0^+ = 23470{,}7$ с вероятностью в 95% накрывает истинное значение. В нашем примере 2018 год, это год контролирующей выборки, и по скольку доверительный интервал накрыл значение эндогенной переменной из контролирующей выборки, то у нас есть основания интерпретировать оценнённую модифицированную модель Самуэльсона-Хикса, как адекватную (по ней можно прогнозировать).
Если доверительные интервалы (с разумными уровнями доворительной вероятности) не накрывают значение эндогенных переменных из контролирующей выборки, то модель признаётся не адекватной и следует искать ошибку вв спецификации (например, в пропуске значащих объясняющих переменных (это понятие будет точно определено на следующем занятии)). Одно из назначений модели это прогнозирование.

$\boxed{\text{ДЗ}}$ Осуществить прогнощирования к модифицированным моделям государственных расходов и инвестиций в России.

$\boxed{\text{ДТЗ}}$ Построение модели делового цикла экономики России с учетом мирового финансового кризиса, санкций и стоимостью денег во времени. Объясняющая переменная в такой модели выбрать привлекая модифицированный коэффициент детерминации, проверить адекватность оценённой модели с помощью интервального прогнозирования на дату контролирующей выборки. Исследовать остуствие автокореляции у случайных возмущений модели. Срок сдачи [через 2 недели (16 декабря)].

Структура $\boxed{\text{ДТЗ}}$:

A. Титульный лист имя автора и проверяющего.

B. Содержание:

- постановка задачи
- методичка решения
- полученные результаты и выводы
- Список литерутуры.
- Приложение

Оценивание модели может быть либо в Excel, либо в RStudio.

Семинар № 13

Поиск незначащих переменных объясняемых переменных в эконометрических моделях: t – тест

План

1. Поиск незначащих переменных при помощи протокола
2. lm () Rstudio
3. $\boxed{\text{ДЗ}}$

Базовая модель эконометрики

$$\begin{cases} y_t = \tilde{a}_0 + \tilde{a}_1 x_{1,t} + \ldots + \tilde{a}_k \cdot x_{k,t} + u_t; \\ R^2 = \ \ldots, Q = \ \ldots, DW = \ \ldots; \\ F = \ \ldots, \ F_{\text{крит}} = \ \ldots; \end{cases}$$

Эти ошибки позволяют проверить $H_0: \ a_j = 0$ (2) против $H_1: \ a_j \neq 0$ (3);

Если справедливо (2), то x_{ij} может быть удалена как незначащая. Процедура проверки H_0 против H_1 называется t – тест и имеет слеждующие шаги:

$$1. \ t_j = \frac{a_j}{S\tilde{a}_j} \tag{4}$$

$$2. \ \alpha \in [0.01; \ 0.05]$$

$t_{\text{крит}}$ – двусторонний квантиль распределения Стьюдента уровня $1 - \alpha$

$$t_{\text{крит}_{1-\alpha}}(n - (k+1))$$

$$3. \ |t_j| \leqslant t \tag{5}$$

Если справедливо, то H_0 может быть принята, как непротиворечащая.

$\boxed{!}$ Опаснее удалить значащую переменную (ошибка II рода), чем сохранить незначащую (I рода).

Если соответсвующая велечина p-value меньше чем принятый уровень α, то гипотеза H_0может быть отклонена в пользу гипотезы H_1. Обратим внимание на последний столбец со звёздочкой, наличие звёздочки (хотя бы одной) служит наглядным сигналом несправедливости неравенства (5). То есть сигналом о значимости объясняющей переменной в моделе.

В нашем примере p-value равно $4.07e - 05$ *** поэтому переменная San_t значащая и её целесообразно сохранить в моделе.

$\boxed{\text{ДЗ}}$ Проверить значимость переменных из домашнего творческого задания.

Семинар №14

Тестирование предпосылки Голдфилда-Кванта

и прогнозирование по оценённой ЛММР

План

1. Тест Голдфилда-Кванта и Дарбина-Уотсона;
2. Прогнозирование и проверка адекватности в R
3. ДЗ

Для прогнозирования и тестирования нам потребуются пакеты ggplot2, lmtest, dplyr, tseries.

Приступаем к тестированию модели модифицированной модели расходов домозяйств Самуэдьсона-Хикса.

```
library(ggplot2)
library(lmtest)
library(dplyr)
library(tseries)

C<-read.table("dataRStudio.txt", sep="", dec=".", header = TRUE)
C

Cmodel<-lm(data = C, Ct~Yt+Crt+Sant)
summary(Cmodel)

# тест Голдфилда-Кванта
gqtest(Cmodel, fraction=0.33, data=C, order.by=C["Yt"])
```

По скольку величина p-value больше чем 0.05, то гипотеза о гомоскедастичности случайного возмущения принимается.

ДЗ Исследовать гипотезу о гомоскедастичности в оригинальной модели рассходов домохозяйств России.

Следующий тест Дарбина-Уотсона. Этот тест запрограммирован в функции dwtest.

```
dwtest(Cmodel, alternative = c("greate"))
```

Рассматривая значение `p-value` мы констатируем, что это значение меньше чем 0.05, следовательно мы должны отколонить гипотезу H_0 в сторону ошибки.

ДЗ Осуществить исследование в R об отсутсвии гетерскедастичности и отсутвия автокорреляции созданной в домашнем творческом задании.

Приступаем к прогнозированию по модели и проверка адекватности.

```
x0<-data.frame(Yt=435333.75, Crt=0, Sant=1)
x0
predeict.lm(Cmodel, newdata = x0, interval = "prediction", level =
0.95)
```

`fit` - точечный прогноз. `lwr`, `upr` - нижняя и верняя границы доверительного интервала. Поскольку доверительный интервал накрыл значение объясняемой переменной расходов домохяйств, то мы имеем основание признать адекватной. Наша модель не прошла 1 тест, тест Дарбина-Уотсона и это заставляет нас искать причину, которая должна быть устранена в окончательной модели.

Семинар №15
План

$$\begin{cases} Y = A \cdot K^{\alpha} \cdot L^{\beta}; \\ A > 0; 0 < \alpha; 0 < \beta; \end{cases} \tag{1}$$

$$\begin{cases} Y = A \cdot K^{\alpha} \cdot L^{\beta} + u; \\ E(u) = 0; Var(u) = \sigma_u^2; \end{cases} \tag{2}$$

$$\left(A, \alpha, \beta; \sigma_u^2\right) \tag{3}$$

$$\ln Y = \ln A + \alpha \cdot \ln K + \beta \cdot \ln L \tag{4}$$

$$\begin{cases} Y = A \cdot K^{\alpha} \cdot L^{\beta} \cdot e^{u}; \\ E(u) = 0; Var(u) = \sigma_u^2; \end{cases} \tag{5}$$

$$\begin{cases} \underset{y}{\ln Y} = \underset{a_0}{\ln A} + \alpha \cdot \underset{x_1}{\ln K} + \beta \cdot \underset{x_2}{\ln L} + u; \\ E(u) = 0; Var(u) = \sigma_u^2; \end{cases} \tag{6}$$

$$(A, \alpha, \beta; \sigma_u) \tag{7}$$

$$\alpha + \beta = 1 \tag{8}$$

$$Y = A \cdot K^{\alpha} \cdot L^{1-\alpha}; \tag{1'}$$

$$\frac{Y}{L} = A \cdot \left(\underset{k}{\frac{K}{L}}\right)^{\alpha} \tag{9}$$

Задача №1. Трансформировать в ситуации (8) модель (5) к линейной модели парной регрессии.

Решение:

$$\ln Y = \ln A + \alpha \cdot \ln K + (1-\alpha) \cdot \ln L$$

$$E(u) = 0; Var(u) = \sigma_u^2;$$

$$\begin{cases} Y = A \cdot K^{\alpha} \cdot L^{1-\alpha} \cdot e^{u}; \\ E(u) = 0; Var(u) = \sigma_u^2; \end{cases} \tag{10}$$

$$\frac{Y}{L} = A \cdot \left(\frac{K}{L}\right)^{\alpha} \cdot e^{u};$$

$$\begin{cases} \underset{y}{\ln \frac{Y}{L}} = \underset{a_0}{\ln A} + \alpha \cdot \underset{X}{\ln \frac{K}{L}} + u; \\ E(u) = 0; Var(u) = \sigma_u^2; \\ (a_0, \alpha, \sigma_u); \end{cases} \tag{11}$$

Задача №2. По данным ВВП основного капитала и труда в России оценить модель производства товара и услуг в стране, предполагая, что значения эластичности α и β удовлетворяют условию (8) "Постоянства отдачи от масштабов производства".

Решение:

Шаг 1. Составляем спецификаю эконометрической модели (10) и затем трансформируем эту модель к модели парной линейной регрессии (11).

Шаг 2. Переносим информацию в Excel

Год	Y(млрд. рублей)	K (млрд. рублей)	L(млн. чел.)
2000	7305	4306	65273
2001	7678	4979	65124
2002	8042	5518	66266
2003	8633	6640	67152
2004	9250	8103	67134
2005	9839	9207	68603
2006	10597	11277	69157
2007	11455	13403	70814
2008	12097	15527	70603
2009	11205	13100	69410
2010	11652	14444	69934
2011	12212	16120	70857

Шаг 3. Готовим уравнения наблюдей в качестве уравнений (11).

Год	LN(Y/L)	LN(K/L)
2000	4.717736	4.189186203
2001	4.769821	4.336691202
2002	4.798756	4.42209381
2003	4.856389	4.593908536
2004	4.925688	4.79329902
2005	4.965773	4.899383078
2006	5.031947	5.09414125
2007	5.086125	5.24317712
2008	5.14364	5.393263086
2009	5.084084	5.24033656
2010	5.115681	5.330482442
2011	5.149511	5.427152255

Шаг 4. Обращаемся к функции ЛИНЕЙН метода наименьших квадратов.

alpha	a0
0.35046	**3.248419898**
0.001476	0.007280466
0.999823	**0.002155382**
56355.51	10
0.261809	4.64567E-05

Шаг 5. Записываем оценённую модель (11)

$$\begin{cases} \ln \frac{Y}{L} = 3.25 + 0.35 \cdot \ln\left(\frac{K}{L}\right) + u \\ R^2 = 0.99 \end{cases} \tag{12}$$

Это обозначает, что относительное влияние на уровень ВВП неучтённых факторов состовляет 0.2.

Шаг 6. От трансформированной модели (12) возвращаемся к оценённой модели (10), для этого найдём оценки A при помощи операции потанцеирования. $A = 25.75$, записываем оценённую модель.

$$\left\{ Y = \underset{S_{\widetilde{A}}=?}{25.75} \cdot \underset{0.0015}{K^{0.35}} \cdot \underset{0.0015}{L^{0.65}} \cdot e^{u = 0.002} \right. \tag{13}$$

$\boxed{\text{ДЗ}}$ По статестическим данным оценить модель со спецификацией (6), не предполагая справедливость равенства (8).

