

Disclaimer

Этот документ является повторением тем, тестов и проблем курса эконометрики 1, «всё в одном и под рукой», является стремлением автора к обобщению материала курса, и содержит в некоторых местах субъективное мнение и понимание автора. Данный документ может и не покрывать полностью всю программу курса, но при соответствующем feedback'е может рано или поздно её покрыть.

Этот документ ни в коем случае нельзя считать самостоятельным пособием для изучения курса, он таковым не задумывался и таковым не будет являться никогда.

Самое лучшее пособие – это лекции, семинары, конспекты, литература и собственная голова.

За основу взяты:

- лекции **Ильи Борисовича Воскобойникова**
- семинары **Андрея Александровича Мамонтова**
- книги
D.N. Gujarati, C. Dougherty и G.S. Maddala.

Следующие замечательные люди заслуживают особой благодарности за ценные замечания и дополнения:

- **Тлехугов Николай Владимирович**
- **Дмитриев Александр Сергеевич**

Теорема Гаусса-Маркова

Если выборочные данные обладают следующими свойствами:

$$\text{Модель: } Y_i = \alpha + \beta X_i + \varepsilon_i$$

1 - Модель данных правильно специфицирована.

1.1 - Модель линейна по параметрам.

1.2 - В модели имеется неслучайная часть и случайная часть, входящая аддитивно.

1.3 - Модель адекватна устройству данных.

2 - Определяющие переменные X не равны между собой и могут быть как стохастическими, так и нет. В случае стохастических регрессоров, регрессоры и ошибки должны быть некоррелированными.

3 - Ошибки не имеют систематический характер $E(\varepsilon_i) = 0; \forall i = 1 \dots n$

4 - Дисперсии всех ошибок равны между собой $Var(\varepsilon_i) = \sigma^2; \forall i = 1 \dots n$

5 - Все ошибки не коррелируют между собой $Cov(\varepsilon_i; \varepsilon_j) = 0; \forall i, j = 1 \dots n; i \neq j$

То оценки МНК (OLS) состоятельны и эффективны в классе линейных несмещённых оценок.

Общие положения

Классическая регрессия в матричной форме со свободным членом и с выполненными предпосылками ТГМ имеет следующий вид:

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}_{n \times 1} \quad X = \begin{pmatrix} 1 & X_{1,1} & X_{2,1} & \dots & X_{k,1} \\ 1 & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1,n} & X_{2,n} & \dots & X_{k,n} \end{pmatrix}_{n \times (k+1)} \quad \beta = \begin{pmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}_{(k+1) \times 1} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}_{n \times 1} \quad \varepsilon_i \sim iid N(0; \sigma^2)$$

$$\text{Model: } Y = X \cdot \beta + \varepsilon$$

$$\hat{\beta} = (X^T \times X)^{-1} \times (X^T \times Y) \quad \hat{Y} = X \cdot \hat{\beta} \quad \hat{\varepsilon} = Y - \hat{Y} = Y - X \cdot \hat{\beta} \quad (\hat{\sigma}^2) = \frac{(\hat{\varepsilon})^T \times \hat{\varepsilon}}{n - (k + 1)}$$

Регрессии со свободным членом по методу OLS порождает несколько «приятных» свойств:

Во-первых:

$$\left. \begin{aligned} RSS &= (\hat{\varepsilon})^T \times \hat{\varepsilon} = \sum_{i=1}^n \left(Y_i - (\alpha + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \dots + \beta_k X_{k,i}) \right)^2 \\ ESS &= \sum_{i=1}^n \left((\alpha + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \dots + \beta_k X_{k,i}) - \bar{Y} \right)^2 \\ TSS &= \sum_{i=1}^n (Y_i - \bar{Y})^2 \end{aligned} \right\} \Rightarrow TSS = ESS + RSS$$

Во-вторых, в силу предпосылок 3, 4 и 5 ТГМ:

$$\begin{aligned} E(\hat{\beta}) &= (X^T \times X)^{-1} \times (X^T \times E(Y)) = (X^T \times X)^{-1} \times (X^T \times X \cdot \beta) = \beta \\ Var(\hat{\beta}) &= \left((X^T \times X)^{-1} \times X^T \right) \times Var(Y) \times \left((X^T \times X)^{-1} \times X^T \right)^T = \\ &= (X^T \times X)^{-1} \times X^T \times (\sigma^2 \cdot I) \times X \times (X^T \times X)^{-1} = (X^T \times X)^{-1} \cdot \sigma^2 \end{aligned}$$

где $Var(\hat{\beta})$ - ковариационная матрица оценок $\hat{\beta}$.

Из этого следует, если предположить «нормальность» остатков, что

$$\hat{\beta} \sim N \left(\begin{pmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}; (X^T \times X)^{-1} \cdot \sigma^2 \right)$$

где N - $(k+1)$ -мерное нормальное распределение.

В-третьих, если предположить «нормальность» остатков, получаем:

$$\frac{\hat{\sigma}^2}{\sigma^2} \times (n - (k + 1)) \sim \chi^2_{n - (k + 1)}$$

из чего можно сделать необходимые расчёты доверительных интервалов дисперсии (например).

Нам известно, что оценки параметров состоятельны, эффективны и, главное, **не смещены**.

Таким образом, можно считать $\hat{\sigma}^2$ хорошей оценкой дисперсии регрессии и использовать её в расчётах.

Оценённая ковариационная матрица оценок $\hat{\beta}$ равна: $Var^{\wedge}(\hat{\beta}) = (X^T \times X)^{-1} \cdot \hat{\sigma}^2$.

$$\begin{aligned} R^2 &= \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} & R^2_{adj} &= 1 - \frac{RSS / n - (k + 1)}{TSS / n - 1} \\ \text{Далее} & & seREGR &= \sqrt{\frac{RSS}{n - (k + 1)}} = \sqrt{(\hat{\sigma}^2)} \\ (\hat{\sigma}^2) &= \frac{RSS}{n - (k + 1)} \end{aligned}$$

Все эти тесты можно отнести к первой предпосылки ТГМ: правильная спецификация модели. Эти тесты позволяют определить на основе выборочных данных необходимость наличия или отсутствия параметров или групп параметров, функциональных преобразований данных и применения dummy переменных. Все распределения справедливы только в случае «нормальности» остатков, в противном случае присутствуют только асимптотические распределения.

Remark: Распределения всех статистик справедливы только в случае, если правдоподобна гипотеза H_0 - поэтому это и называется проверкой гипотез. Если наша расчётная статистика лежит вне критических значений соответствующих распределений для соответствующего уровня значимости, то считаем, что наше рассчитанное значение статистики неправдоподобно высоко или низко для гипотезы H_0 , и отклоняем её.

Тест значимости отдельного коэффициента

$$H_0 : \beta_m = 0 \quad H_1 : \beta_m \neq 0$$

1 - Рассчитываем исходную регрессию и получаем всевозможные оценки параметров, их дисперсии и ковариации.

2 - Рассчитываем следующую статистику:

$$\frac{\hat{\beta}_m - 0}{\sqrt{\text{Var}(\hat{\beta}_m)}} \sim t_{n-(k+1)} \quad \text{given } H_0 = \text{true}$$

3 - Получаем либо P_v , либо критические значения t распределения для нужного уровня значимости.

4 - Принимаем решение: если уровень значимости больше P_v , или рассчитанное значение лежит вне пределов критических значений, то отвергаем H_0 .

Этот тест по результатам совпадает с тестом значимости линейных ограничений для одного линейного ограничения одного отдельного параметра.

Тест значимости регрессии

(Goodness of Fit Test)

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0 \quad H_1 : \exists j \geq 1 \text{ s.t. } \beta_j \neq 0$$

или $H_0 : R^2 = 0 \quad H_1 : R^2 \neq 0$ (условная запись).

1 - Рассчитываем исходную регрессию и получаем всевозможные оценки параметров, их дисперсии и ковариации.

2 - Рассчитываем следующую статистику:

$$\frac{\text{ESS} / (k+1) - 1}{\text{RSS} / n - (k+1)} \sim F((k+1) - 1; n - (k+1)) \quad \text{given } H_0 = \text{true}$$

3 - Получаем либо P_v , либо критические значения F распределения для нужного уровня значимости.

4 - Принимаем решение: если уровень значимости больше P_v , или рассчитанное значение лежит вне пределов критических значений, то отвергаем H_0 .

Этот тест является частным случаем следующего теста для k линейных ограничений на все параметры $\beta_1; \dots; \beta_k$.

Тест значимости линейных ограничений

(я его называю **Improvement of Fit Test**)

H_0 : m линейных ограничений значимы; H_1 : иначе

1 - Рассчитываем исходную (без ограничений) регрессию и получаем RSS_{UR} .

2 - Рассчитываем регрессию с m независимыми линейными ограничениями на параметры $\beta_1; \dots; \beta_k$ и получаем RSS_R .

3 - Рассчитываем следующую статистику:

$$\frac{(RSS_R - RSS_{UR})/m}{RSS_{UR}/(n - (k + 1))} \sim F(m; n - (k + 1)) \quad \text{given } H_0 = \text{true}$$

4 - Получаем либо P_v , либо критические значения F распределения для нужного уровня значимости.

5 - Принимаем решение: если уровень значимости больше P_v , или рассчитанное значение лежит вне пределов критических значений, то отвергаем H_0 .

Иными словами, если расхождения в RSS 'ах велики, то какие-то линейные ограничения, скорее всего, не верны (не эффективны), следовательно, отвергаем H_0 . А если расхождения статистически не значимы, то с вероятностью P_v , скорее всего, эти m линейных ограничений эффективны, следовательно, НЕ отвергаем H_0 .

Тест стабильности структуры выборки или параметров

(**Chow Test**)

Remark: Иными словами, тест на отсутствие необходимости использования dummy переменных, для разграничения выборок, иными словами, для разграничения линейных зависимостей в разных выборках.

Итак, пусть имеется две (ну или более) выборки (можно разбить одну выборку на две части). Зададимся вопросом, равны ли структурно эти выборки, или, что аналогично, стабильны ли параметры в моделях по этим двум (нескольким) выборкам, или, что опять-таки аналогично, но есть сама походу оригинальная идея Chow, следует ли вводить dummy переменные для разграничения данных из одной выборки от данных из другой (третьей, пятой, m 'ой, $(m+1)$ 'ой и т.д.)?

Следует отметить следующее: если каким-то **единичным** данным из общей выборки соответствуют **индивидуальные** dummy, то RSS регрессии с такими dummy будет равен RSS 'у регрессии посчитанной по выборке без этих наблюдений (это, кстати, и есть underlying принцип в тесте на устойчивость предсказании, который мы рассмотрим далее по документу).

Предпосылки: остатки во всех выборках «нормальны» [$\sim N(...; ...)$] и не зависимы друг от друга как между собой в пределах выборок, так и между выборками.

Рассмотрим случай с m выборками размера. Можно ли их «смешать», или всё-таки ввести dummy для каждой? Для начала совместим все выборки, и полученную выборку назовём P .

H_0 : $\beta_{1,1} = \beta_{2,1} = \dots = \beta_{m,1}; \dots; \beta_{1,k} = \beta_{2,k} = \dots = \beta_{m,k}$; H_1 : иначе

Словами интерпретировать можно так:

H_0 : зависимость везде линейная одинаковая

H_1 : зависимость у каждой выборки своя (своя, но линейная)

1 - Рассчитываем регрессию, приняв гипотезу H_0 , и получаем RSS_P . То есть, оцениваем регрессию вида

$$Y^{<P>}_i = \alpha + \beta_1 X^{<P>}_{1,i} + \beta_2 X^{<P>}_{2,i} + \dots + \beta_k X^{<P>}_{k,i} + \varepsilon_i$$

по совмещённой выборке Р.

2 - Оцениваем регрессию по каждой выборке отдельно, используя модель, и получаем RSS_m , m - номер выборки,

$$Y^{<m>}_i = \alpha + \beta_{m,1} X^{<m>}_{1,i} + \beta_{m,2} X^{<m>}_{2,i} + \dots + \beta_{m,k} X^{<m>}_{k,i} + \varepsilon_i$$

3 - Рассчитываем следующую статистику:

$$\frac{(RSS_p - (RSS_1 + RSS_2 + \dots + RSS_m))}{\frac{(n_1 + n_2 + \dots + n_m - (k+1)) - ((n_1 + n_2 + \dots + n_m) - m \times (k+1))}{RSS_1 + RSS_2 + \dots + RSS_m}} \sim F((m-1) \times (k+1); ((n_1 + n_2 + \dots + n_m) - m \times (k+1)))$$

given $H_0 = true$

4 - Получаем либо P_v , либо критические значения F распределения для нужного уровня значимости.

5 - Принимаем решение: если уровень значимости больше P_v , или рассчитанное значение лежит вне пределов критических значений, то отвергаем H_0 .

Каким образом это связано с dummy? Таким, что, к примеру, в случае с двумя выборками коэффициентами при введённой неким образом Dummy будут поправки (или разницы) к соответствующим коэффициентам некой базовой, присутствующей во всех выборках, модели. Не отвергая гипотезу H_0 , мы говорим, что в этих поправках нет необходимости, и их можно исключить из спецификации модели. Таким, образом, тестом Chow, можно проверить следует ли вводить dummy или нет, то есть, стоит ли отличать выборки или нет (вспомните ДЗ№3 и ДЗ№4).

Тест Предсказания Чоу

(Chow Forecast Test)

Этот тест помогает нам посмотреть, одинакова ли модель для всей выборки размера N, или выборку можно «разбить» на две различные по структуре части. Итак, приступим. Разобьём нашу выборку на две части, I с M первыми наблюдениями и II с последними (N-M) наблюдениями. Ставим вопрос: содержит ли вторая часть выборки хоть какую-нибудь линейную зависимость, похожую на зависимость из первой части, или наблюдения во второй части, настолько плохо себя ведут, что нам придётся всем (N-M) наблюдениям присваивать индивидуальные dummy переменные? Выходит, что нам необходимо проверить гипотезу о равенстве нулю этих (N-M) dummy переменных при наблюдениях из второй части выборки. Ну а для этого мы можем использовать тест значимости линейных ограничений, описанный выше. Но, подождите, вдруг удастся его оптимизировать? Итак, рассмотрим $RSS_{M+(N-M)}^D = RSS_N^D - RSS$ нашей исходной модели с (K+1) параметрами, но дополненной индивидуальными dummy переменными для каждого наблюдения из второй части выборки - то есть плюс ещё N-M параметров. Оценки зависимой переменной последних N-M наблюдений будут «идеальными» с той точки зрения, что оценки зависимой переменной будут совпадать абсолютно точно с выборочными данными зависимой переменной, как раз за счёт наличия индивидуальных dummy переменных. Опять-таки за счёт наличия таких dummy, последние N-M наблюдений не будут влиять на оценивание основных параметров (параметров не при dummy, то есть тех параметров, которые описывают зависимость [структуру] первых M наблюдений - поскольку dummy абсолютно точно описывают все

наблюдения из вторых N-M наблюдений). Поэтому, построив регрессию только по M наблюдениям из первой части, мы получим RSS_M , который численно совпадает с RSS_N^D . Преобразуем нужную нам для проверки основной гипотезы статистику:

$$\frac{(RSS_N - RSS_N^D) / (N - M)}{RSS_N^D / N - (k + 1) - (N - M)} = \frac{(RSS_{M+(N-M)} - RSS_M) / (N - M)}{RSS_M / M - (k + 1)}$$

Итак, получившийся алгоритм:

H₀: N-M параметров при dummy равны нулю; **H₁**: иначе.

Иными словами:

H₀: зависимость везде линейная одинаковая

H₁: линейная зависимость есть только в первой части выборки

1 - Рассчитываем исходную (без ограничений) регрессию и получаем RSS_N .

2 - Оценим исходную модель по наблюдениям из первой части выборки, размером M, и получаем RSS_M .

3 - Рассчитываем следующую статистику:

$$\frac{(RSS_{M+(N-M)} - RSS_M) / (N - M)}{RSS_M / M - (k + 1)} \sim F((N - M); M - (k + 1)) \quad \text{given } H_0 = \text{true}$$

4 - Получаем либо P_v, либо критические значения F распределения для нужного уровня значимости.

5 - Принимаем решение: если уровень значимости больше P_v, или рассчитанное значение лежит вне пределов критических значений, то отвергаем H_0 .

Тест о выборе типа функционального преобразования

(Zarembka Scaling, 1968)

В эконометрической теории существует много видов линейных по параметрам уравнений, которые следует использовать, или исходя из теоретических предпосылок, или из предпосылок качества оценивания. Рассмотрим функциональные формы, релевантные к данному тесту.

Lin-Lin (Linear model, Линейная модель):

$$Y_i = \alpha + \beta_1 \cdot X_{1,i} + \beta_k \cdot X_{k,i} + \dots + \beta_k \cdot X_{k,i} + \varepsilon_i$$

$$\beta_m = \frac{\partial Y}{\partial X_m} \quad - \text{ абсолютное изменение } Y \text{ к абсолютному изменению } X_m$$

Semi-Log (Полулогарифмические модели)

1] **Lin-Log**

$$Y_i = \alpha + \beta_1 \cdot \ln(X_{1,i}) + \beta_k \cdot \ln(X_{k,i}) + \dots + \beta_k \cdot \ln(X_{k,i}) + \varepsilon_i$$

$$\beta_m = \frac{\partial Y}{\partial (\ln(X_m))} = \frac{\partial Y}{\partial X_m / X_m} \quad - \text{ абсолютное изменение } Y \text{ к относительному изменению } X_m$$

2] **Log-Lin**

$$\ln(Y_i) = \alpha + \beta_1 \cdot X_{1,i} + \beta_k \cdot X_{k,i} + \dots + \beta_k \cdot X_{k,i} + \varepsilon_i$$

$$\beta_m = \frac{\partial (\ln(Y))}{\partial X_m} = \frac{\partial Y / Y}{\partial X_m} \quad - \text{ относительное изменение } Y \text{ к абсолютному изменению } X_m,$$

полуэластичность, «темп роста» Y при изменении X_m .

Log-Log (Log-Linear, Double Log):

$$\ln(Y_i) = \alpha + \beta_1 \cdot \ln(X_{1,i}) + \beta_k \cdot \ln(X_{k,i}) + \dots + \beta_k \cdot \ln(X_{k,i}) + \varepsilon_i$$

$$\beta_m = \frac{\partial(\ln(Y))}{\partial(\ln(X_m))} = \frac{X}{Y} \cdot \frac{\partial Y}{\partial X_m} - \text{эластичность } Y \text{ по } X_m$$

В некоторых случаях (например, в модели типа Cobb-Douglas, или некоторых линейных зверях) бывает интересно и необходимо сравнить и, соответственно, выбрать модель в исходных данных и модель в логарифмированных данных (вспомним ДЗ№3, там модель в log и исходная у меня лично давали ОЧЕНЬ схожие результаты). К сожалению, сам тест и его свойства нетривиальны и требуют чтения оригинальной работы этих авторов. Тест Зарембки рассчитан на сравнение следующих пар моделей: **Log-Log** против **Lin-Log** (сравнение 2), и **Log-Lin** против **Lin-Lin** (сравнение 1). В нём требуется нормальность остатков!

Заметьте $Z_{m,i} = X_{m,i}$ в случае сравнения 1 и $Z_{m,i} = \ln(X_{m,i})$ в случае сравнения 2.

Изложу алгоритм этого теста:

H_0 : нет различий между моделями; H_1 : иначе

1 - Рассчитываем геометрическое среднее зависимых данных:

$$Y^G = \left(\prod_{i=1}^n Y_i \right)^{1/n}$$

2 - Преобразовываем зависимые Y_i следующим образом:

$$\tilde{Y}_i = \frac{Y_i}{Y^G}$$

3 - Строим регрессию вида

$$\tilde{Y}_i = \alpha + \beta_1 Z_{1,i} + \beta_2 Z_{2,i} + \dots + \beta_k Z_{k,i} + \varepsilon_i \text{ и получаем } RSS_1.$$

4 - Строим регрессию вида

$$\ln(\tilde{Y}_i) = \alpha + \beta_1 Z_{1,i} + \beta_2 Z_{2,i} + \dots + \beta_k Z_{k,i} + \varepsilon_i \text{ и получаем } RSS_2.$$

5 - Рассчитываем статистику следующего вида:

$$n/2 \times \left| \ln \left(\frac{RSS_1}{RSS_2} \right) \right| \sim \chi^2 df = 1$$

6 - Получаем либо P_v , либо критические значения χ^2 распределения с одной степенью свободы для нужного уровня значимости.

7 - Принимаем решение: если уровень значимости больше P_v , или рассчитанное значение лежит вне пределов критических значений, то отвергаем H_0 .

Преобразования такого рода как деление на геометрическое среднее в этом случае даёт возможность сравнения RSS' ов этих двух моделей. Если так получилось, что основная гипотеза отвергается, то выбор типа сводится к простому выбору той модели, которая даёт наименьший RSS .

Remark: Этот тест предназначен только для выбора одного из двух типов модели - все промежуточные оценки и значения, кроме RSS рассматривать не рекомендуется и не следует. Важное ограничение: все регрессоры и $Y_i > 0$.

Общий тест типа функционального преобразования Бокса-Кокса

(General Box-Cox Test, 1964)

Этот тест предназначен для выбора типа модели: log-линейной, просто линейной или какой-то средней. Основу теста составляет поиск оценки нужного параметра преобразования модели по методу Максимального Правдоподобия. Но сами Box-Cox

отметили, что могут возникнуть серьёзные трудности при работе с оценкой ММП и предложили альтернативу – поиск оптимального параметра по решётке (Grid Search). Оптимальный параметр – тот, регрессия с использованием которого даёт наименьшее значение RSS. Кстати это найденное значение параметра для данной выборки будет оценкой ММП.