$\boxed{\text{ДЗ Факультативная}}$ Вернёмся к оценённым моделям (12) и (13). Оценка $\widetilde{A}$ вычислена по правилу (14):

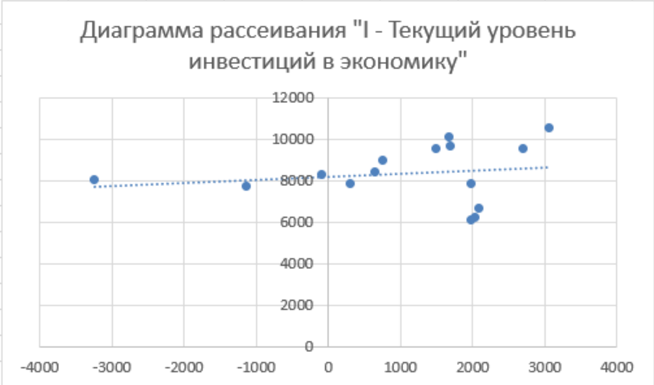
$$\widetilde{A} = e^{\tilde{a}_0},\ S\tilde{a}_0 = 0.007 \tag{14}$$

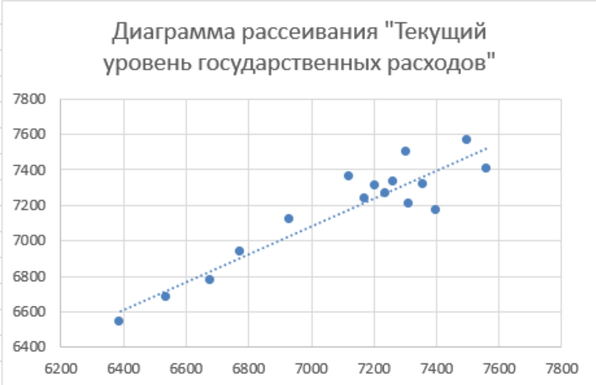
Рассуждая в дифференциалах получить ошибку $\widetilde{A}$.

Эконометрика. Домашняя работа №3
Аверьянов Тимофей ПМ 3-1

Задача №1. Осуществить аналогичное исследование модели Самуэльсона-Хикса для уравнения инвестиций и государственных расходов.

Решение:





Рассматривая, эти диаграммы мы можем сделать следующие выводы:

1. Точки реальных данных разметились вдоль восходящей прямой;
2. Точки реальных данных хоатично разместились вдоль восходящей прямой, то выше, то ниже, то ближе, то дальше.

Вот аналитические записи данных диаграмм:

$$\begin{cases} I_t = b \cdot (Y_{t-1} - Y_{t-2}) + v_t \\ E(v_t) = 0; Var(v_t) = \sigma_v^2 \end{cases} \quad (1)$$

$$\begin{cases} G_t = g \cdot G_{t-1} + w_t \\ E(w_t) = 0; Var(w_t) = \sigma_w^2 \end{cases} \quad (2)$$

Итог: Выражение (1) и (2) более адекватно описывают текущие уровени инвестиций в экономику и государственных расходов.

Эконометрика. Домашняя работа №4
Аверьянов Тимофей ПМ 3-1

Задача №1. Доказать (по правилами дифференцирования сложной функции), что величина $\widetilde{g}$, обеспечивающая эксремум $ESS\left(\widetilde{g}\right)$ может быть вычислена в процессе решения линейного алгербраического уравнения $R \cdot \widetilde{g} = S$, где $R = \sum_{t=2003}^{T} G_{t-1}^2$, $S = \sum_{t=2003}^{T} G_{t-1} \cdot G_t$.

Решение:

$$ESS\left(\widetilde{g}\right) = \sum_{t=2003}^{t=T} \left(G_t - \widetilde{g} \cdot G_{t-1}\right)^2 \to \min$$

$$\frac{\partial ESS\left(\widetilde{g}\right)}{\partial \widetilde{g}} = -2 \sum_{t=2003}^{T} \left(G_t - \widetilde{g} \cdot G_{t-1}\right) \cdot G_{t-1} = 0 \Rightarrow$$

$$\Rightarrow \widetilde{g} \sum_{t=2003}^{T} G_{t-1}^2 = \sum_{t=2003}^{T} G_t \cdot G_{t-1}$$

$$\widetilde{g} = \frac{\sum_{t=2003}^{T} G_t \cdot G_{t-1}}{\sum_{t=2003}^{T} G_{t-1}^2} = \frac{S}{R} \blacksquare$$

Задача №2. Завершить обсуждение **2 этапа** рассчётом: $\widetilde{\sigma}_w = \sqrt{\dfrac{\sum\left(\widetilde{w}_t^2\right)}{n(=15)-1}}$.

Решение:

Вычислим оценку среднеквадратичного отклонения случайного возмущения или меру влияния неучтённых факторов. Для это вычислим квадратичные оценки случайных фаткоров $\widetilde{w}_t$:

	wt2
	10200.39
	7821.715
	2013.469
	10905.66
	18339.76
	33992.52
	10418.54
	27265.16
	2071.716
	18424.92
	38.82768
	47920.98
	83092.57
	151.7303
	883.7483
Sum =	273541.7

А затем по формуле $\widetilde{\sigma}_w = \sqrt{\dfrac{\sum\left(\widetilde{w}_t^2\right)}{n(=15)-1}}$

= КОРЕНЬ(AB19/(15 - 1))

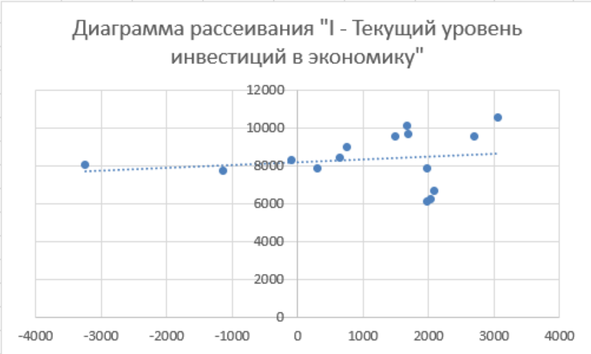
U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
		t	Gt	Gt-1	wt		wt2			Gt-12		G_t*G_{t-1}	
		2003	6540.2	6390	100.997		10200.39			40832100		41791878	
		2004	6679	6540.2	88.44046		7821.715			42774216.04		43681995.8	
		2005	6775.3	6679	44.8717		2013.469			44609041		45252228.7	
		2006	6931.9	6775.3	104.4302		10905.66			45904690.09		46965702.07	
		2007	7120.7	6931.9	135.4244		18339.76			48051237.61		49359980.33	
		2008	7359.9	7120.7	184.3706		33992.52			50704368.49		52407639.93	
		2009	7314.5	7359.9	-102.0712		10418.54			54168128.01		53833988.55	
		2010	7205.7	7314.5	-165.1217		27265.16			53501910.25		52706092.65	
		2011	7306.7	7205.7	45.51611		2071.716			51922112.49		52649888.19	
		2012	7498.7	7306.7	135.7384		18424.92			53387864.89		54790751.29	
		2013	7562.671	7498.7	6.231186		38.82768			56230501.69		56710202.35	
		2014	7401.995	7562.671	-218.9086		47920.98			57193995.32		55978855.19	
		2015	7170.733	7401.995	-288.2578		83092.57			54789531.85		53077728.23	
		2016	7238.265	7170.733	12.31788		151.7303			51419406.94		51903664.63	
		2017	7264.272	7238.265	-29.72791		883.7483			52392482.96		52580726.62	
				AVG =	3.616713	Sum =	273541.7		R =	757881587.6	S =	763691322.5	
					g=S/R=R-1*S=		1.007666			Sigma=	1))		
=	7322.386												

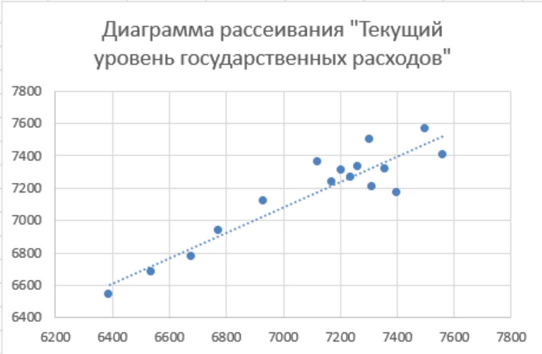
Таким образом $\widetilde{\sigma}_w \approx 140$.

Эконометрика. Домашняя работа №5

Аверьянов Тимофей ПМ 3-1

Задача №1. Проанализировать диаграмму рассеивания остальных элементов диаграммы Самуэльсона-Хикса и если есть основания модифицировать отразив на них влияние кризиса и санкций. Воспользоваться функцией линейн и оценить параматеры двух остальных параметров модели Самуэльсона-Хикса.





Рассматривая диаграммы можем сделать следующий вывод: первый очевидный выброс датируется 2009 годом это значит, что его причиной является мировой финансвый кризис. Остальные выбросы датируется следующими годами: 2015, 2016, 2017. И их причиной являются санкции западных стран.

Модель Самуэльсона-Хикса нужно модифицировать отразив в ней воздействие мирового кризиса и санкции западных стран.

Обозначим фиктивную переменную свзязанную с мировым финансовым кризисом Gr_t она равна 0, если в период t кризис отсутсвует и 1, если существует

$$Gr_t = \begin{cases} 0, & \text{если в период t кризис отсутсвует} \\ 1, & \text{если существует} \end{cases}$$

Gr_t – это *индикатор кризиса*.