По ходу дела этот тест чисто механический (хотя в G.S.Maddala я нашёл упоминания о двух гипотезах равенства параметра нулю и единице). Вох-Сох предложили следующее непрерывное по параметру преобразование данных:

1 - Рассчитываем геометрическое среднее зависимых данных Y_i :

$$Y^G = \left(\prod_{i=1}^n Y_i \right)^{1/n}$$

2 - Преобразовываем переменные Y_i следующим образом:

$$\tilde{Y}_i = \frac{Y_i}{Y^G}$$

3 - Применяем только к \tilde{Y} следующее преобразование (предполагая, что все переменные больше нуля), а $X_1; X_2; \dots; X_m$ оставляем неизменными, необработанными:

$$y_i(\lambda) = \frac{(\tilde{Y}_i)^\lambda - 1}{\lambda}, \text{ где } \lambda - \text{параметр } (0; 1]. \text{ А при } \lambda \rightarrow 0 \quad y_i(\lambda) \rightarrow \ln(\tilde{Y}_i).$$

Remark: В Johnston указаны другие пределы λ от -1 до 1, не включая нуль. -1 обозначает регрессию $\frac{1}{\tilde{Y}_i}$ по $X_1; X_2; \dots; X_m$.

4 - Проводим Grid Search оптимального λ : каждый раз строим регрессию

$$y_i(\lambda) = \frac{(\tilde{Y}_i)^\lambda - 1}{\lambda} \text{ по } X_1; X_2; \dots; X_m \text{ со свободным членом, для заданного } \lambda \text{ и запоминаем RSS. То есть строим таблицу RSS'ов для каждого значения параметра, те столбец } \lambda \text{ и столбец RSS.}$$

5 - Выбираем такое λ , при котором RSS будет самым наименьшим, дабы не свалиться в локальный минимум (такие случаи бывают). При таком параметре преобразования модель будет очень хорошо описывать имеющиеся у нас данные: дисперсии оценок и регрессии будут наименьшими.

Remark: Тест Zarembka, описанный выше, есть частный случай этого теста при крайних значениях λ , 0 и 1. Кстати, те две гипотезы, о которых упоминалось в G.S. Maddala, как раз и являются гипотезами теста Zarembka. По-моему, предыдущий тест, скорее всего, применяют сначала, и в случае неопределённого результата (log-... и lin-... модели эквивалентны, то есть гипотеза H_0 верна), определяют преобразование данных (тип модели) вот этим самым тестом Вох-Сох.

Remark: Можно сначала применить тест Zarembka или Вох-Сох для необработанных $X_1; X_2; \dots; X_m$ (lin-lin vs log-lin), а потом для $\ln(X_1); \ln(X_2); \dots; \ln(X_m)$ (lin-log vs log-log). Если имеются проблемы с безразличием моделей, то я считаю надо выбрать ту модель, которую потом проще интерпретировать – раз, и с наименьшим RSS – два (так я делал в экзаменационной домашке).

Мультиколлинеарность (МК)

Эта проблема возникает, когда исходные данные достаточно сильно коррелируют между собой, и её результатом является завышение дисперсий оценок и, соответственно, их не стабильность. Сильная корреляция означает, что можно найти такие множители, что минимальное значение линейной комбинации векторов исходных данных будет не нуль, но близко к нулю (сравнительно). Таким образом, определитель $\Delta = \det(X^T \times X)$, может принять близкое к нулю значение, и элементы обратной матрицы $(X^T \times X)^{-1}$, определяющей как раз и сами оценки параметров и их дисперсий, будут достаточно большими (по модулю).

Увидеть на практике присутствие МК можно по следующим её последствиям:

1. Числа в ковариационной матрице оценок параметров достаточно велики, и соответственно, сами оценки не стабильны.
2. Незначимость многих параметров, хотя R^2 может быть очень высоким, что **ложно** свидетельствует о хорошем качестве регрессии.
3. Оценки параметров высоко чувствительны к даже незначительным изменениям выборочных данных (к выбору другой выборки из той же генеральной совокупности).
4. Высокие парные корреляции входных данных, или высокие R^2 регрессий входных данных по самим себе.
5. Число обусловленности матрицы $X^T \times X$, равное корню из отношения максимального к минимальному собственному значению матрицы, достаточно велико (Gujarati: если число в пределах от 10 до $10\sqrt{10}$ то средняя/сильная мультиколлинеарность, выше $10\sqrt{10}$ - очень сильная мультиколлинеарность). Самое важное это то, что собственные числа $X^T X$ при МК близки к 0 (а некоторые наоборот очень большие) и, следовательно, матрица практически необратима. Как следствие отсюда уже нестабильность, большие значения обратной к $X^T \times X$ и т.д.
6. $VIF(X_m) = \frac{1}{1 - R^2_{X_m|X_{-m}}}$ - фактор вздутия вариации - хороший индикатор МК. Если больше 3 (Мамонтов), то средняя/высокая МК; если больше 10 (Gujarati) то высокая МК.

Лечить мультиколлинеарность можно несколькими способами:

1. Можно увеличить размеры выборки, причём не на единицы наблюдений, а в разы. Тогда, скорее всего, проблема мультиколлинеарности исходных данных может исчезнуть.
2. Можно использовать априорную информацию о зависимостях в исходных данных, и соответствующим образом преобразовывать их.
3. Построить геометрически ортогональное преобразование векторов исходных данных, таким образом: 1 - найти проекцию некоего одного вектора исходных данных (X_m) на гиперплоскость, построенную на векторах всех кроме X_m (X_{-m}) (строим регрессию X_m по X_{-m} и получаем \hat{X}_m , саму проекцию). Затем, находим разность векторов $X_m - \hat{X}_m = X_m^{\perp}$ и получаем, что новый вектор X_m^{\perp} отчищен от влияния всех остальных векторов X_{-m} . Таки образом его корреляция с последними значительно снижается (вспомните ДЗ№5).

4. В случае с полиномиальной регрессией, преобразованием можно сделать простой перенос начала координат (центрирование) (вспомните ДЗ№6). К примеру, можно центрировать исходные данные значением параметра в точке среднего арифметического зависимых данных (\bar{Y}). Или можно в какую иную точку.
5. Можно использовать ортогональные полиномы Чебышёва. Но для этого наши исходные данные должны быть регулярными и равномерными (в них не должно быть уплотнений или дыр в значениях, если их отобразить на числовую прямую).
6. Можно воспользоваться методом Главных Компонент (нам о нём АА Мамонтов рассказывал, но не рекомендовал его использовать в силу сложности интерпретации коэффициентов).

Remark: Скорее всего эту проблему можно отнести к пункту 2 ТГМ (моё предположение).

Гетероскедастичность (ГС)

Эта проблема связана с предпосылкой 4 ТГМ. Как можно догадаться, в реальном мире гомоскедастичности (ТГМ) ожидать не следует. Она может быть как в генеральной совокупности, так и в силу некоей особой структуры выборочных данных (по опыту своих домашних могу сказать, что выборки у меня были такие, что оценки нужных параметров конкретно расходились с теми «идеальными» значениями, которые были у Андрея Александровича Мамонтова).

В общем, теория такова: $Var(\varepsilon_i) = \sigma_i^2; \forall i=1...n$ и скорее всего все σ_i^2 не равны между собой. ГС приводит к тому, что в общем случае оценки наших параметров не эффективны в классе линейных несмещённых.

Так вот, если использовать GLS (обобщённый метод наименьших квадратов) то проблем с неэффективностью можно избежать.

GLS

$$Y_i = \alpha + \beta_1 \cdot X_{1,i} + \beta_k \cdot X_{k,i} + \dots + \beta_k \cdot X_{k,i} + \varepsilon_i$$

Если предположить, что $Var(\varepsilon_i) = \sigma_i^2 = Z_i^2 \times \sigma^2; \forall i=1...n$, то, преобразовав модель, следующим образом получим выполнение предпосылки 4 ТГМ: так как

$$Var(\varepsilon_i / Z_i) = \sigma^2; \forall i=1...n, \quad \text{то} \quad Y_i / Z_i = \alpha \times 1 / Z_i + \beta_1 \cdot X_{1,i} / Z_i + \beta_k \cdot X_{k,i} / Z_i + \dots + \beta_k \cdot X_{k,i} / Z_i + \varepsilon_i / Z_i, \quad \text{где}$$

ошибки обладают свойством гомоскедастичности (вспомним ДЗ№6). Но, к сожалению, такая модель может и не содержать свободного члена, что приводит к соответствующим трудностям (ну, к примеру, если $\sigma_i^2 = f(X_{m_1,i}; \dots; X_{m_p,i})$ то свободного члена нет, а если $\sigma_i^2 = f(X_{m,i}) = \gamma \times X_{m,i}$ то свободный член есть). Кстати, в принципе, не обязательна зависимость дисперсии только от одного Z_i , но об этом позже.

Как оценки возможных дисперсий можно использовать абсолютные значения (модули) оценок ошибок, или квадраты этих оценок (**White**) (остатки, residuals) регрессии при достаточно большом размере выборки (Gujarati, лекции), но, к сожалению, в общем случае, остатки могут не быть независимыми, что усложняет их использование. Илья Борисович ещё указывает на существование рекурсивных остатков, которые обладают свойствами независимости, но мы в курсе эконометрики 1 их не рассматривали. На их основе существуют тесты, аналогичные приведённым ниже.

Выявить гетероскедастичность можно с помощью следующих способов и тестов:

1. **Графический метод:** взглянув на графики остатков по разным векторам исходных данных, включая объясняющие и объясняемые переменные, можно сделать вывод о наличии неких функциональных зависимостей. Но этот метод является неформальным и служит отправной точкой в исследовании ГС, «первой догадкой» о наличии гетероскедастичности.

Remark: Я считаю, что таким способом мы можем перепутать наличие гетероскедастичности с недоопределённостью модели (например, в домашке №6 мы по графику остатков определяли, что степень нашего полинома должна быть выше 1й [хотя существует теорема гласящая, что через n точек можно провести полиномиальную кривую n-1 порядка...]).

2. Тест Парка

(Park Test, 1966)

Парк предложил рассматривать следующую зависимость σ_i^2 от X_m .

$$\sigma_i^2 = \sigma^2 \cdot (X_{m,i})^\beta \cdot e^{v_i}, \text{ где } v_i - \text{белый шум.}$$

Park предлагает использовать квадраты оценок остатков (\hat{u}_i^2) и работать со следующей моделью:

$$\ln(\hat{u}_i^2) = \ln(\sigma^2) + \beta \times \ln(X_{m,i}) + v_i$$

Построив регрессию согласно этой модели, Park оценивает, является ли параметр β незначимым. Если да, то предположение о наличии гетероскедастичности мы принимаем. Выходит, что тест Парка – двухшаговый тест:

1 – Строим регрессию, используя модель

$$Y_i = \alpha + \beta_1 \cdot X_{1,i} + \beta_k \cdot X_{k,i} + \dots + \beta_k \cdot X_{k,i} + \varepsilon_i \text{ и находим остатки } \hat{u}_i.$$

1 и 1/2 – АА Мамонтов предлагает отфильтровать полученные остатки по следующей схеме: если остаток по модулю меньше чем 1/3 от стандартной ошибки

регрессии $\sqrt{(\hat{\sigma}^2)}$, то вместо него берём $\frac{\sqrt{(\hat{\sigma}^2)}}{3}$, в противном случае берём \hat{u}_i .

Нового обозначения отфильтрованным остаткам мы давать не будем.

2 – Оцениваем параметры регрессии

$$\ln(\hat{u}_i^2) = \ln(\sigma^2) + \beta \times \ln(X_m) + v_i \text{ по всем входным данным } X_{1..k}.$$

3 – Проверяем значимость угловых коэффициентов (β) и выносим суждение об отсутствии или наличии ГС.

3. Тест Глейзера

(Glejser Test, 1969)

Этот тест похож по алгоритму на тест Park, но имеет пару отличий. Glejser предлагает вместо \hat{u}_i^2 использовать $|\hat{u}_i|$ и несколько иные функциональные формы зависимости от исходных объясняющих данных:

$$|\hat{u}_i| = \beta_1 + \beta_2 \times (X_{m,i})^\gamma + v_i \text{ для } \gamma = -2; -1; -1/2; 1/2; 1; 2$$

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 \times X_{m,i}} + v_i$$

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 \times X_{m,i}^2} + v_i, \text{ где } v_i - \text{белый шум.}$$

Единственное – возникают проблемы с оценкой параметров в последних двух моделях, поскольку он не линейны по параметрам и требуют аппарата нелинейного оценивания. Glejser обнаружил, что первые регрессии ($\gamma = -2; -1; -1/2; 1/2; 1; 2$) в общем случае дают удовлетворительные результаты в обнаружении ГС (Gujarati). Таки образом весь тест сводится к определению значимости параметров (как и Park Test).

4. Тест Ранговой Корреляции Спирмена

(Spearman's rank correlation Test)

Этот тест относится к семейству тестов, не опирающихся на предположение о каком-то определённом законе распределения данных из генеральной совокупности. Сам Spearman только указал на существование этого теста и на

его простоту применимости. В 1935 году Hotelling и Pabst доказали асимптотическое стандартное нормальное распределение этого теста скорректированного на множитель $\sqrt{n-1}$ где n - размер выборки.

$$r = 1 - \frac{6 \times \sum_i d_i^2}{n(n^2 - 1)}$$

где d_i - разница между рангами, присвоенными двум пара разным характеристикам наблюдения с номером i , а n - размер выборки. Тогда $r * \sqrt{n-1} \stackrel{asy}{\sim} N(0;1)$.

Алгоритм теста следующий:

1 - Оцениваем основную регрессию

$$Y_i = \alpha + \beta_1 \cdot X_{1,i} + \beta_k \cdot X_{k,i} + \dots + \beta_k \cdot X_{k,i} + \varepsilon_i \text{ и находим остатки } \hat{u}_i.$$

2 - Ранжируем \hat{u}_i по их абсолютному значению, и (сохраняя «спаренность» наблюдений из выборки, то есть, не нарушаем исходную нумерацию данных) ранжируем $X_{m,i}$ или \hat{Y}_i . Считаем квадраты разностей рангов и вычисляем значение r .

3 - Проверяем следующую гипотезу, посчитав статистику по указанной ниже формуле:

$$H_0 : \rho = 0 \quad H_1 : \rho \neq 0$$

$$\frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}$$

4 - Получаем либо P_v , либо критические значения t распределения для нужного уровня значимости.

5 - Принимаем решение: если уровень значимости больше P_v , или рассчитанное значение лежит вне пределов критических значений, то отвергаем H_0 .

Результатом этого теста является утверждение о гетероскедастичности, если мы отвергаем H_0 .

Remark: Кстати можно использовать ранговую корреляцию Кендалла (это на любителя).

5. Тест Голдфельда-Квандта

(Goldfeld-Quandt Test, 1965)

Этот тест основывается на предположении о том, что $\sigma_i^2 = f(X_{m,i})$ существует и является неубывающей функцией. Например, $\sigma_i^2 = \sigma^2 \times (X_{m,i})^2$, то есть чем больше $X_{m,i}$, тем больше разброс. Goldfeld-Quandt предлагают следующую процедуру для оценки такого рода предположения о возрастании дисперсий.

1 - Упорядочим наши наблюдения по возрастанию данных $X_{m,i}$ (кстати можно и не только по $X_{m,i}$, а по любой переменной, даже не входящей в модель, к примеру по таким, переменным, по которым мы строим тест Breusch-Godfery, \hat{Y}_i и т.д.).

2 - Исключим c центральных наблюдений. Оставшиеся $n-c$ наблюдений разбиваем на две группы: группу соответствующую малыми значениям X_m и группу с большими X_m (в каждой группе по $\frac{n-c}{2}$ наблюдения).

3 - Обсчитываем основную модель регрессии

$$Y_i = \alpha + \beta_1 \cdot X_{1,i} + \beta_k \cdot X_{k,i} + \dots + \beta_k \cdot X_{k,i} + \varepsilon_i$$

по первой группе и по второй, и оцениваем RSS (Residual Sums of Squares). Получаем соответственно RSS_1 и RSS_2 , где RSS_2 соответствует выборки с большими X_m . У каждого RSS'а по $\frac{n-c}{2} - k - 1$ степени свободы.

4 - Теперь проверяем гипотезу о равенстве дисперсий двух выборок при неизвестных дисперсиях генеральной совокупности (обыкновенный тест из курса математической статистики), предположив, что ошибки нормально распределены:

$$H_0: \sigma^2_1 = \sigma^2_2 \quad H_1: \sigma^2_1 \neq \sigma^2_2$$

$$\lambda = \frac{RSS_2 / df_2}{RSS_1 / df_1} = \frac{RSS_2 / \left[\frac{(n-c)}{2} - k - 1 \right]}{RSS_1 / \left[\frac{(n-c)}{2} - k - 1 \right]} \sim F\left(\frac{(n-c)}{2} - k - 1; \frac{(n-c)}{2} - k - 1\right)$$

5 - Получаем либо P_v , либо критические значения F распределения для нужного уровня значимости.

6 - Принимаем решение: если уровень значимости больше P_v , или рассчитанное значение лежит вне пределов критических значений, то отвергаем H_0 .

Не отвергнув гипотезу H_0 , мы делаем вывод о том, что дисперсии ошибок положительно не зависят от тех данных $X_{m,i}$, по которым мы упорядочивали наши наблюдения, и, соответственно, присутствует гомоскедастичность.

Вообще говоря, мощность и успех теста зависит от c и от выбранного X_m , по которому производилось упорядочивание, поэтому для более точных результатов необходимо провести тестирование для нескольких различных c и для разных X_m .

Remark: В общем случае, не обязательно выбрасывать именно центральные наблюдения, можно разбить отсортированную выборку на две подвыборки размера n_1 и n_2 , причём необязательно что в сумме n_1 и n_2 дают размер исходной выборки. И теперь сравнивать дисперсии этих двух выборок, то есть проверять $H_0: \sigma^2_1 = \sigma^2_2 \quad H_1: \sigma^2_1 \neq \sigma^2_2$ но для этих нами полученных подвыборок. Правда в этом случае проверять основную гипотезу при помощи

$$\lambda = \frac{RSS_2 / df_2}{RSS_1 / df_1} = \frac{RSS_2 / [n_2 - k - 1]}{RSS_1 / [n_1 - k - 1]} \sim F(n_2 - k - 1; n_1 - k - 1)$$

Remark: Вообще говоря, сортировать данные выборки совсем не обязательно по какому-то X_m , определённого в модели (её спецификации) - можно взять какую-либо другую переменную Z_p (любой ряд не стохастических данных, ну, скажем, как в тесте

Бройша-Пэгана-Годффри) в качестве того, по чему мы сортируем наблюдения и проверяем неравенства дисперсий подвыборок.

6. Тест Бройша-Пэгана-Годффри

(Breusch-Pagan-Godfrey Test)

Этот тест предполагает одну очень хорошую и полезную вещь: дисперсии могут зависеть от абсолютно любого переменного, группы переменных или функционально преобразованных переменных. Итак, пусть основная модель выглядит так, как и раньше:

$$Y_i = \alpha + \beta_1 \cdot X_{1,i} + \beta_2 \cdot X_{2,i} + \dots + \beta_k \cdot X_{k,i} + \varepsilon_i$$

BPG предположили, что дисперсии ошибок описываются следующей функцией:

$$\sigma_i^2 = f(\theta_0 + \theta_1 \cdot Z_{1,i} + \theta_2 \cdot Z_{2,i} + \dots + \theta_m \cdot Z_{m,i}),$$

где Z_p - любой ряд не стохастических данных (некоторыми Z могут быть X , \hat{Y} , \hat{Y}^2).

В частном случае можно рассматривать следующую линейную зависимость дисперсий от Z вида $\sigma_i^2 = \theta_0 + \theta_1 \cdot Z_{1,i} + \theta_2 \cdot Z_{2,i} + \dots + \theta_m \cdot Z_{m,i}$ и если $\theta_1 = \theta_2 = \dots = \theta_m = 0$ то $\sigma_i^2 = \theta_0$, что есть константа, и соответственно ошибки обладают свойством гомоскедастичности. Иными словами, остаётся проверить гипотезу о $\theta_1 = \theta_2 = \dots = \theta_m = 0$, и сразу же получается результат о наличии или отсутствии гетероскедастичности. Итак, порядок тестирования следующий:

1 - Оцениваем исходную модель и получаем оценки остатков \hat{u}_i

2 - Считаем оценку дисперсии по методу максимального правдоподобия

$$\hat{\sigma}^2 = \frac{\sum_i \hat{u}_i^2}{n}$$

3 - Строим следующий ряд чисел $p_i = \frac{\hat{u}_i^2}{\hat{\sigma}^2}$.

4 - Оцениваем регрессию $p_i = \theta_0 + \theta_1 \cdot Z_{1,i} + \theta_2 \cdot Z_{2,i} + \dots + \theta_m \cdot Z_{m,i} + v_i$ и получаем ESS (объяснённая сумма квадратов).

5 - Рассчитываем следующую статистику, которая при выполнении основной гипотезы $H_0: \theta_1 = \theta_2 = \dots = \theta_m = 0$ и предпосылки о нормальности ошибок в исходной модели, асимптотически распределена по закону χ_{m+1-1}^2 :

$$\Theta = \frac{1}{2} \times ESS$$

6 - Получаем либо P_v , либо критические значения χ^2 распределения для нужного уровня значимости.

7 - Принимаем решение: если уровень значимости больше P_v , или рассчитанное значение лежит вне пределов критических значений, то отвергаем H_0 , и делаем вывод, что присутствует гетероскедастичность.

7. Общий тест гетероскедастичности Уайта

(White's General Heteroscedasticity Test)

White предлагает оценить регрессию \hat{u}_i^2 по исходным векторам данных, их квадратам и всем попарным произведениям. То есть оценить следующую регрессию по всем $\frac{k(k+3)}{2}+1$ параметрам θ :

$$\hat{u}_i^2 = \theta_0 + \sum_{m=1}^k \theta_m \cdot X_{m,i} + \sum_{m=1}^k \theta_{m+k} \cdot X_{m,i}^2 + \sum_{m=1}^k \sum_{p>m}^k \theta_{m+2k+g(m,p)} \cdot X_{m,i} X_{p,i},$$

где $g(m,p) = (m-1) \cdot k + p$ - некая функция, производящая индексы (это на самом деле и не важно какая это функция, я использовал её только для того, чтобы не вводить двумерные индексы...). Но, надо учесть, что количество оставшихся степеней свободы очень быстро уменьшается, что ведёт к менее точным оценкам параметров и нестабильным выводам. Алгоритм теста следующий:

- 1 - Оцениваем исходную модель и получаем оценки остатков \hat{u}_i
- 2 - Обсчитываем вышеуказанную регрессию и получаем R^2 .
- 3 - Проверим гипотезу о качестве вышеуказанной модели. При выполнении предпосылки о гомоскедастичности статистика $n \times R^2$ распределена асимптотически по закону χ_{df}^2 , где df равно количеству параметров в вышеуказанной регрессии за исключением свободного члена.
- 4 - Если основная гипотеза о гомоскедастичности не отвергается на заданном уровне значимости, то все параметры вышеуказанной регрессии за исключением свободного члена равны нулю. В противном случае, приходится принимать гипотезу о гетероскедастичности.