Аналогично индикатор кризиса:

$$San_t = \begin{cases} 1, & \text{если существует} \\ 0, & \text{если отсутвует} \end{cases}$$

С помощью этих величин модифицируем модель Самуэльсона-Хикса для уравнения инвестиций и государственных расходов:

$$\begin{cases} I_t = b_0 + b_1 \cdot (Y_{t-1} - Y_{t-2}) + b_2 \cdot Cr_t + b_3 \cdot San_t + v_t; \\ E(v_t) = 0; \ Var(v_t) = \sigma_v^2 \end{cases}$$

$$\begin{cases} G_t = g_1 \cdot G_{t-1} + g_2 \cdot Cr_t + g_3 \cdot San_t + w_t; \\ E(w_t) = 0; \ Var(w_t) = \sigma_w^2 \end{cases}$$

Вычислим значение констант и запишем системы уравнений с вычесленными параметрами

Д3

t	It	triangleYt-1	Crt	Sant
2003	5396.9	1992.6	0	0
2004	6056.2	2102.9	0	0
2005	6631.1	2002.7	0	0
2006	7806.4	2724.1	0	0
2007	9526.5	3084.1	0	0
2008	10526.1	2058.1	0	0
2009	6209.8	-3228.2	1	0
2010	7982.2	1713.6	0	0
2011	9656.3	1695.6	0	0
2012	10084.86296	1515.7	0	0
2013	9525.04786	767.20411	0	0
2014	8947.736489	323.09334	0	0
2015	7848.354778	-1118.5164	0	1
2016	7700.652187	-74.136787	0	1
2017	8269.508	662.60574	0	1

b3	b2	b1	b0
-1059.82436	**-3743.4451**	**-0.3126**	**8944.098**
1584.788669	3427.5468	0.589281	1178.232
0.150751915	**1633.654**	#Н/Д	#Н/Д
0.650878147	11	#Н/Д	#Н/Д
5211240.505	29357080	#Н/Д	#Н/Д

t	Gt	Gt-1	Crt	Sant
2003	6540.2	6390	0	0
2004	6679	6540.2	0	0
2005	6775.3	6679	0	0
2006	6931.9	6775.3	0	0
2007	7120.7	6931.9	0	0
2008	7359.9	7120.7	0	0
2009	7314.5	7359.9	1	0
2010	7205.7	7314.5	0	0
2011	7306.7	7205.7	0	0
2012	7498.7	7306.7	0	0
2013	7562.671	7498.7	0	0
2014	7401.995	7562.671	0	0
2015	7170.733	7401.995	0	1
2016	7238.265	7170.733	0	1
2017	7264.272	7238.265	0	1

g2	g1	g0
-141.20014	**-141.866**	**1.013107**
88.01584075	140.6853	0.005747
0.999719361	**134.1769**	#Н/Д
14249.18953	12	#Н/Д
769603093	216041.2	#Н/Д

$$\begin{cases} I_t = 8944.1 + -0.3 \cdot (Y_{t-1} - Y_{t-2}) + -3743.4451 \cdot Cr_t + -1059.83 \cdot San_t + v_t; \\ E(v_t) = 0;\ Var(v_t) = 1633.654 \end{cases}$$

$$\begin{cases} G_t = 1.01 \cdot G_{t-1} + -141.866 \cdot Cr_t + -141.2 \cdot San_t + w_t; \\ E(w_t) = 0;\ Var(w_t) = 134.1769 \end{cases}$$

Домашняя работа №6 (Аверьянов Тимофей ПМ 3-1)

Задача №1. Составить элементы компактной запись, остальных двух элементов модифицированной модели Самуэльсона-Хикса.

Решение: Модифицированная модель Самуэльсона-Хикса выглядит следующим образом:

$$\begin{cases} C_t = a_0 + a_1 \cdot Y_{t-1} + a_2 \cdot Cr_t + a_3 \cdot San_t + u_t; \\ I_t = b_0 + b_1 \cdot (Y_{t-1} - Y_{t-2}) + b_2 \cdot Cr_t + b_3 \cdot San_t + u_t; \\ G_t = g_1 \cdot G_{t-1} + g_2 \cdot Cr_t + g_3 \cdot San_t + u_t; \\ Y_t = C_t + I_t + G_t; \end{cases}$$

Для первого уравнение элементы компактной записи были составлены на Семинаре №6. Запишем элементы компактной записи для остальных двух, начнём с объёма инвестиций страны I:

$$I_t = b_0 + b_1 \cdot (Y_{t-1} - Y_{t-2}) + b_2 \cdot Cr_t + b_3 \cdot San_t + u_t;$$

Компактная запись будет выглядить следующим образом:

$$\vec{I} = Y \cdot \vec{b} + \vec{u};$$

, элементами которой являются:

$$\vec{I} = \begin{pmatrix} I_{2003} \\ \dots \\ I_{2017} \end{pmatrix}; \vec{b} = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{pmatrix}; Y = \begin{pmatrix} 1 & Y_{2002} - Y_{2001} & Gr_{2003} & San_{2003} \\ \dots & \dots & \dots & \dots \\ 1 & Y_{2016} - Y_{2015} & Gr_{2017} & San_{2017} \end{pmatrix}; \vec{u} = \begin{pmatrix} u_{2003} \\ \dots \\ u_{2017} \end{pmatrix}$$

Теперь запишем компактную запись для государственных расходов G:

$$G_t = g_1 \cdot G_{t-1} + g_2 \cdot Cr_t + g_3 \cdot San_t + u_t;$$

Компактная запись будет выглядить следующим образом:

$$\vec{G} = Z \cdot \vec{g} + \vec{u};$$

, элементами которой являются:

$$\vec{G} = \begin{pmatrix} G_{2003} \\ \dots \\ G_{2017} \end{pmatrix}; \vec{g} = \begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix}; Z = \begin{pmatrix} G_{2002} & Gr_{2003} & San_{2003} \\ \dots & \dots & \dots \\ G_{2016} & Gr_{2017} & San_{2017} \end{pmatrix}; \vec{u} = \begin{pmatrix} u_{2003} \\ \dots \\ u_{2017} \end{pmatrix}$$

Задача №2.

$$\tilde{\vec{a}} = \left(X^T \cdot X\right)^{-1} \cdot X^T \cdot \vec{y}; \qquad (*)$$

Показать, что вектор $\tilde{\vec{a}}$ является случайным и опираясь на формулы (9) найти ожидаемое значение вектора ($E\left(\tilde{\vec{a}}\right) - ?$) и его коварионную матрицу ($Cov\left(\tilde{\vec{a}}, \tilde{\vec{a}}\right) - ?$).

$$\begin{cases} E(\vec{v}) = A \cdot E(\vec{u}) + b; \\ Cov(\vec{u}, \vec{u}) = A \cdot Cov(\vec{u}, \vec{u}) \cdot A^T; \end{cases} \qquad (9)$$

Решение: В правой части уравнения (*) матрица X является константой или фиксированной величиной, а вектор $\vec{y}$ является случайным значением в силу того, что:

$$\vec{y} = X \cdot \vec{a} + \vec{u}; \qquad (**)$$

в правой части (**) первое слагаемое $X \cdot \vec{a} = \alpha$ − это вектор констант, а второй вектор случайный $\vec{u}$ и мы можем трактовать вектор $\vec{y}$ выражения (**), как афинное преобразование ветора $\vec{u}$.
Следовательно, так как $\vec{y}$ является случайным, то и оптимальные оценки коэффициентов вектора $\vec{a}$ из уравнения коэффициентов (**), так же будут случайными оценками.
Наёдем математическое ожидание от оценок коэффициентов $\vec{a}$.
Для нахождения математического ожидания от оценок коэффициентов $\vec{a}$.
Перепишем уравнение (*) в другую форму:

$$\tilde{\vec{a}} = \left(X^T X\right)^{-1} \cdot X^T \cdot (X\vec{a} + \vec{u}) =$$
$$= \left(X^T X\right)^{-1} \cdot \left(X^T X\right) \cdot \vec{a} + \left(X^T X\right)^{-1} X^T \vec{u} =$$
$$= \left[\text{Вспомним, что } \left(X^T X\right)^{-1} \cdot \left(X^T X\right) = I\left(\text{еденичной матрице}\right)\right] =$$
$$= \vec{a} + \left(X^T X\right)^{-1} X^T \vec{u}$$

Возьмём математическое ожидание от левой и правой части и получим, вспоминая свойство вектора $\vec{u}$, а именно, что $E(\vec{u}) = 0$:

$$E\left(\tilde{\vec{a}}\right) = E(\vec{a}) + E\left(\left(X^T X\right)^{-1} X^T \vec{u}\right) = E(\vec{a}) + \left(X^T X\right)^{-1} X^T E(\vec{u}) = E(\vec{a}) = \vec{a}$$

Коварионная матрица вектора $\tilde{\vec{a}}$ по определению равна:

$$Cov\left(\tilde{\vec{a}}, \tilde{\vec{a}}\right) = E\left(\left(\tilde{\vec{a}} - \vec{a}\right)\left(\tilde{\vec{a}} - \vec{a}\right)^T\right) = \begin{pmatrix} \sigma_{\tilde{\vec{a}}_1}^{\ 2} & \sigma_{\tilde{\vec{a}}_1, \tilde{\vec{a}}_2} & \dots & \sigma_{\tilde{\vec{a}}_1, \tilde{\vec{a}}_n} \\ \sigma_{\tilde{\vec{a}}_2, \tilde{\vec{a}}_1} & \sigma_{\tilde{\vec{a}}_2}^{\ 2} & \dots & \sigma_{\tilde{\vec{a}}_2, \tilde{\vec{a}}_n} \\ \dots & \dots & \ddots & \dots \\ \sigma_{\tilde{\vec{a}}_n, \tilde{\vec{a}}_1} & \sigma_{\tilde{\vec{a}}_n, \tilde{\vec{a}}_2} & \dots & \sigma_{\tilde{\vec{a}}_n}^{\ 2} \end{pmatrix}$$

Её диагональные элементы равны $\sigma_{\tilde{\vec{a}}_1} = Var\left(\tilde{\vec{a}}_1\right)$ дисперсиям оценок отдельных коэффициентов. А диагональные элементы равны ковариациям оценок $\sigma_{\tilde{\vec{a}}_1, \tilde{\vec{a}}_2} = Cov\left(\tilde{\vec{a}}_1, \tilde{\vec{a}}_2\right)$. Заметим, что $\sigma_{\tilde{\vec{a}}_i, \tilde{\vec{a}}_j} = \sigma_{\tilde{\vec{a}}_j, \tilde{\vec{a}}_i}$, то есть матрица $Cov\left(\tilde{\vec{a}}, \tilde{\vec{a}}\right)$ симметричная относительно главной диагонали. Далее по выведенной нами формуле вычислим:

$$\tilde{\vec{a}} - \vec{a} = \left(X^T X\right)^{-1} X^T \vec{u}$$

И подстваим в формулу ковариации выше получим:

$$E\left(\left(\left(X^T X\right)^{-1} X^T \vec{u}\right)\left(\left(X^T X\right)^{-1} X^T \vec{u}\right)^T\right) =$$

$$= E\left(\left(X^T X\right)^{-1} X^T \left(\vec{u} \cdot \vec{u}^T\right) X \left(X^T X\right)^{-1}\right) =$$

$$= \left(X^T X\right)^{-1} X^T E\left(\vec{u} \cdot \vec{u}^T\right) X \left(X^T X\right)^{-1} =$$

$$= \left[\text{вспомним, что } E\left(\vec{u} \cdot \vec{u}^T\right) = \sigma_u^2 \cdot I\left(\text{еденичная матрица}\right)\right] =$$

$$= \sigma^2 \left(X^T X\right)^{-1} X^T X \left(X^T X\right)^{-1} = \sigma^2 \left(X^T X\right)^{-1}.$$

Сдедовательно ковариация равна:

$$Cov\left(\tilde{\vec{a}}, \tilde{\vec{a}}\right) = \sigma^2 \left(X^T X\right)^{-1}$$

In [1]:

```
# библиотеки/для latex
import IPython
import pandas as pd
import statsmodels.formula.api as smf
import numpy as np

# преобразлвание в latex таблицу
def _repr_latex_(self):
    return self.to_latex()
```

Уровень расхода домохозяйств

In [2]:

```
# загружаем базу данных
csvtab = pd.read_csv('C:/Users/timha/OneDrive/Рабочий стол/data2.csv', sep =
 ↪  ';', decimal='.', engine='python')
csvtab
```

Out [2]:

	t	Ct	Yt	Crt	Sant
0	2003	11159.80000	27312.30000	0	0
1	2004	12550.70000	29304.90000	0	0
2	2005	14087.40000	31407.80000	0	0
3	2006	15799.70000	33410.50000	0	0
4	2007	18060.80000	36134.60000	0	0
5	2008	19967.00000	39218.70000	0	0
6	2009	18946.60000	41276.80000	1	0
7	2010	19993.80000	38048.60000	0	0
8	2011	21356.20000	39762.20000	0	0
9	2012	23053.80000	41457.80000	0	0
10	2013	24263.15976	42973.50000	0	0
11	2014	24736.37084	43740.70411	0	0
12	2015	22418.53533	44063.79745	0	1
13	2016	21780.76321	42945.28104	0	1
14	2017	22511.92051	42871.14425	0	1

In [3]:

```
# Отбираем экзогенные и эндогенные переменные
Y = csvtab['Ct']
X = sm.add_constant(csvtab[['Yt', 'Crt', 'Sant']]) # добавляем
 ↪  intercept(const)
results = sm.OLS(Y, X).fit() # линейная регрессия
print('Residual standard error:', np.sqrt(results.scale)) # вычисление
 ↪  Residual standard error
results.summary()
```

```
Residual standard error: 406.519064173296
```

Out [3]:

```
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                     Ct   R-squared:                       0.993
Model:                            OLS   Adj. R-squared:                  0.991
Method:                 Least Squares   F-statistic:                     505.5
Date:                Sun, 20 Oct 2019   Prob (F-statistic):           4.62e-12
Time:                        15:59:07   Log-Likelihood:                -109.07
No. Observations:                  15   AIC:                             226.1
Df Residuals:                      11   BIC:                             229.0
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const      -1.211e+04    851.717    -14.214      0.000    -1.4e+04   -1.02e+04
Yt             0.8397      0.023     36.478      0.000       0.789       0.890
Crt        -3606.3103    437.942     -8.235      0.000   -4570.215   -2642.406
Sant       -2009.1504    306.166     -6.562      0.000   -2683.018   -1335.283
==============================================================================
Omnibus:                        3.036   Durbin-Watson:                   1.392
Prob(Omnibus):                  0.219   Jarque-Bera (JB):                1.156
Skew:                          -0.621   Prob(JB):                        0.561
Kurtosis:                       3.553   Cond. No.                     3.20e+05
==============================================================================
"""
```

Сравним с Excel:

a3	a2	a1	a0
-2009.150407	-3606.31029	0.839684	-12106.6
306.1664607	437.9424085	0.023019	851.7167
0.992798771	406.519065	#Н/Д	#Н/Д
505.5056988	11	#Н/Д	#Н/Д
250616203.5	1817835.253	#Н/Д	#Н/Д

Как мы видим результаты совпали

Объём инвестиция страны I

In [4]:

```
I = pd.read_csv('C:/Users/timha/OneDrive/Рабочий стол/data3.csv', sep = ';',
↪ decimal='.', engine='python')
I
```

Out [4]:

```
      t            It  triangleYt-1  Crt  Sant
0  2003   5396.900000   1992.600000    0     0
1  2004   6056.200000   2102.900000    0     0
2  2005   6631.100000   2002.700000    0     0
3  2006   7806.400000   2724.100000    0     0
4  2007   9526.500000   3084.100000    0     0
5  2008  10526.100000   2058.100000    0     0
6  2009   6209.800000  -3228.200000    1     0
7  2010   7982.200000   1713.600000    0     0
8  2011   9656.300000   1695.600000    0     0
```

```
9   2012  10084.862960   1515.700000    0     0
10  2013   9525.047860    767.204112    0     0
11  2014   8947.736489    323.093343    0     0
12  2015   7848.354778  -1118.516417    0     1
13  2016   7700.652187    -74.136787    0     1
14  2017   8269.508000    662.605739    0     1
```

In [5]:

```python
# Отбираем экзогенные и эндогенные переменные
Y = I['It']
X = sm.add_constant(I[['triangleYt-1', 'Crt', 'Sant']]) # добавляем
 ↪ intercept(const)
results = sm.OLS(Y, X).fit() # линейная регрессия
print('Residual standard error:', np.sqrt(results.scale)) # вычисление
 ↪ Residual standard error
results.summary()
```

```
Residual standard error: 1633.6540201758721
```

Out [5]:

```
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                     It   R-squared:                       0.151
Model:                            OLS   Adj. R-squared:                 -0.081
Method:                 Least Squares   F-statistic:                    0.6509
Date:                Sun, 20 Oct 2019   Prob (F-statistic):              0.599
Time:                        16:16:29   Log-Likelihood:                -129.94
No. Observations:                  15   AIC:                             267.9
Df Residuals:                      11   BIC:                             270.7
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const         8944.0977   1178.232      7.591      0.000    6350.827    1.15e+04
triangleYt-1    -0.3126      0.589     -0.530      0.606      -1.610       0.984
Crt          -3743.4451   3427.547     -1.092      0.298   -1.13e+04    3800.535
Sant         -1059.8244   1584.789     -0.669      0.517   -4547.921    2428.272
==============================================================================
Omnibus:                        0.910   Durbin-Watson:                   0.568
Prob(Omnibus):                  0.635   Jarque-Bera (JB):                0.693
Skew:                          -0.474   Prob(JB):                        0.707
Kurtosis:                       2.542   Cond. No.                     1.70e+04
==============================================================================
"""
```

Сравним с Excel:

b3	b2	b1	b0
-1059.82436	-3743.4451	-0.3126	8944.098
1584.788669	3427.5468	0.589281	1178.232
0.150751915	1633.654	#Н/Д	#Н/Д
0.650878147	11	#Н/Д	#Н/Д
5211240.505	29357080	#Н/Д	#Н/Д

Как мы видим результаты совпали

Государственные расходы G

In [6]:

```
G = pd.read_csv('C:/Users/timha/OneDrive/Рабочий стол/data4.csv', sep = ';',
 ↪ decimal='.', engine='python')
G
```

Out [6]:

	t	Gt	Gt-1	Crt	Sant
0	2003	6540.200000	6390.000000	0	0
1	2004	6679.000000	6540.200000	0	0
2	2005	6775.300000	6679.000000	0	0
3	2006	6931.900000	6775.300000	0	0
4	2007	7120.700000	6931.900000	0	0
5	2008	7359.900000	7120.700000	0	0
6	2009	7314.500000	7359.900000	1	0
7	2010	7205.700000	7314.500000	0	0
8	2011	7306.700000	7205.700000	0	0
9	2012	7498.700000	7306.700000	0	0
10	2013	7562.671176	7498.700000	0	0
11	2014	7401.995126	7562.671176	0	0
12	2015	7170.732664	7401.995126	0	1
13	2016	7238.265190	7170.732664	0	1
14	2017	7264.271927	7238.265190	0	1

In [7]:

```
# Отбираем экзогенные и эндогенные переменные
Y = G['Gt']
X = G[['Gt-1', 'Crt', 'Sant']] # intercept(const) здесь не нужно
results = sm.OLS(Y, X).fit() # линейная регрессия
print('Residual standard error:', np.sqrt(results.scale)) # вычисление
 ↪ Residual standard error
results.summary()
```

```
Residual standard error: 134.17688134260314
```

Out [7]:

```
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                     Gt   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 1.425e+04
Date:                Sun, 20 Oct 2019   Prob (F-statistic):           1.43e-21
Time:                        16:22:21   Log-Likelihood:                -93.098
No. Observations:                  15   AIC:                             192.2
Df Residuals:                      12   BIC:                             194.3
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Gt-1           1.0131      0.006    176.292      0.000       1.001       1.026
Crt         -141.8664    140.685     -1.008      0.333    -448.393     164.661
Sant        -141.2001     88.016     -1.604      0.135    -332.970      50.570
```

```
==============================================================================
Omnibus:                        4.116   Durbin-Watson:                   1.346
Prob(Omnibus):                  0.128   Jarque-Bera (JB):                2.742
Skew:                          -1.043   Prob(JB):                        0.254
Kurtosis:                       2.806   Cond. No.                     2.90e+04
==============================================================================
"""
```

Сравним с Excel:

g2	g1	g0
-141.20014	-141.866	1.013107
88.01584075	140.6853	0.005747
0.999719361	134.1769	#Н/Д
14249.18953	12	#Н/Д
769603093	216041.2	#Н/Д

Как мы видим результаты совпали

Эконометрика. Домашняя работа №8
Аверьянов Тимофей ПМ 3-1

Задание №1. Протестировать для двух остальных оригинальных фрагментов модели Самуэльсона-Хикса.

Решение:

Для решения поставленной задачи необходимо выполнить несколько шагов:

Шаг 1. Составление уравнений наблюдений в рамках тестируемой модели:

$$\begin{cases} I_{2003} = b_0 + b_1 \cdot (Y_{2002} - Y_{2001}) + v_{2003} \\ \dots \\ I_{2018} = b_0 + b_1 \cdot (Y_{2017} - Y_{2016}) + v_{2018} \end{cases}$$

Шаг 2. Вычисляются абсолютные значения объясняющей перменной и уравнения наблюдений упорядычеваются по возрастанию велечины $Z_t = |Y_{t-1} - Y_{t-2}|,\ t\ = 2003,\ \dots, 2018$.

Шаг 3. В упорядоченной системе уравнений наблюдений отмечаются n_1 первых уравненений и n_1 последних уравнений. Оптимальное значение: $n_1\ = \frac{1}{3}n$ (от общего кол-ва уравнений).