Remark: Как можно заметить, тест Уайта есть нечто похожее на тест Breusch-Pagan-Godfrey, и в принципе можно выполнить BPG Test по исходным векторам данных, их квадратам и всем попарным произведениям и сделать вывод о гетероскедастичности. И, казалось бы, что White test как бы «лишний», но этот тест играет важную роль в определении ошибок спецификации модели.

Устранить гетероскедастичность можно тем вот методом, о котором я рассказывал в начале раздела.

В общем, поскольку гетероскедастичность лишь делает наши оценки параметров неэффективными, если встаёт проблема выбора функционального вида модели, можно не обращать на ГС особо пристального внимания при этом. Но оценки, всё же, придётся считать уже каким-либо образом, избавившись от ГС.

Автокорреляция (АС) ошибок

В этом разделе рассмотрим последствия и решения проблем, связанных с нарушением предпосылки 5 ТГМ о независимости ошибок между собой ($Cov(\varepsilon_i, \varepsilon_j) \neq 0; \exists i, j = 1 \dots n; i \neq j$). Ужасающие последствия автокорреляции такие же, как и в случае гетероскедастичности: оценки параметров регрессии в общем случае не эффективны в классе линейных несмещённых.

Тесты, выявляющие АС, я изложу ниже, но сейчас отмечу, что каждый тест использует разный подход к обнаружению АС и, поэтому их результаты не обязательно должны совпадать:

1. **Графический тест:** по графикам остатков можно определить присутствие АС. Если нарисовать точечную диаграмму, где точками выступают $(e[t], e[t-1])$, то по тому, какую форму принимают скопления этих точек можно судить о наличии или отсутствии схемы AR(1) в остатках (АС). Также можно построить график остатков по номеру наблюдения: наличие резких изменений знака остатка может свидетельствовать об отрицательной АС; если же знаки остатков и их абсолютные значения изменяются не сильно, то можно ожидать положительную автокорреляцию.

Remark: Графический тест является просто картинкой и на его основе можно получить первое впечатление о наличии автокорреляции. Поэтому формальные тесты делать тоже необходимо.

Remark: В общем случае график остатков можно строить по $(e[t], e[t-i])$, но это нужно для иллюстрирования наличия АС порядка i .

2. Тест Серий

(Runs Test [Geary test], 1970)

Этот тест не опирается на априорное знание вида функции распределения исходных данных, поэтому затруднений с его применением быть не может. Алгоритм этого теста следующий:

- 1 - Оцениваем нашу основную регрессионную модель, и получаем оценки ошибок.
- 2 - Обращая внимание только на знаки этих остатков, определяем количество серий K в них. Серия - это непрерывная последовательность (даже единичной длины) неких атрибутов или символов, в нашем случае + и -. Также считаем, сколько значений оценок остатков больше не меньше и строго меньше нуля: N^+ и N^- . В сумме, очевидно, они дадут объём исходной выборки: $N = N^+ + N^-$.
- 3 - Считаем следующую статистику, которая при выполнении основной гипотезы асимптотически стандартно нормально распределена:

$$H_0 : AC = 0 \quad H_1 : AC \neq 0$$

$$K - \frac{2N^+N^- + N}{N} \stackrel{asy}{\sim} N(0;1)$$

$$\sqrt{\frac{2N^+N^- \cdot (2N^+N^- - N)}{N^2(N-1)}}$$

Очевидно, что достаточно большое количество серий характерно для отрицательной автокорреляции первого порядка, слишком маленькое - положительной АС. Очень простой и понятный тест.

3. Общий Тест Бройша-Годфри

(Breusch-Godfrey General Test, 1978)

Этот тест следует использовать для определения максимального порядка лага в авторегрессионной схеме остатков. Этот тест есть частный случай LM теста. Правда LM тест требует достаточно большой по объёму выборки. Итак, рассмотрим исходную модель

$$Y_t = \alpha + \beta_1 \cdot X_{1,t} + \beta_2 \cdot X_{2,t} + \dots + \beta_k \cdot X_{k,t} + u_t \quad t=1,2,\dots,n,$$

где X_m могут быть лаговыми зависимыми переменными, и ошибки u_t , образованные процессом $AR(p)$, где некоторые из ρ_m могут быть априорно равными нулю (те отсутствовать вовсе):

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \varepsilon_t \quad \varepsilon_t \sim iid N(0; \sigma^2)$$

Remark: Ошибки u_t могут быть образованы процессом $ARMA(p; q)$.

Нас интересует проверка гипотезы об одновременном равенстве нулю всех $\rho_1, \rho_2, \dots, \rho_p$, поскольку тогда автокорреляция отсутствует. Итак, алгоритм теста следующий:

1 - Оцениваем методом OLS регрессию согласно исходной модели. Получим остатки \hat{u}_t (и RSS^I).

2 - Оцениваем R^2 (и RSS^{II}) следующей регрессии:

$$\hat{u}_t = \alpha + \sum_{m=1}^k \beta_m \cdot X_{m,t} + \sum_{m=1}^p \rho_m \cdot u_{t-m} + v_t \quad t = p+1, p+2, \dots, n$$

3.1 - Maddala, Воскобойников:

Теперь считаем значение статистики, непосредственно проверяющей гипотезу H_0 : например, статистики теста значимости (см. выше) для p

штук линейных ограничений,
$$\hat{F} = \frac{RSS^I - RSS^{II} / p}{RSS^{II} / (n - (k+1) - p)},$$
 - и, для проверки

основной гипотезы, используем следующее свойство $p \cdot \hat{F} \stackrel{asy}{\sim} \chi_p^2$, поскольку размеры выборки достаточно велики.

3.2 - Gujarati, Мамонтов:

Теперь считаем следующую статистику: $(n-p) \cdot R^2 \stackrel{asy}{\sim} \chi_p^2$, при достаточно большой по размеру выборке.

4 - Получаем либо p_v , либо критические значения χ^2 распределения для нужного уровня значимости.

5 - Принимаем решение: если уровень значимости больше p_v , или рассчитанное значение лежит вне пределов критических значений, то отвергаем H_0 , и делаем вывод, что присутствует автокорреляция остатков.

Аналогично, мы можем оценить регрессию $Y_t = \alpha + \sum_{m=1}^k \beta_m \cdot X_{m,t} + \sum_{m=1}^p \rho_m \cdot u_{t-m} + v_t$, и проверять гипотезу $H_0: \rho_1 = \rho_2 = \dots = \rho_p = 0$ аналогично пункту 3.1 в алгоритме этого теста.

4. h-Тест Дарбина

(Durbin's h test)

Этот тест позволяет выявить автокорреляцию первого порядка при условии, что в спецификации модели допускается зависимость от лаговой объясняемой переменной. Таким образом, основная модель регрессии:

$$Y_t = \alpha + \sum_{m=1}^k \beta_m \cdot X_{m,t} + \gamma_1 Y_{t-1} + u_t \quad u_t = \rho u_{t-1} + v_t, \text{ где } v_t \sim iid N(0;1).$$

В общем случае, оценки этой модели, полученные методом OLS, не являются даже состоятельными. В любом случае, Durbin предлагает проверять наличие/отсутствие автокорреляции первого порядка следующим образом:

- 1 - По методу OLS всё-таки оцениваем параметры исходной модели, и получаем остатки \hat{u}_t и оценку дисперсии параметра γ_1 (при $Y[t-1]$).
- 2 - По остаткам оцениваем, согласно процессу AR(1), ρ .
- 3 - Вычисляем значение h-статистики, которая при выполнении гипотезы об отсутствии автокорреляции, распределена стандартно нормально.

$h = \hat{\rho} \times \sqrt{\frac{n}{1 - n \times \hat{V}(\hat{\gamma}_1)}}$, где $\hat{\rho}$ - оценка автокорреляции остатков и $\hat{V}(\hat{\gamma}_1)$ - оценённая дисперсия параметра при лаговой зависимой переменной.

- 4 - Получаем либо P_v , либо критические значения $N(0;1)$ распределения для нужного уровня значимости.
- 5 - Принимаем решение: если уровень значимости больше P_v , или рассчитанное значение лежит вне пределов критических значений, то отвергаем H_0 , и делаем вывод, что присутствует автокорреляция остатков.

Если не удаётся рассчитать h, то есть $1 - n \times \hat{V}(\hat{\gamma}_1) \leq 0$, Дарбин предлагает альтернативный тест:

- 1 - По методу OLS оцениваем параметры исходной модели, и получаем остатки \hat{u}_t .
- 2 - Строим регрессию остатков \hat{u}_t по \hat{u}_{t-1} , Y_{t-1} и $X_{m,t}$ (любые m, то есть, все объясняющие переменные).
- 3 - Проверяем значимость параметра при \hat{u}_{t-1} , что будет эквивалентно проверке гипотезы H_0 о нулевой автокорреляции.

Remark: В качестве оценки $\hat{\rho}$ автокорреляции остатков первого порядка можно использовать оценку из статистики DW или оценку методом OLS.

Этот тест является более легко применимым тестом, по сравнению с тестом Дарбина-Уотсона.

5. Тест Дарбина-Уотсона

(Durbin-Watson Test, 1951)

Итак, тест содержит несколько шагов:

- 1 - Оценить исходную регрессию, и получить оценки остатков.

2 - Посчитать $d = \frac{\sum_{t=2}^T (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^T \hat{u}_t^2}$ - статистику Дарбина-Уотсона

(от 0 (AC=1) до 4 (AC=-1)).

- 3 - Для заданных количества параметров, размера выборки и уровня значимости определяем верхнее и нижнее d значение (d_U и d_L).

4 - Проверяем следующие четыре гипотезы:

H_0^+ : положительная АС; H_1^+ : неположительная АС;

H_0^- : отрицательная АС; H_1^- : неотрицательная АС;

- причём все четыре гипотезы являются основными и «равноправными».

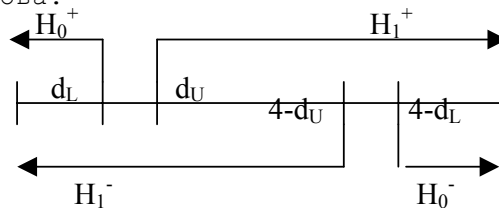
Одна из проблем с этим тестом возникает именно здесь, при проверке гипотез об автокорреляции, поскольку в области определения этой статистики (от нуля до четырех) находятся две области неопределённости, введенные Дарбином и Уотсоном, относительно знака и наличия АС. Эти области неопределённости не позволяют делать каких-либо определённых выводов.

Распределение этой статистики асимптотически симметрично относительно числа 2, поэтому в таблицах этого распределения указаны только верхние и нижние границы областей неопределённости меньшие 2 (то есть для гипотез H_0^+ и H_1^+).

Remark: Математическое ожидание статистики $E(DW) = 2 + 2 \cdot (m-1) / (N-m)$, где m - количество параметров, включая свободный член, N - размер выборки.