Шаг 4. По первым n_1 уравнений вычисляется:

$$\begin{cases} ESS_1 = \sum \tilde{u}_i^{(1)^2} \\ ESS_2 = \sum \tilde{u}_i^{(2)^2} \end{cases}$$

По велечинам вычисляется величина:

$$GQ\ = \frac{ESS_1}{ESS_2} - \text{статистика Гольфилта-Кванта}$$

Шаг 5. Проверяется справедливость двух неравенств:

$$\begin{cases} GQ\ \le F_{\text{крит}} \\ \dfrac{1}{GQ}\ \le F_{\text{крит}} \end{cases}$$

Выполним все эти шаги получим следующее решение:

t	It	triangleYt-1	z = |triangleYt-1|
2016	7700.652	-74.13678663	74.13678663
2014	8947.736	323.0933428	323.0933428
2017	8269.508	662.6057386	662.6057386
2013	9525.048	767.2041115	767.2041115
2015	7848.355	-1118.516417	1118.516417
2012	10084.86	1515.7	1515.7
2011	9656.3	1695.6	1695.6
2010	7982.2	1713.6	1713.6
2003	5396.9	1992.6	1992.6
2005	6631.1	2002.7	2002.7
2008	10526.1	2058.1	2058.1
2004	6056.2	2102.9	2102.9
2006	7806.4	2724.1	2724.1
2007	9526.5	3084.1	3084.1
2009	6209.8	-3228.2	3228.2

b1	b0	
0.683572	8381.66557	
0.427034	295.1554554	
0.460662	651.2572304	
2.562379	3	m_1
1086797	**1272407.94**	
	ESS_1	

b1	b0
0.404943	7479.056258
0.374269	1004.430148
0.280685	1941.996403
1.170632	3
4414864	**11314050.09**
	ESS_2

GQ	0.112463
1/GQ	8.891842
a	0.05
$F_{крит}$	9.276628

Таким образом проверим условие **шага 5**:

$$\begin{cases} GQ \ \le F_{\text{крит}} \\ \dfrac{1}{GQ} \ \le F_{\text{крит}} \end{cases}$$

$$H_0: \ Var(v_1) = \ \dots \ = Var(v_n) = \sigma_v^2 \qquad (1)$$

Получается, что гипотеза или предпосылка (1) принимается, как не противоречащая реальным данным. И случайные возмущения формулируются, как *гомоскедастичные*.

Аналогчино поступим и для государственных расходов, получим следующее решение:

t	Gt	Gt-1	Z = |Gt_1|
2003	6540.2	6390	6390
2004	6679	6540.2	6540.2
2005	6775.3	6679	6679
2006	6931.9	6775.3	6775.3
2007	7120.7	6931.9	6931.9
2008	7359.9	7120.7	7120.7
2016	7238.265	7170.732664	7170.7327
2011	7306.7	7205.7	7205.7
2017	7264.272	7238.26519	7238.2652
2012	7498.7	7306.7	7306.7
2010	7205.7	7314.5	7314.5
2009	7314.5	7359.9	7359.9
2015	7170.733	7401.995126	7401.9951
2013	7562.671	7498.7	7498.7
2014	7401.995	7562.671176	7562.6712

g	g0	
1.021967286	0	
0.002136581	#Н/Д	
0.999982517	31.84661	
228789.1459	**4**	m_1
232039496.4	**4056.827**	
	ESS_1	

g	g0
0.987033683	0
0.00679196	#Н/Д
0.999810633	112.8128
21119.01831	**4**
268776049.9	**50906.92**
	ESS_2

GQ	0.079691
1/GQ	12.54846
a	0.05
$F_{крит}$	6.388233

Таким образом проверим условие **шага 5**:

$$\begin{cases} GQ \le F_{\text{крит}} \\ \dfrac{1}{GQ} \le F_{\text{крит}} \end{cases}$$

$$H_0: Var(w_1) = \ldots = Var(w_n) = \sigma_w^2 \qquad (2)$$

Получается, что гипотеза или предпосылка (2) не принимается, как противоречащая реальным данным. И случайные возмущения формулируются, как *гетероскедастичные*.

Эконометрика. Домашняя работа №9

Аверьянов Тимофей ПМ 3-1

Задача № 1. Исследовать гипотезу об отсутвии автокореляции случайного возмущения в инвистиционном фрагменте в модифицированной модели Самуэльсона-Хикса экономики России.

Решение:

Инвистиционный фрагмент в модифицированной модели Самуэльсона-Хикса экономики России имеет следующий вид:

$$I_t = b_0 + b_1 \cdot (Y_{t-1} - Y_{t-2}) + b_2 \cdot Cr_t + b_3 \cdot San_t + v_t;$$

Проверяемая гипотеза

$$H_0 : Cov(u_{t+1}, u_t) = 0$$

Альтернативная гипотеза состоит в положительной корреляции состоит в положительной корреляции в два соседних момента времени.

$$H_1 : Cov(u_{t+1}, u_t) > 0$$

Тест осуществляется в виде следующих шагов:

Шаг 1. Создавамая модель оценивается МНК и вычисляются: оценки остатков и величину $\sum_{i=1}^{n} \tilde{v}_i^2$.

Откроем файл, который Excel вычислим оценки остатков и величину $\sum_{i=1}^{n} \tilde{v}_i^2$.

ДЗ

t	It	triangleYt-1	Crt	Sant
2004	6056.2	2102.9	0	0
2005	6631.1	2002.7	0	0
2006	7806.4	2724.1	0	0
2007	9526.5	3084.1	0	0
2008	10526.1	2058.1	0	0
2009	6209.8	-3228.2	1	0
2010	7982.2	1713.6	0	0
2011	9656.3	1695.6	0	0
2012	10084.86	1515.7	0	0
2013	9525.048	767.204111	0	0
2014	8947.736	323.093343	0	0
2015	7848.355	-1118.5164	0	1
2016	7700.652	-74.136787	0	1
2017	8269.508	662.605739	0	1

b3	b2	b1	b0
-1205.87	**-3663.3644**	**-0.2385**	**9103.238**
1370.406	2960.56432	0.510091	1020.245
0.248166	**1410.96979**	#Н/Д	#Н/Д
1.100268	10	#Н/Д	#Н/Д
6571358	**19908357.4**	#Н/Д	#Н/Д

Таким образом, шаг 1 выполнен рассмотрим протокол оценивания данной модели.

Величина $\sum_{i=1}^{n} \tilde{v}_i^2$ находится в пятой строчке второго столбца протокола.

Шаг 2. Вычисляется статистика по правилу: $DW = \dfrac{\sum\limits_{i=1}^{n-1}\left(\tilde{v}_{i+1} - \tilde{v}_i\right)^2}{\sum\limits_{i=1}^{n}\left(\tilde{v}_i\right)^2}$

Вычислим числитель дроби:

ДЗ

t	It	triangleYt-1	Crt	Sant	vt	$(v_{t+1}-v_t)^2$
2004	6056.2	2102.9	0	0	-2545.5	
2005	6631.1	2002.7	0	0	-1994.49	303603.497
2006	7806.4	2724.1	0	0	-647.139	1815363.2
2007	9526.5	3084.1	0	0	1158.821	3261491.97
2008	10526.1	2058.1	0	0	1913.72	569871.969
2009	6209.8	-3228.2	1	0	0	3662323.99
2010	7982.2	1713.6	0	0	-712.343	507433.169
2011	9656.3	1695.6	0	0	957.4636	2788255.4
2012	10084.86	1515.7	0	0	1343.12	148731.128
2013	9525.048	767.204111	0	0	604.7887	545133.592
2014	8947.736	323.093343	0	0	-78.4433	466805.886
2015	7848.355	-1118.5164	0	1	-315.778	56327.7431
2016	7700.652	-74.136787	0	1	-214.396	10278.3725
2017	8269.508	662.605739	0	1	530.1735	554383.232
					Сумма	14690003.1

b3	b2	b1	b0
-1205.87	**-3663.3644**	**-0.2385**	**9103.238**
1370.406	2960.56432	0.510091	1020.245
0.248166	**1410.96979**	#Н/Д	#Н/Д
1.100268	10	#Н/Д	#Н/Д
6571358	**19908357.4**	#Н/Д	#Н/Д

Получаем, что:

$$\sum_{i=1}^{n-1}\left(\tilde{v}_{i+1} - \tilde{v}_i\right)^2 = 14690003.1364207$$

Знаменатель этой дроби вычислен в шаге 1. Таким образом DW:

$$DW = \frac{14690003.1364207}{19908357.4294107} \approx 0.7378812234262531$$

При помощи Excel:

DW=	0.73788122

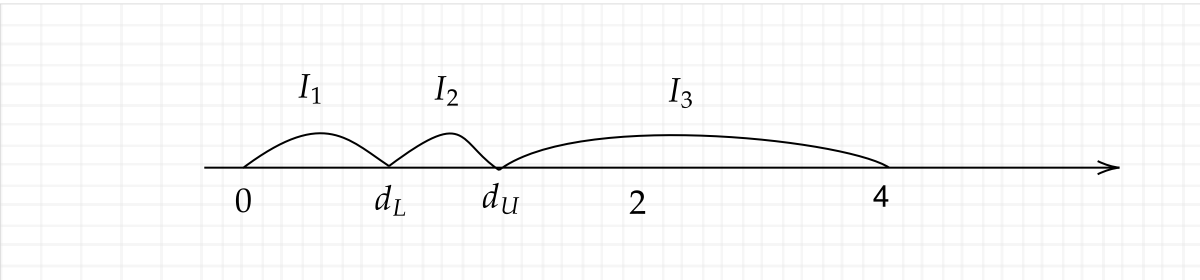
Шаг 3. По таблице Дарбина-Уотсена по аргументам (n, k) определяются две величины d_L, d_U.

n	$k^1 = 1$		$k^1 = 2$		$k^1 = 3$		$k^1 = 4$		$k^1 = 5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
6	0,61	1,40	–	–	–	–				
7	0,70	1,36	0,47	1,90	–	–				
8	0,76	1,33	0,56	1,78	0,37	2,29				
9	0,82	1,32	0,63	1,70	0,46	2,13				
10	0,88	1,32	0,70	1,64	0,53	2,02				
11	0,93	1,32	0,66	1,60	0,60	1,93				
12	0,97	1,33	0,81	1,58	0,66	1,86				
13	1,01	1,34	0,86	1,56	0,72	1,82				
14	1.05	1.35	0.91	1.55	0.77	1.78				
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83

Pic 1: Значения статистики Дарбина - Уотсона

Число $k = 3,\ n = 14$ из таблицы выбираем $d_L = 0.77,\ d_U = 1.78$.

Шаг 4. Определяется интервал куда попала статистика DW.



В нашем случае статистика DW попала в интервал I_1, следовательно гипотеза H_0 отвергается в пользу гипотезы H_1.

Вывод: Подчеркнём, что если эта предпосылка нарушается, то процедура наименьших квадратов теряет свойство оптимальности, а во-вторых, причиной нарушения этой предпосылки чаще всего служит ошибка спецификации модели, например пропуск значащих объясняющих переменных модели. По этим причинам экономисты всегда тестируют эту предпосылку.

Эконометрика. Домашняя работа № 10

Аверьянов Тимофей ПМ 3-1

Задача. Исследовать качество модификации в двух оставшихся моделей.