Область неопределённости для H_0^- и H_1^- примерно симметрична предыдущей относительно числа 2. Если для H_0^+ и H_1^+ это d_L и d_U , то для H_0^- и H_1^- это $4 - d_U$ и $4 - d_L$. Если DW меньше d_L то мы отвергаем H_1^+ и можем утверждать, что $AC > 0$. Если DW больше d_U то мы отвергаем H_0^+ и можем утверждать, что не наблюдается положительной АС (АС может быть меньшей, либо равной нулю). Если DW больше d_U , но меньше $(4 - d_U)$ то мы отвергаем H_0^+ и H_0^- , и можем утверждать, что не наблюдается ни положительной, ни отрицательной АС.

В общем, иллюстрация такова:



Remark: мы предполагаем, что $d_L < d_U < (4 - d_U) < (4 - d_L)$, хотя Мамонтов указывал на существование таких случаев когда это свойство нарушается.

У теста Durbin-Watson есть следующие ограничения, которые иногда забываются:

- 1) *Обязательное наличие свободного члена в спецификации модели.* Если не выполняется это условие, то необходимо использовать таблицы Farebrother теста (в общем случае d_L будут просто ниже тех, которые в таблицах DW, а d_U не изменятся - то есть расширится область неопределённости)
- 2) Не стохастические регрессоры.
- 3) *В остатках присутствует схема AR порядка не выше 1.* Наличие проблем в этом пункте выявляется общим тестом Бройша-Годфри. Или путём использования некоей другой статистики.
- 4) *Должны отсутствовать лаговые зависимые переменные.* В этом пункте, в случае его нарушения, нам может помочь Durbin's h Test.
- 5) Выборочные данные должны обладать регулярностью (данные не выпадают).
- 6) Ошибки должны быть нормально распределёнными случайными величинами.

Remark: Вообще говоря, основная гипотеза этого теста это наличие только нулевой автокорреляции. Вспомним, как мы проверяли обычные гипотезы в математической статистике: обсчитывали некую формулу, статистику, и проверяли, лежит ли нами найденное значение в пределах критических значений, для какого-то уровня значимости (конечно P_v - более «гибкое» решение проблемы оценки стат. гипотез, но и обычным методом не грех воспользоваться, поскольку он нагляднее).

Критические значения мы брали согласно тому закону распределения, по которому распределена наша статистика при выполнении основной гипотезы. Так вот, в DW тоже самое-то по идее, но вот это вот «по идее» здесь не пройдёт. Оказывается, потому, что критическое значение статистики Durbin-Watson зависит не только от размера выборки и количества регрессоров (без свободного члена), а ещё и от того какие значения примут объясняющие переменные. Поскольку, очевидно, не удастся построить точные таблицы распределения DW для любых выборок любого размера, учитывая любое количество регрессоров (без свободного члена) для любого уровня значимости, нам приходится пользоваться оценками этого критического для нас критического значения статистики DW при заданном уровне значимости. Поэтому и появляются нижнее и верхнее значения в таблицах, в чьих пределах лежит нужное нам критическое значение (конечно, ещё стоит вопрос о вероятности того, что при данной выборке критическое значение лежит в пределах этих d_L d_U). Так вот, поскольку не понятно, где именно в пределах d_L d_U лежит это значение, то область от d_L d_U мы нарекаем областью неопределённости и разводим руками, не зная, что делать при такой несправедливости. Теперь, две области неопределённости возникают в силу того, что нам не безразлично, какая автокорреляция у нас присутствует, при отвержении основной гипотезы. Интервал (d_L ; d_U) содержит критическое значение для альтернативной гипотезы об положительной АС, а ($4 - d_U$; $4 - d_L$) – критическое значение для альтернативной гипотезы об отрицательной АС.

Итак, наконец, можно приступить к теме борьбы с автокорреляцией ошибок, то есть к непосредственному оцениванию параметров модели. Кстати, эти методы тоже имеют различные обоснования и подходы к оцениванию, и таким образом, не исключена ситуация, когда их результаты различны.

Койчк

Это преобразование использовал **Койчк** для анализа модели **DL(+∞)** (смотри **Koyck's Approach** в разделе о динамических моделях). Но при преобразовании такого рода, теряется одно наблюдение, и, соответственно, уменьшается количество степеней свободы. Посему, **Praise-Winston** предложили не терять первое самое наблюдение и преобразовывать его, но несколько по-другому.

Итак, если ошибки в основной модели есть результат модели AR(1), и нам удалось оценить этот злосчастный параметр схемы AR(1), то модель

$$Y_t = \alpha + \beta_1 \cdot X_{1,t} + \beta_2 \cdot X_{2,t} + \dots + \beta_k \cdot X_{k,t} + u_t \quad u_t = \rho \cdot u_{t-1} + \varepsilon_t \quad \varepsilon_t \sim iid N(0; \sigma^2)$$

можно преобразовать к следующему виду, который можно оценивать с помощью OLS,

$$Y_t - \rho \cdot Y_{t-1} = \alpha \cdot (1 - \rho) + \beta_1 \cdot (X_{1,t} - \rho \cdot X_{1,t-1}) + \beta_2 \cdot (X_{2,t} - \rho \cdot X_{2,t-1}) + \dots + \beta_k \cdot (X_{k,t} - \rho \cdot X_{k,t-1}) + \varepsilon_t,$$

или переобозначив переменные $\tilde{Y}_t = Y_t - \rho \cdot Y_{t-1}$, $\tilde{X}_{1,t} = X_{1,t} - \rho \cdot X_{1,t-1}$, $\tilde{X}_{2,t} = X_{2,t} - \rho \cdot X_{2,t-1}$, ..., $\tilde{X}_{k,t} = X_{k,t} - \rho \cdot X_{k,t-1}$, $\tilde{\varepsilon}_t = \varepsilon_t$.

Единственный минус – теперь наблюдений у нас в наличии не n , а $n-1$, и нашу модель нужно будет оценивать OLS без свободного члена. **Praise-Winston** предложили не терять это одиночное наблюдение и сконструировать его следующим образом, дополнив вышеуказанное преобразование («поправить» преобразование – поправка **Praise-Winston**):

$$\begin{aligned} \tilde{Y}_t &= Y_t - \rho \cdot Y_{t-1} & \tilde{X}_{m,t} &= X_{m,t} - \rho \cdot X_{m,t-1} & \tilde{\gamma}_t &= 1 - \rho & t &\geq 2 \\ \tilde{Y}_1 &= \sqrt{1 - \rho^2} \cdot Y_1 & \tilde{X}_{m,1} &= \sqrt{1 - \rho^2} \cdot X_{m,1} & \tilde{\gamma}_1 &= \sqrt{1 - \rho^2} \end{aligned}$$

Модель и данные, преобразованные таким способом, не противоречат условиям ТГМ и не теряют наблюдений из выборки. Соответственно методом OLS (без свободного

члена) можно получить состоятельные и эффективные в классе линейных несмещённых оценки параметров $\alpha, \beta_1, \dots, \beta_k$, что, как раз, нам и надо. Это преобразование всем хорошо, только оно требует знания параметра ρ или его оценки.

Hildreth-Lu

Этот метод оценки параметров регрессии основан на так называемом методе поиска по решётке – методе приближений параметра схемы AR(1) путём перебора. Как и раньше модель наша такова:

$$Y_t = \alpha + \beta_1 \cdot X_{1,t} + \beta_2 \cdot X_{2,t} + \dots + \beta_k \cdot X_{k,t} + u_t \quad u_t = \rho \cdot u_{t-1} + \varepsilon_t \quad \varepsilon_t \sim iid N(0; \sigma^2)$$

Суть поиска такова:

0 – Очевидно, что схема AR(p) есть некое динамическое уравнение со случайной компонентой. Например, для AR(1) уравнение выглядит так:

$$u_t = \rho \cdot u_{t-1} + \varepsilon_t \quad \varepsilon_t \sim iid(0; \sigma^2)$$

При $\rho = 0$, процесс AR(1) превратится в простую случайную величину (чистый случайный процесс), то есть $u_t = \varepsilon_t$ ($\varepsilon_t \sim iid(0; \sigma^2)$), с нулевым мат. ожиданием, конечной дисперсией и нулевыми ковариациями. Такой процесс, кстати называется сильно стационарным.

При $|\rho| < 1$, процесс AR(1) становится слабо стационарным и его мат. ожидание равно нулю, дисперсия конечна и равна $\frac{\sigma^2}{1-\rho^2}$ и автоковариация равна

$$COV(u_t; u_{t-s}) = \frac{\rho^s \cdot \sigma^2}{1-\rho^2} \quad \forall s \geq 1.$$

А при $|\rho| = 1$, процесс AR(1) будет иметь следующие мат. ожидание и дисперсию: $E(u_t) = t \times 0$ $VAR(u_t) = t \times \sigma^2$. Этот процесс будет не стационарен и будет чистым видом случайного блуждания.

При $|\rho| > 1$, процесс AR(1) будет иметь «взрывную» динамику и его дисперсия будет бесконечной.

Поскольку в нашем курсе мы рассматривали слабо стационарные процессы, то область нашего перебора сужается до $|\rho| < 1$.

1 – Строим таблицу первых приближений ρ . К примеру, можно взять значения ρ от -0.9 до 0.9 (включая границы) с шагом 0.1. Взяв некое ρ из этого интервала, применяем преобразование **Koyck** с поправкой **Praise-Winston** (см. выше) и строим регрессию нашей модели с поправкой и получаем её RSS и R^2 . Выбираем такое ρ , которое даёт минимальное RSS (если два равных минимальных значения, то с большим R^2), и обозначим его ρ_1^* .

2 – Делаем всё аналогично первому пункту, только теперь с шагом 0.01, но в интервале от $\rho_1^* - 0.1 + 0.01$ до $\rho_1^* + 0.1 - 0.01$. Выбираем ρ_2^* аналогичным образом.

3 – Теперь перебираем все ρ с шагом 0.001 от $\rho_2^* - 0.01 + 0.001$ до $\rho_2^* + 0.01 - 0.001$ и аналогично выбираем ρ_3^* .

4 - И так далее, до той точности, которая нам нужна.

Полученная таким образом оценка ρ будет эффективной, и таким образом той нужной нам оценкой. Теперь уже можно применить преобразование **Koyck** с поправкой **Praise-Winston** (см. выше) и получить эффективные оценки параметров нашей модели. Этот метод очень хороший, поскольку решает в лоб задачу оценки ρ , но, как и любое решение в лоб, не является эффективным с вычислительной точки зрения.

Cochrane-Orcutt

Эта процедура находит нужный нам параметр модели

$$Y_t = \alpha + \beta_1 \cdot X_{1,t} + \beta_2 \cdot X_{2,t} + \dots + \beta_k \cdot X_{k,t} + u_t \quad u_t = \rho \cdot u_{t-1} + \varepsilon_t \quad \varepsilon_t \sim iid N(0; \sigma^2)$$

методом последовательного приближения параметра схемы AR(1) путём итераций.

Алгоритм его таков:

1 - Основная модель $Y_t = \alpha + \beta_1 \cdot X_{1,t} + \beta_2 \cdot X_{2,t} + \dots + \beta_k \cdot X_{k,t} + u_t$. Находим оценки её параметров $\alpha, \beta_1, \dots, \beta_k$ (обозначим $\alpha^0, \beta_1^0, \dots, \beta_k^0$) методом OLS.