Решение: Начнём с объёма инвестиций страны I:

$$\begin{cases} I_t = b_0 + b_1 \cdot (Y_{t-1} - Y_{t-2}) + b_2 \cdot Cr_t + b_3 \cdot San_t + u_t; \\ E(u_t) = 0; Var(u_t) = \sigma_u^2 \end{cases} \quad (1)$$

Вычислим коэффициент детерминации для данной модели:

	b3	b2	b1	b0
	-1205.87	**-3663.3644**	**-0.2385**	**9103.238**
	1370.406	2960.56432	0.510091	1020.245
R2=	**0.248166**	**1410.96979**	#Н/Д	#Н/Д
	1.100268	10	#Н/Д	#Н/Д
	6571358	**19908357.4**	#Н/Д	#Н/Д

Таблица 1: Протокол объёма инвестиций страны I

В нашем примере коэффициент детерминации равен 0.24 (3 строчка 1 столбец) и это значит, что в модели (1) лаговый ВВП страны, индикаторы кризиса и санкции объясняют на 24% текущий уровень инвестиций страны. Добавим, что статистика критерия $H_0: a_1 = a_2 = a_3 = 0$ о неудовлетворительной спецификации модели находится в 4 строчке 1 стобца и в нашем примере равна 1.100268.

Отметим правило по которому рассчитано значение статистики:

$$F = \frac{R^2 / k}{\left(1 - R^2\right) / (n - (k+1))}$$

Гипотеза о неудовлетворительной спецификации должна быть отвергнута, если величина F превышает критический уровень $F_{\text{крит}}$. Этот уровень имеет смысл квантили F распределения уровня $1 - \alpha$ с количеством степеней свободы $k = 3$ (экзогенные переменные) и $n\left(\text{кол-во ур-ний наблюдений}\right) - k + 1$; величина $n - k + 1$ автоматически расчитывается функцией ЛИНЕЙН и всегда расположено в 4 строчке 2 столбца протокола. И у нас она равна 10. Рассчитаем $F_{\text{крит}}$ с помощью функции F.ОБР

	b3	b2	b1	b0			
	-1205.87	**-3663.3644**	**-0.2385**	**9103.238**			
	1370.406	2960.56432	0.510091	1020.245		DW=	**0.737881**
R2=	**0.248166**	**1410.96979**	#Н/Д	#Н/Д		Fкрит=	**3.708265**
	1.100268	10	#Н/Д	#Н/Д			
	6571358	**19908357.4**	#Н/Д	#Н/Д			

Если $F > F_{\text{крит}}$ то мы отвергаем гипотезу. В нашем случае это не так $F_{\text{крит}} > F$, следовательно гипотеза о неудовлетворительной принимается.

Теперь осуществим F тест в $R\ Studio$.

```
# Оценивание множественной регресси
```

```
getwd()
C <-read.table("I_invest.txt", sep="", dec=".", header = TRUE)
C
Cmodel<-lm(data = C, It~triangleYt+Crt+Sant)
summary(Cmodel)
```

Гипотеза о неудовлетворительной спецификации отвергается, если в протоколе $RStudio$ величина $p-value$ меньше чем 0.05. В нашем случае $p-value$: 0.3938. Если же $p-value$ больше чем 0.05, то гипотеза не может быть отвергнута. В нашем случае гипотеза не может быть отвергнута.

В протоколе функции $RStudio$ скоректированный коэффициент детерминации находится в предпоследней строчке протокола справа и в нашем примере он имеет значение 0.02262.

Удалим сомнительную переменная например идикатор санкции.

```
# Шаг 2
Cmodel2<-lm(data = C, It~triangleYt+Crt)
summary(Cmodel2)
```

После удаления из модели San_t скорректированный коэффициент детерминации увеличился 0.04267 , следовательно переменная НЕ является значащей её МОЖНО удалять из модели.

Удалим следующую сомнительную переменную идикатор кризиса.

```
# Шаг 3
Cmodel3<-lm(data = C, It~triangleYt+Sant)
summary(Cmodel3)
```

После удаления из модели Cr_t скорректированный коэффициент детерминации уменьшился -0.02458 , следовательно переменная является значащей её нельзя удалять из модели.

Проверим качество модификации государственных расходов G:

$$\begin{cases} G_t = g_1 \cdot G_{t-1} + g_2 \cdot Cr_t + g_3 \cdot San_t + u_t; \\ E(u_t) = 0; Var(u_t) = \sigma_u^2 \end{cases} \tag{2}$$

Вычислим коэффициент детерминации для данной модели:

	g3	g2	g1
	-141.2	**-141.866**	**1.013107**
	88.01584	140.6853	0.005747
R^2=	**0.999719**	**134.1769**	#Н/Д
F=	**14249.19**	**12**	#Н/Д
	7.7E+08	216041.2	#Н/Д

Таблица 2: Протокол государственных расходов G

В нашем примере коэффициент детерминации равен 0.999729 (3 строчка 1 столбец)

и это значит, что в модели (2) лаговый уровень государственных расходов, индикаторы кризиса и санкции объясняют на 99% текущий уровень инвестиций страны. Добавим, что статистика критерия $H_0: a_1 = a_2 = a_3 = 0$ о неудовлетворительной спецификации модели находится в 4 строчке 1 стобца и в нашем примере равна 14249.19.
Рассчитаем $F_{\text{крит}}$ с помощью функции F.ОБР:

	g3	g2	g1			
	-141.2	-141.866	1.013107			
	88.01584	140.6853	0.005747		Fкрит =	3.490295
R^2=	0.999719	134.1769	#Н/Д			
F=	14249.19	12	#Н/Д			
	7.7E+08	216041.2	#Н/Д			

Так как $F > F_{\text{крит}}$ то мы отвергаем гипотезу о неудовлетворительной спецификации модели.
Теперь осуществим F тест в $R\ Studio$.

```
# Оценивание множественной регресси
getwd()
C <-read.table("G_gos_ros.txt", sep="", dec=".", header = TRUE)
C
Cmodel<-lm(data = C, Gt~0+Gt_1+Crt+Sant)
summary(Cmodel)
```

Гипотеза о неудовлетворительной спецификации отвергается, если в протоколе $RStudio$ величина $p-value$ меньше чем 0.05. В нашем случае $p-value: 2.716e-05$. Если же $p-value$ больше чем 0.05, то гипотеза не может быть отвергнута. В данном случае гипотеза отвергается.
В протоколе функции $RStudio$ скоректированный коэффициент детерминации находится в предпоследней строчке протокола справа и в нашем примере он имеет значение 0.9996.
Удалим сомнительную переменная например идикатор санкции.

```
# Шаг 2
Cmodel2<-lm(data = C, Gt~0+Gt_1+Crt)
summary(Cmodel2)
```

После удаления из модели San_t скорректированный коэффициент детерминации уменьшился 0.9996, следовательно переменная НЕ является значащей её МОЖНО удалять из модели.
Удалим следующую сомнительную переменную идикатор кризиса.

```
# Шаг 3
Cmodel3<-lm(data = C, Gt~0+Gt_1+Sant)
summary(Cmodel3)
```

После удаления из модели Cr_t скорректированный коэффициент детерминации увеличился 0.9996, следовательно переменная НЕ является значащей её МОЖНО удалять из модели.

Эконометрика. Домашняя работа № 11
Аверьянов Тимофей ПМ 3-1

Задача №1. Вывести формулу (*) и (**):

$$E\left(\triangle\widetilde{y}_0\right) = 0 \qquad (*)$$

$$Var\left(\triangle\widetilde{y}_0\right) = E\left(\triangle\widetilde{y}_0^2\right) = \sigma_u^2\cdot\vec{x}_0^T\cdot Q\cdot\vec{x}_0 + \sigma_u^2 = \sigma_u^2(1+q_0) \qquad (**)$$

Решение:

$$E\left(\triangle\widetilde{y}_0\right) =$$

$$= \left[\text{мы знаем, что: } \widetilde{y}_0 = \widetilde{a}_0 + \widetilde{a}_1\cdot x_{1,0} + \ldots + \widetilde{a}_k\cdot x_{k,0} = \vec{\widetilde{a}}^T\cdot\vec{x}_0\right] =$$

$$= E\left(\triangle\widetilde{y}_0\right) = E\left(\widetilde{y}_0\right) - E(y_0) = E\left(\widetilde{y}_0\right) - a_0 - a_1\cdot x_{1,0} - \ldots - a_k\cdot x_{k,0} =$$

$$= E\left(\widetilde{y}_0\right) - \vec{a}^T\cdot\vec{x}_0 = \vec{\widetilde{a}}^T\cdot\vec{x}_0 - \vec{a}^T\cdot\vec{x}_0 = \left(\vec{\widetilde{a}} - \vec{a}\right)^T\cdot\vec{x}_0 =$$

[в силу свойства несмещенности оценок коэффициентов $E\left(\vec{a}^T\right) = \vec{\widetilde{a}}$ математическое ожидание ошибки $E\left(\triangle\widetilde{y}_0\right) = 0$] =

$$= \left(\vec{\widetilde{a}} - \vec{\widetilde{a}}\right)^T\cdot\vec{x}_0 = 0 \ \blacksquare \qquad (*')$$

$$Var\left(\triangle\widetilde{y}_0\right) = E\left\{\left(\widetilde{y}_0 - y_0\right)^2\right\} = E\left\{\left[\vec{\widetilde{a}}^T\cdot\vec{x}_0 + u_t - \vec{a}^T\cdot\vec{x}_0\right]^T\cdot\left[\vec{\widetilde{a}}^T\cdot\vec{x}_0 + u_t - \vec{a}^T\cdot\vec{x}_0\right]\right\} =$$

$$= E\left\{\left[\left(\vec{\widetilde{a}} - \vec{a}\right)^T\cdot\vec{x}_0 + u_t\right]^T\cdot\left[\left(\vec{\widetilde{a}} - \vec{a}\right)^T\cdot\vec{x}_0 + u_t\right]\right\} =$$

$$= E\left\{u_t^2\right\} + \vec{x}_0^T\cdot E\left\{\left(\vec{a} - \vec{\widetilde{a}}\right)^T\cdot\left(\vec{a} - \vec{\widetilde{a}}\right)\right\}\cdot\vec{x}_0 + 2\cdot\vec{x}_0^T E\left\{\left(\vec{a} - \vec{\widetilde{a}}\right)\cdot u_t\right\} =$$

$$= E\left\{u_t^2\right\} + \sigma_u^2\cdot\vec{x}_0^T\left(X^T\cdot X\right)^{-1}\cdot\vec{x}_0 + 2\cdot\vec{x}_0^T E\{\vec{a}\cdot u_t\} =$$

$$= \left[\sigma_u^2 = E\left\{u_t^2\right\}; \vec{a} \text{ и } u_t \text{ некоррелированы}\right] =$$

$$= \sigma_u^2 + \sigma_u^2\cdot\vec{x}_0^T\left(X^T\cdot X\right)^{-1}\cdot\vec{x}_0 = \left[Q = \left(X^T\cdot X\right)^{-1}\right] = \sigma_u^2\left(1 + \vec{x}_0^T Q\cdot\vec{x}_0\right)$$

$$= \left[q_0 = \vec{x}_0^T\cdot Q\cdot\vec{x}_0\right] = \sigma_u^2(1 + q_0) \ \blacksquare \qquad (**')$$

Задача №2. Вычислить прогнозы гос. расходов и инвестиций по моделе Самуэльсона-Хикса по контролирующей выборке.

Решение:

Вычислим прогноз для инвестиций в моделе Самуэльсона-Хикса. Нам потребуются значения объясняющих переменных:

		ΔY_{2017}	Cr_{2017}	San_{2017}
$x_0^T=$	1	662.6057	0	1
$I^P_{2018}=$	7739.334			
$I_{2018}=$	8393.411			

Шаг 1. Как мы видим наш прогноз отличается от реального значения, поэтому вычислим стандартную ошибку прогноза. Формируем матрицу X у которой 14 строк и 4 столбца:

Шаг №1				
X	1	1992.6	0	0
	1	2102.9	0	0
	1	2002.7	0	0
	1	2724.1	0	0
	1	3084.1	0	0
	1	2058.1	1	0
	1	-3228.2	0	0
	1	1713.6	0	0
	1	1695.6	0	0
	1	1515.7	0	0
	1	767.2041	0	0
	1	323.0933	0	1
	1	-1118.52	0	1
	1	-74.1368	0	1

И транспонированную к ней X^T.

X^T													
1	1	1	1	1	1	1	1	1	1	1	1	1	1
1992.6	2102.9	2002.7	2724.1	3084.1	2058.1	-3228.2	1713.6	1695.6	1515.7	767.204111	323.0933	-1118.52	-74.1368
0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	1	1

Шаг 2. Вычислим матрицу $X^T \cdot X$.

Шаг №2				
X^TX	14	15558.84425	1	3
	15558.84425	54051292.44	2058.1	-869.559861
	1	2058.1	1	0
	3	-869.559861	0	3

Шаг 3. Рассчитаем матицу Q.

Шаг №3				
$Q = (X^TX)^{-1}$	0.171423338	-4.9702E-05	-0.06913158	-0.18582964
	-4.9702E-05	3.45866E-08	-2.1481E-05	5.9727E-05
	-0.069131577	-2.14807E-05	1.113341011	0.06290532
	-0.185829637	5.97271E-05	0.062905324	0.53647506

Шаг 4. Рассчитаем $q_0 = \vec{x}_0^T \cdot Q \cdot \vec{x}_0$

	Шаг №4						
x_0	1		x_0^T*Q	-0.047339153	3.29423E-05	-0.020459489	0.390220934
	662.6057386						
	0						
	1		$q_0 = x_0^T*Q*x_0$	0.364709568			

Шаг 5. Рассчитаем значение стандартной ошибки.

Шаг №5	
SI^{P}_{2018}	1633.223948

Вычислим прогноз для государственных расходов в моделе Самуэльсона-Хикса.

	G_{2017}	Cr_{2017}	San_{2017}
x_0^{T}=	7264.272	0	1
G^{P}_{2018}=	**7218.285**		
G_{2018}=	**7326.382**		

Шаг 1. Как мы видим наш прогноз отличается от реального значения, поэтому вычислим стандартную ошибку прогноза. Формируем матрицу X у которой 15 строк и 3 столбца:

Шаг №1			
X	6390	0	0
	6540.2	0	0
	6679	0	0
	6775.3	0	0
	6931.9	0	0
	7120.7	0	0
	7359.9	1	0
	7314.5	0	0
	7205.7	0	0
	7306.7	0	0
	7498.7	0	0
	7562.671	0	0
	7401.995	0	1
	7170.733	0	1
	7238.265	0	1

И транспонированную к ней X^T.

X^T														
6390	6540.2	6679	6775.3	6931.9	7120.7	7359.9	7314.5	7205.7	7306.7	7498.7	7562.671176	7401.99513	7170.733	7238.265
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	1	1	1

Шаг 2. Вычислим матрицу $X^T \cdot X$.

Шаг №2			
X^TX	757881587.6	7359.9	21810.993
	7359.9	1	0
	21810.99298	0	3

Шаг 3. Рассчитаем матицу Q.

Шаг №3			
$Q = (X^TX)^{-1}$	1.83439E-09	-1.35009E-05	-1.33366E-05
	-1.35009E-05	1.099365477	0.098156212
	-1.33366E-05	0.098156212	0.430294997

Шаг 4. Рассчитаем $q_0 = \vec{x}_0^T \cdot Q \cdot \vec{x}_0$.

	Шаг №4					
x_0	7264.272		x_0^T*Q	-1.11147E-08	8.1803E-05	0.333414141
	0					
	1		$q_0 = x_0^T*Q*x_0$	0.333333401		

Шаг 5. Рассчитаем значение стандартной ошибки.

Шаг №5	
SG^p_{2018}	154.934121

Эконометрика. Домашняя работа № 14
Аверьянов Тимофей ПМ 3-1

Задача № 1. Исследовать гипотезу о гомоскедастичности в оригинальной модели расходов домохозяйств России.

Решение:

Оригинальная модель расходов домохозяйств России выглядит следующим образом:

$$\begin{cases} C_t = a_0 + a_1 Y_{t-1} + u_t; \\ E(u_t) = 0; Var(u_t) = \sigma_u^2; \end{cases} \qquad (*)$$

Исследуем гипотезу о гомоскедастичности,

$$Var(u_1) = Var(u_2) = \ldots = Var(u_n)$$

для выполним следующий код:

```
library(ggplot2)
library(lmtest)
library(dplyr)
library(tseries)

C<-read.table("dataRStudio.txt", sep="", dec=".", header = TRUE)
C

Cmodel<-lm(data = C, Ct~Yt)
summary(Cmodel)

# тест Голдфилда-Кванта
gqtest(Cmodel, fraction=0.33, data=C, order.by=C["Yt"])
```

```
	Goldfeld-Quandt test

data:  Cmodel
GQ = 88.367, df1 = 4, df2 = 3, p-value = 0.00192
alternative hypothesis: variance increases from segment 1 to 2
```

По скольку величина `p-value` меньше чем 0.05, то гипотеза о гомоскедастичности случайного возмущения отклоняется. Следовательно, в оригинальной моделе расходов домохозяйств России случайные возмущения являются гетероскедастичными.

Задача № 2. Осуществить исследование в R об отсутсвии гетерскедастичности и отсутвия автокорреляции созданной в домашнем творческом задании.

Решение:

$$
\begin{cases}
C_t \;=\; a_0 \;+\; a_1 \;\cdot Y_{t-1} \;+\; a_2 \;\cdot Cr_t \;+\; a_3 \;\cdot San_t + a_4 \cdot Cost_t + \; u_t; \\
I_t \;=\; b_0 \;+ b_1 \;\cdot\; (Y_{t-1} \;-\; Y_{t-2}) \;+\; b_2 \;\cdot Cr_t \;+\; b_3 \;\cdot San_t + b_4 \cdot Cost_t \;+\; v_t; \\
G_t \;=\; g_1 \;\cdot\; G_{t-1} \;+\; g_2 \;\cdot Cr_t \;+\; g_3 \;\cdot San_t + g_4 \cdot Cost_t + \; w_t; \\
Y_t \;=\; C_t \;+\; I_t \;+\; G_t;
\end{cases}
\qquad (**)
$$

Проведём исследование на гомоскедастичность случайных возмущений *уровня расходов домозяйств* в модифицированной модели Самуэльсона-Хикса (**), а так же проведём тест Дарбина-Уотсона об отсутвии автокорреляции случайных возмущений:

```r
library(ggplot2)
library(lmtest)
library(dplyr)
library(tseries)

C <-read.table("data1.txt", sep="", dec=".", header = TRUE)
C

Cmodel<-lm(data = C, Ct~Yt_1+Crt+Sant+Costt)
summary(Cmodel)

# тест Голдфилда-Кванта
gqtest(Cmodel, fraction=0.2, data=C, order.by = C["Yt_1"]) # проведем GQ тест выкинув посередине 20% наблюдений
# тест Дарбина-Уотсона
dwtest(Cmodel, alternative = c("greate"))
```

```
        Goldfeld-Quandt test

data:  Cmodel
GQ = 2.3157, df1 = 1, df2 = 1, p-value = 0.3701
alternative hypothesis: variance increases from segment 1 to 2

        Durbin-Watson test

data:  Cmodel
DW = 2.4744, p-value = 0.5787
alternative hypothesis: true autocorrelation is greater than 0
```

Вывод: Получаем, что фрагмент уровня расходов домохозяйств модифицированной модели Самуэльсона-Хикса является гомоскедастичным, а так как в тесте Дарбина-Уотсона `p-value` > 0.05, то гипотеза H_0 об отсутсвии автокорреляции считается справедливой. Что совпало с ДТЗ.

Так как в инвестиционной модели $k = 14$, а у фрагмента государственных расходов отсутсвует свободный член, то проводить тест Дарбина-Уотсона бессмысленно. Исходя из этого проведём для оставшихся двух фрагментов исследование на гомоскедастичность случайных возмущений.

Для *инвестиционного* фрагмента:

```
C <-read.table("data2.txt", sep="", dec=".", header = TRUE)
C

Cmodel<-lm(data = C, It~deltaYt_1+Crt+Sant+Costt)
summary(Cmodel)

# тест Голдфилда-Кванта
gqtest(Cmodel, fraction=0.13, data=C, order.by = C["deltaYt_1"]) # проведем GQ тест выкинув посередине 13% наблюдений
```

```
        Goldfeld-Quandt test

data:  Cmodel
GQ = 227, df1 = 2, df2 = 1, p-value = 0.04688
alternative hypothesis: variance increases from segment 1 to 2
```

Вывод: $\texttt{p-value} < 0.05$ следовательно гипотеза о гомоскедастичности случайных возмущений отклоняется в пользу гипотезы о гетероскедастичности случайных возмущений.

Для фрагмента *госудасвенных расходов*:

```
C <-read.table("data3.txt", sep="", dec=".", header = TRUE)
C

Cmodel<-lm(data = C, Gt~0+Gt_1+Crt+Sant+Costt)
summary(Cmodel)

# тест Голдфилда-Кванта
gqtest(Cmodel, fraction=0.33, data=C, order.by=C["Gt_1"]) # проведем GQ тест выкинув посередине 33% наблюдений
```

```
        Goldfeld-Quandt test
```

```
data:  Cmodel
GQ = 2.6397, df1 = 2, df2 = 1, p-value = 0.3991
alternative hypothesis: variance increases from segment 1 to 2
```

Вывод: `p-value` > 0.05 следовательно гипотеза о гомоскедастичности случайных возмущений принимается.