2 - Получаем остатки $\hat{u}_t^i = Y_t - (\alpha^i + \beta_1^i X_{1,t} + \beta_2^i X_{2,t} + \dots + \beta_k^i X_{k,t})$, где $i = 0, 1, 2, \dots$ есть номер текущей итерации. Заметьте, что остатки считаются по исходным данным (без всяких преобразований).

3 - По остаткам \hat{u}_t^i считаем оценку параметра ρ (обозначим её $\hat{\rho}^i$) методом ML (максимального правдоподобия), которые в случае наших предпосылок совпадают с OLS (с точностью до смещённости оценки дисперсии, в случае ML она смещённая, а в OLS - нет). Короче пользуемся OLS для оценки параметра ρ . Если $|\hat{\rho}^i - \hat{\rho}^{i-1}| < 10^{-k}$, где k - точность оценивания или число итераций i превысило 30 (А.А. Мамонтов говорит, что этим числом ограничиваются все пакеты статистического анализа), то обозначаем только что полученный $\hat{\rho}^i$ символами $\hat{\rho}^{LAST}$, прерываем поиск и переходим к пункту 5 сего списка. В противном случае нас ожидает пункт 4.

4 - Получив $\hat{\rho}^i$, применяем преобразование **Koyck** с поправкой **Praise-Winston** (см. выше). Методом OLS по поправленным данным оцениваем $\alpha, \beta_1, \dots, \beta_k$ (обозначим $\alpha^{i+1}, \beta_1^{i+1}, \dots, \beta_k^{i+1}$). Теперь переходим к пункту 2 (возвращаемся к нему), и находим остатки \hat{u}_t^i с учётом новых оценок параметров $\alpha^{i+1}, \beta_1^{i+1}, \dots, \beta_k^{i+1}$, и не забываем, что по «старым» данным (исходным, непреобразованным, нетронутым, начальным, original, unmodified, unspoiled by Praise-Winston Transformation, и так далее... <сюда вставить ещё целую кучу синонимов со смыслом «не изменённый» :>).

5 - Если мы попали сюда, в этот пункт (а мы в него рано или поздно попадём), то, используя последнюю нами полученную оценку $\hat{\rho}^{LAST}$, применяем преобразование **Koyck** с поправкой **Praise-Winston** (см. выше) и методом OLS получаем эффективные оценки параметров $\alpha, \beta_1, \dots, \beta_k$ нашей модели.

Этот метод тоже очень хорош, но попадаются такие выборки, когда число проделанных итераций до сходимости с нужной точностью превышает 2, 3, 5 (Число указанное А.А. Мамонтовым) (например у меня сходиться стала с 18 шага), или когда на некой итерации оценка ρ превышает единицу. А так, метод имеет право на бытие. **Ура** товарищи!.

2 Step Durbin

Сей метод ещё проще, чем Cochrane-Orcutt. Дарбин предложил найти первое приближение параметра ρ , оценив следующую регрессию простым методом OLS (не задумываясь об ограничениях, связанных с применением OLS):

$$Y_t = \alpha \cdot (1 - \rho) + \rho \cdot Y_{t-1} + \beta_1 \cdot (X_{1,t} - \rho \cdot X_{1,t-1}) + \beta_2 \cdot (X_{2,t} - \rho \cdot X_{2,t-1}) + \dots + \beta_k \cdot (X_{k,t} - \rho \cdot X_{k,t-1}) + \varepsilon_t$$

Найдя оценку параметра при Y_{t-1} , $\hat{\rho}$, Durbin предлагает применить преобразование **Koyck** с поправкой **Praise-Winston** (см. выше) и методом OLS получить эффективные оценки параметров $\alpha, \beta_1, \dots, \beta_k$ нашей модели.

Durbin-Watson AR(1) parameter estimate

Можно показать, что при выполнении предпосылок применения статистики DW и при очень больших размерах выборки, $DW \approx 2 \cdot (1 - \rho)$. Соответственно, посчитав DW можно оценить ρ таким образом: $\hat{\rho} = 1 - \frac{DW}{2}$. Ну а теперь, уже можно применить преобразование **Koyck** с поправкой **Praise-Winston** (см. выше) и методом OLS получить эффективные оценки параметров $\alpha, \beta_1, \dots, \beta_k$ нашей модели (дёшево и сердито! :)).

Динамические модели в эконометрике

Обозначения. Определим оператор лага для упрощения и компактности изложения.

$$\begin{aligned} L(X_t) &= X_{t-1} & L^2(X_t) &= L(L(X_t)) = L(X_{t-1}) = X_{t-2} & L^k(X_t) &= X_{t-k} \\ L(\alpha \times X_t) &= \alpha \times L(X_t) & L^k(\alpha \times X_t) &= \alpha \times L^k(X_t) = \alpha \times X_{t-k} \\ L(\alpha) &= 1 \cdot \alpha. \end{aligned}$$

Обозначим многочлен степени m от x как $A_m(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \dots + \alpha_m x^m$. Частный случай многочлена степени m от x обозначим как $B_m(x) = 1 - \alpha_1 x - \alpha_2 x^2 - \dots - \alpha_m x^m$ и $C_m(x) = 1 + \alpha_1 x + \alpha_2 x^2 + \dots + \alpha_m x^m$.

Вполне очевидно, что

$$A_m(L)X_t = (\alpha_0 + \alpha_1 L + \alpha_2 L^2 + \dots + \alpha_m L^m)X_t = \alpha_0 X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_m X_{t-m}.$$

Далее, u_t - случайное «возмущение», ошибка, шум, причём iid (а вот распределения мы не знаем). Все модели рассматриваем только с одним регрессором, поскольку очевидно, как можно обобщить всё нижеизложенное. Итак, предположим, что X_t регрессор, при том «хороший» (некоррелирующий с ошибками, в случае стохастического регрессора).

Поехали!

Модель с распределёнными лагами порядка m - **DL**(m): $Y_t = D + A_m(L)X_t + u_t$, иными словами $Y_t = D + \alpha_0 X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_m X_{t-m} + u_t$.

Авторегрессионная модель порядка m - **AR**(m): $B_m(L)Y_t = D + u_t$, иными словами $Y_t = D + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_m Y_{t-m} + u_t$.

Модель скользящей средней порядка m - **MA**(m): $Y_t = D + C_m(L)u_t$, иными словами $Y_t = D + u_t + \alpha_1 u_{t-1} + \alpha_2 u_{t-2} + \dots + \alpha_m u_{t-m}$.

Авторегрессионная модель с распределёнными лагами порядка m и n - **ADL**($m; n$):

$$B_m(L)Y_t = A_n(L)X_t + u_t$$

или

$$Y_t = D + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_m Y_{t-m} + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_n X_{t-n} + u_t$$

Оценивание моделей.

Вообще, можно считать, что все X_t ($X_t; X_{t-1}; X_{t-2}; \dots; X_{t-k}; \dots$) которые мы могли наблюдать ранее, влияют на текущий Y_t . То есть, мы имеем дело с моделью **DL**($+\infty$). Итак рассматривать будем модель $Y_t = D + \alpha_0 X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_k X_{t-k} + \dots + u_t$

Alt-Tinbergen

Вообще говоря, это процедура поиска порядка лага в модели **DL**(...). Очевидно, её можно использовать и для случая с конечным числом лагов **DL**(m).

Процедура поиска следующая, начинаем с $m=0$:

1 - Оцениваем по исходным данным модель **DL**(m) методом **OLS**, если позволяет количество наблюдений. Если не позволяют - переходим на шаг 3.

2 - Вспомним оценки модели с $m-1$ лагами оценки. Если в текущих оценках наблюдается высокая нестабильность (высокие стандартные ошибки оценок

параметров или их незначимость) или что хотя бы знак одной из оценок для первых $m-1$ лаговых регрессоров **текущей** модели поменялся на противоположенный по сравнению с моделью с $m-1$ лагами, то переходим на шаг 3. В противном случае прибавляем к m единицу, и переходим на шаг 1.

3 – Если мы оказались здесь, то возможны две ситуации:

Нам не хватает наблюдений и соответственно количество степеней свободы очень мало, что приводит к нестабильным оценкам параметров.

Мы столкнулись либо с переменной знаков, либо с последствиями мультиколлинеарности (не стабильность оценок).

В первой ситуации, нам остаётся развести руками и попросить выдать нам ещё пару тройку десятков наблюдений, поскольку модель может иметь порядок высший, чем тот, на котором мы остановились.

Во второй – мы избавляемся (любыми средствами) от последней полученной нами модели. Выбираем ту модель, в которой все параметры при лагах значимы, и которую можно интерпретировать без проблем (вопрос, как интерпретировать знак минус при параметре лага, скажем, 5 порядка назад? Как найти ему объяснение?) Минусы у этой процедуры очевидны: не известно, сколько моделей с последовательно увеличивающимся количеством лагов нам придётся рассмотреть (предположим, что размер выборки достаточно велик); уродливая мультиколлинеарность, появляющаяся в моделях с большим количеством лагов; быстро уменьшающееся количество степеней свободы.

Koysck's Approach

Этот подход основывается на достаточно логичном умозаключении, что влияние прошлых событий со временем уменьшается, можно сказать, до нуля. Коуск предположил, что в $DL(+\infty)$ $Y_t = D + \alpha_0 X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_k X_{t-k} + \dots + u_t$ параметры априори являются убывающими функциями от времени (в бесконечности до нуля):

$\alpha_k = \alpha_0 \cdot \lambda^K$ для всех целых K не меньших нуля, $0 < \lambda < 1$. В данном случае новый параметр модели λ есть темп распада вещества, из которого состоит параметр (шутка, он состоит из чистого математического вещества, из чистой абстракции:)). Короче λ есть темп уменьшения параметра во времени. С таким предположением Коуск преобразует модель $DL(+\infty)$ $Y_t = D + \alpha_0 X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_m X_{t-m} + \dots + u_t$ в модель, которая по непонятности метода оценивания (бесконечное количество лагов, что сами знаете чем чревато) осталась прежней.

$$Y_t = D + \alpha_0 X_t + \alpha_0 \cdot \lambda^1 X_{t-1} + \alpha_0 \cdot \lambda^2 X_{t-2} + \dots + \alpha_0 \cdot \lambda^m X_{t-m} + \dots + u_t$$

Кстати, с такой предпосылкой о поведении параметров во времени мы исключаем всякие казусы, типа: знак минус при каком-то из параметров; необъяснимо высокий параметр при каком-то лаге где-то далеко от настоящего момента; сумма всех параметров при лагах, долгосрочный мультипликатор, конечна и равна $\frac{\alpha_0}{1-\lambda}$.