Эконометрика. Домашняя работа № 15
Аверьянов Тимофей ПМ 3-1

Задача №1. По статистическим данным оценить модель со спецификацией (6), не предполагая справедливость равенства (8).

$$\begin{cases} \underbrace{\ln Y}_{y} = \underbrace{\ln A}_{a_0} + \alpha \cdot \underbrace{\ln K}_{x_1} + \beta \cdot \underbrace{\ln L}_{x_2} + u; \\ E(u) = 0; Var(u) = \sigma_u^2; \end{cases} \tag{6}$$

$$\alpha + \beta = 1 \tag{8}$$

Решение:

Шаг 1. Составляем спецификацию модели (6).

Шаг 2. Переносим статистические данные в Excel.

Год	Y(млрд. рублей)	K (млрд. рублей)	L(млн. чел.)
2000	7305	4306	65.273
2001	7678	4979	65.124
2002	8042	5518	66.266
2003	8633	6640	67.152
2004	9250	8103	67.134
2005	9839	9207	68.603
2006	10597	11277	69.157
2007	11455	13403	70.814
2008	12097	15527	70.603
2009	11205	13100	69.41
2010	11652	14444	69.934
2011	12212	16120	70.857

Шаг 3. Готовим уравнения наблюдей в качестве уравнений (6).

Год	LN(Y)	LN(K)	LN(L)
2000	8.896314324	8.367764678	4.17857847
2001	8.946114376	8.512984347	4.17629314
2002	8.992433087	8.615770755	4.19367695
2003	9.063347348	8.800867242	4.20695871
2004	9.132378831	8.999989642	4.20669062
2005	9.194109359	9.127719343	4.22833627
2006	9.268326221	9.330520532	4.23637928
2007	9.346181595	9.503233841	4.26005672
2008	9.400712767	9.650335723	4.25707264
2009	9.324115386	9.480367509	4.24003095
2010	9.363233118	9.578034382	4.24755194
2011	9.410174354	9.687816016	4.26066376

Шаг 4. Обращаемся к функции ЛИНЕЙН метода наименьших квадратов.

beta	alpha	ln(A)
0.697474	**0.347372645**	**3.074134087**
0.109976	0.007244649	0.399918086
0.999879	**0.002248364**	#Н/Д
37112.98	9	#Н/Д
0.375223	4.54963E-05	#Н/Д

Шаг 5. Записываем оценённую модель (6).

$$\begin{cases} \ln Y = 3.0741 + 0.3474 \cdot \ln K + 0.6975 \cdot \ln L + u; \\ E(u) = 0; Var(u) = \sigma_u^2; \end{cases} \tag{*}$$

Шаг 6. От трансформированной модели (*) возвращаемся к модели (**):

$$\begin{cases} Y = A \cdot K^{\alpha} \cdot L^{\beta} + u; \\ E(u) = 0; Var(u) = \sigma_u^2; \end{cases} \tag{**}$$

для этого найдём оценки A при помощи операции потенцирования:

A=	**21.63114**

$A = 21.63114$, записываем оценённую модель:

$$\begin{cases} Y = 21.63114 \cdot K^{0.3474} \cdot L^{0.6975} + u; \\ E(u) = 0; Var(u) = \sigma_u^2; \end{cases} \tag{**'}$$

Эконометрика. Домашняя работа № 16
Аверьянов Тимофей ПМ 3-1

Задача №1. Исследовать на нормальность модель полученную в ДТ3.

Решение:

$$\begin{cases} C_t \ = \ a_0 \ + \ a_1 \ \cdot Y_{t-1} \ + \ a_2 \ \cdot Cr_t \ + \ a_3 \ \cdot San_t + a_4 \cdot Cost_t + \ u_t; \\ I_t \ = \ b_0 \ + b_1 \ \cdot \ (Y_{t-1} \ - \ Y_{t-2}) \ + \ b_2 \ \cdot Cr_t \ + \ b_3 \ \cdot San_t + b_4 \cdot Cost_t \ + \ v_t; \\ G_t \ = \ g_1 \ \cdot \ G_{t-1} \ + \ g_2 \ \cdot Cr_t \ + \ g_3 \ \cdot San_t + g_4 \cdot Cost_t + \ w_t; \\ Y_t \ = \ C_t \ + \ I_t \ + \ G_t; \end{cases} \qquad (*)$$

Проведём тест Харке — Бера для модели государственных расходов домохозяйств модифицированной модели Самуэльсона-Хикса (*). В этом тесте исследуется гипотеза о том, что в уравнениях наблюдений объекта случайные возмущения имеют нормальный закон распределения с 0 математическим ожиданием с одной и тойже дисперсией и являются независимыми случайными переменными. Тест состоит из следующих шагов:

Шаг 1. Модель оценивается методом наименьших квадратов и вычисляются оценки случайных возмущений (остатки). Что было сделано в ДТ3.

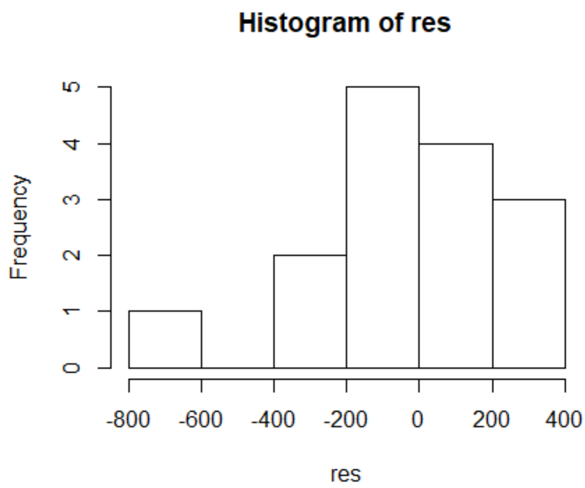
Шаг 2. Воспользуемся функцией `jarque.bera.test` и выполним следующие действия:

```r
library(ggplot2)
library(lmtest)
library(dplyr)
library(tseries)

C <-read.table("data1.txt", sep="", dec=".", header = TRUE)
C
Cmodel<-lm(data = C, Ct~Yt_1+Crt+Sant+Costt)
summary(Cmodel)

# тестирование нормального закона
# распределения случайного возмущения
res<-residuals(Cmodel)
jarque.bera.test(res)
hist(res)
```

```
Output:
        Jarque Bera Test
data:  res
X-squared = 1.5789, df = 2, p-value = 0.4541
```

Поскольку `p-value` больше чем 0.05, то гипотеза о нормальном распределении принимается.

Используя, функцию `hist` построим гистограмму:



Аналогично, выполним действия для уровня инвестиций модифицированной модели Самуэльсона-Хикса (*).

Шаг 1. Модель оценивается методом наименьших квадратов и вычисляются оценки случайных возмущений (остатки). Что было сделано в ДТЗ.

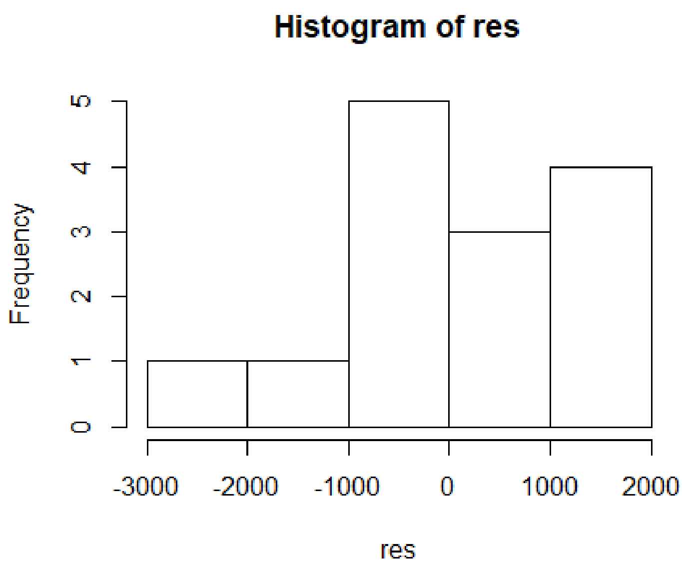
Шаг 2. Воспользуемся функцией `jarque.bera.test` и выполним следующие действия:

```
C <-read.table("data2.txt", sep="", dec=".", header = TRUE)
C

Cmodel<-lm(data = C, It~deltaYt_1+Crt+Sant+Costt)
summary(Cmodel)


# тестирование нормального закона
# распределения случайного возмущения
res<-residuals(Cmodel)
jarque.bera.test(res)
hist(res)
```

```
Output:
        Jarque Bera Test
data:  res
X-squared = 1.1163, df = 2, p-value = 0.5723
```

Поскольку `p-value` больше чем 0.05, то гипотеза о нормальном распределении принимается.

Используя, функцию `hist` построим гистограмму:



И наконец проведём тест Харке — Бера для государственных расходов модифицированной модели Самуэльсона-Хикса (*):

Шаг 1. Модель оценивается методом наименьших квадратов и вычисляются оценки случайных возмущений (остатки). Что было сделано в ДТЗ.

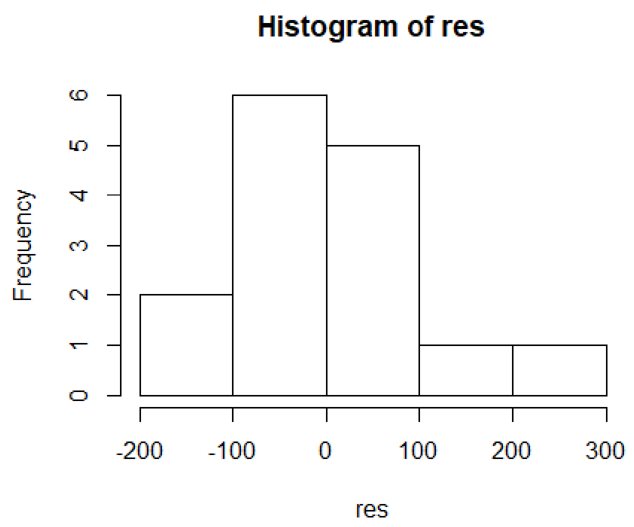
Шаг 2. Воспользуемся функцией `jarque.bera.test` и выполним следующие действия:

```
C <-read.table("data3.txt", sep="", dec=".", header = TRUE)
C

Cmodel<-lm(data = C, Gt~0+Gt_1+Crt+Sant+Costt)
summary(Cmodel)

# тестирование нормального закона
# распределения случайного возмущения
res<-residuals(Cmodel)
jarque.bera.test(res)
hist(res)
```

```
Output:
        Jarque Bera Test
data:  res
X-squared = 0.43314, df = 2, p-value = 0.8053
```

Поскольку `p-value` больше чем 0.05, то гипотеза о нормальном распределении принимается.

Используя, функцию `hist` построим гистограмму:


Histogram of res
Frequency
0
1
2
3
4
5
6
-200
-100
0
100
200
300
res