Koysck придумал гениальное решения проблемы оценивания этой модели: взять ту же самую модель но в шаге позади от текущего момента, помножить на λ и вычесть из модели для текущего момента. Правда, появились проблемы другого рода. Коуск превратил модель $DL(+\infty)$ в модель ARMA(1,1) с регрессором.

$$Y_t - \lambda Y_{t-1} = D \cdot (1 - \lambda) + \alpha_0 X_t + v_t \quad v_t = (u_t - \lambda \cdot u_{t-1})$$

В этой модели не будет мультиколлинеарности (все $X_{t-1}; X_{t-2}; \dots$ заменены одним $\lambda \cdot Y_{t-1}$), нам нужно оценить не бесконечное число параметров, а всего лишь три $D; \alpha_0; \lambda$, что есть хорошо. Но, теперь нам придётся учесть, что у нас появился стохастичный регрессор Y_{t-1} , который коррелирует с остатками, и что остатки обладают корреляцией между собой, что влечёт невозможность использования теста Durbin-Watson (но у нас есть тест h Дарбина) – вообще мы не можем использовать OLS для оценки этой модели. Хотя старые u_t и обладают свойствами:

$E(u_t) = 0$, $VAR(u_t) = \sigma^2$ и $COV(u_t; u_{t+s}) = 0$ для любого $s > 0$ – наши новые ошибки v_t не обладают ими, что не может радовать. Например $COV(v_t; v_{t+1}) = -\lambda \cdot \sigma^2$. Наш регрессор Y_{t-1} порождает следующее неприятное свойство: $COV(v_t; Y_{t-1}) = -\lambda \cdot \sigma^2$. Таким образом, OLS оценки не являются даже состоятельными. Но есть метод, который может нам помочь.

Instrumental Variables Method

Если нам удастся найти такой Z_t , заменитель Y_{t-1} (проксу для Y_{t-1}), обладающий следующими свойствами $p \lim_{n \rightarrow \infty} \left(\frac{\sum_t Z_t v_t}{n} \right) = 0$ и $p \lim_{n \rightarrow \infty} \left(\frac{\sum_t Z_t Y_{t-1}}{n} \right) = q \quad |q| > 0$, то оценки нашей модели будут состоятельными, даже с использованием метода OLS. Единственная проблема – найти достаточно хороший заменитель Y_{t-1} . Liviatan (1963) предлагает использовать в качестве IV X_{t-1} , поскольку он по определению не может коррелировать с X_t и, тем более, с ошибками v_t . Анализ и тестирование моделей с IV аналогичны простой линейной регрессии согласованна с ТГМ, но нам приходится пользоваться асимптотическим распределениями и свойствами.

Almon's Approach

Если коэффициенты при лагах не подчиняются геометрическому распаду, то можно использовать подход Almon. Её идея в следующем: пусть дана модель **DL(m)**.

Рассмотрим её в виде $Y_t = \alpha + \sum_{i=0}^m \beta_i X_{t-i} + u_t$. Almon предлагает вспомнить теорему, о том,

что через n точек можно провести полиномиальную кривую $n-1$ порядка, и использовать $\beta_i = A_K(i) = \alpha_0 + \alpha_1 \cdot i^1 + \alpha_2 \cdot i^2 + \dots + \alpha_K \cdot i^K$ (теорему Вейерштрасса). Подставив сие равенство

в модель DL(m) получаем $Y_t = \alpha + \sum_{i=0}^m (\alpha_0 + \alpha_1 \cdot i^1 + \alpha_2 \cdot i^2 + \dots + \alpha_K \cdot i^K) X_{t-i} + u_t$ и преобразовываем

до $Y_t = \alpha + \alpha_0 \sum_{i=0}^m X_{t-i} + \alpha_1 \sum_{i=0}^m i X_{t-i} + \alpha_2 \sum_{i=0}^m i^2 X_{t-i} + \dots + \alpha_K \sum_{i=0}^m i^K \cdot X_{t-i} + u_t$, что в свою очередь содержит $K+2$ параметра (а не неограниченное их число), правда, от изменённых

исходных данных. Обозначим $\sum_{i=0}^m i^s X_{t-i}$ как $Z_{s,t}$, и, таким образом модель, которую нам

придётся оценить упрощается до $Y_t = \alpha + \alpha_0 Z_{0,t} + \alpha_1 Z_{1,t} + \alpha_2 Z_{2,t} + \dots + \alpha_K Z_{K,t} + u_t$. Как мы можем заметить, степень полинома, которым мы пытаемся сгладить параметры модели (а ведь Almon именно это и предложила), не должна превышать порядка модели DL(...).

Оценивать эту модель достаточно просто - методом OLS. Итак, оценки новых параметров преобразованной модели эффективны и не смещены, и, следовательно, поскольку параметры $\beta_i = A_K(i)$ линейны относительно оцененных коэффициентов $\alpha_0; \alpha_1; \alpha_2; \dots; \alpha_K$, то оценки β_i можно с лёгкостью получить. Проблемы, связанные с этим подходом:

- 1 - Нужно определить порядок m модели DL.
- 2 - Нужно определить степень полинома, сглаживающего наши параметры, K .
- 3 - Возможно новые регрессоры Z обладают мультиколлинеарностью, что чревато нестабильностью оценок параметров.

Краткосрочные и долгосрочные эффекты (отклики) в модели ADL.

На лекциях мы рассматривали ADL(1;1) и ЕСМ, построенную на её основе. Рассмотрим ADL(1;1) и построим ЕСМ на её основе:

$y_t = m + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + u_t$ u_t - белый шум, а строчные буквы - натуральные логарифмы соответствующих заглавных букв.

Рассчитаем долгосрочное равновесие для $E(y_t) = \bar{y}$ в этой модели для некоего \bar{x} :

$$\bar{y} = E(m + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + u_t) = m + \alpha_1 \bar{y} + \beta_0 \bar{x} + \beta_1 \bar{x} + 0, \text{ таким образом, равновесием}$$

является следующее уравнение $\bar{y} = \frac{m}{1-\alpha_1} + \frac{\beta_0 + \beta_1}{1-\alpha_1} \bar{x}$. То же самое можно получить,

просуммировав все отклики текущего y_t на изменении x_{t-s} для любого s большего или равного нулю. Для этого в нашу модель последовательно подставим выражения для y_{t-1} , y_{t-2} , y_{t-3} и так далее. Получим следующее «чудо»:

$$y_t = m(1 + \alpha_1 + \alpha_1^2 + \dots) + (1 + \alpha_1 L + \alpha_1^2 L^2 + \dots)(\beta_0 x_t + \beta_1 x_{t-1}) + u_t$$

Используя определения оператора лага, перемножив и перегруппировав слагаемые из скобок, получим $y_t = m \frac{1}{1-\alpha_1} + [\beta_0 + (\beta_1 + \beta_0 \alpha_1)(L + \alpha_1 L^2 + \alpha_1^2 L^3 + \dots)]x_t + u_t$ из чего легко получить отклики текущего y_t на изменении x_{t-s} , для любого s большего или равного нулю.

$$\frac{\partial y_t}{\partial x_t} = \beta_0 \quad \frac{\partial y_t}{\partial x_{t-1}} = \beta_1 + \beta_0 \alpha_1 \quad \frac{\partial y_t}{\partial x_{t-2}} = \beta_1 \alpha_1 + \beta_0 \alpha_1^2 \quad \frac{\partial y_t}{\partial x_{t-k}} = \beta_1 \alpha_1^{k-1} + \beta_0 \alpha_1^k$$

Просуммировав все отклики, получаем $\sum_{k=0}^{\infty} [\beta_1 \alpha_1^{k-1} + \beta_0 \alpha_1^k] + \beta_0 = \frac{\beta_1}{1-\alpha_1} + \frac{\beta_0 \alpha_1}{1-\alpha_1} + \beta_0 = \frac{\beta_1 + \beta_0}{1-\alpha_1}$

что, заметьте равно $\frac{A_1(1)}{B_1(1)}$. В принципе, долгосрочный эффект от изменения текущего x_t

для модели ADL(m;n) равен $\frac{A_m(1)}{B_n(1)}$. Получаем, что в равновесии эластичность Y по X

постоянна и равна $\frac{\beta_1 + \beta_0}{1-\alpha_1} = \gamma$ и, соответственно, $\ln(Y) = \ln(A) + \gamma \times \ln(X)$ в долгосрочном равновесии.

Выведем теперь из всего вышесказанного модель ЕСМ. Подставим $\Delta y_t = y_t - y_{t-1}$ $\Delta x_t = x_t - x_{t-1}$ в ADL(1;1): $\Delta y_t = m + [\alpha_1 - 1]y_{t-1} + \beta_0 \Delta x_t + (\beta_1 + \beta_0)x_{t-1} + u_t$ - что упрощается до

$\Delta y_t = m + \beta_0 \Delta x_t + [\alpha_1 - 1](y_{t-1} - \gamma x_{t-1}) + u_t$ - **ЕСМ**. Кстати, с помощью такой модели проще оценить возможные значения параметра γ (см. выше). Нужно всего лишь вернуться к модели $\Delta y_t = m + [\alpha_1 - 1]y_{t-1} + \beta_0 \Delta x_t + (\beta_1 + \beta_0)x_{t-1} + u_t$ и проверить значимость равенства суммы параметра при y_{t-1} с параметром при x_{t-1} некоему числу с помощью обыкновенного «двуххвостого» t теста. Например, таким образом можно проверить гипотезу о единичной эластичности Y по X в равновесии.

Пример 1:

Предположим, что для модели $Y_i = \alpha + \beta_1 \cdot X_i + \varepsilon_i$ справедливы все предпосылки классической нормальной регрессионной модели, за исключением того, что дисперсии ошибок удовлетворяют соотношениям $\sigma^2_i = \mu + \gamma \cdot X_i$. Можно использовать следующую процедуру:

1 - Оценить уравнение $Y_i = \alpha + \beta_1 \cdot X_i + \varepsilon_i$ методом OLS и получим остатки u_i .

2 - Оценить уравнение $u^2_i = \mu + \gamma \cdot X_i + v_i$ и берём в качестве оценок дисперсии $\hat{\sigma}^2_i$ величины $\hat{\sigma}^2_i = \hat{\mu} + \hat{\gamma} \cdot X_i$.

3 - Применим метод взвешенных наименьших квадратов к исходной модели $Y_i = \alpha + \beta_1 \cdot X_i + \varepsilon_i$, т.е. строим регрессию $\frac{Y_i}{\hat{\sigma}^2_i} = \alpha \frac{1}{\hat{\sigma}^2_i} + \beta_1 \cdot \frac{X_i}{\hat{\sigma}^2_i} + \frac{\varepsilon_i}{\hat{\sigma}^2_i}$ по которой оцениваем параметры α и β_1 . Полученные оценки будут состоятельными.

Пример 2:

Рассмотрим модель $Y_i = \alpha + \beta_1 \cdot X_i + \varepsilon_i$, где $\varepsilon_i = \rho \cdot \varepsilon_{i-1} + u_i$

$E(u_i) = 0$, $Var(u_i) = \sigma^2$ и $Cov(u_i; u_s) = 0$ для $i \neq s$.

Оценить β_1 можно следующими действиями, причём оценка получится состоятельной, линейной и несмещённой, но не эффективной. Введём обозначение $\Delta Y_i = Y_i - Y_{i-1}$. Регрессия первой разности: $\Delta Y_i = \beta_1 \cdot \Delta X_i + \varepsilon^*_i$, где $\varepsilon^*_i = \varepsilon_i - \varepsilon_{i-1} = (\rho - 1) \cdot \varepsilon_{i-1} + u_i$.

Итак, будем считать, что $\varepsilon_1 = u_1$. Тогда $\varepsilon_i = \sum_{s=1}^i \rho^{i-s} \cdot u_s$. Обозначая

$\varepsilon^*_i = \varepsilon_i - \varepsilon_{i-1} = (\rho - 1) \cdot \varepsilon_{i-1} + u_i$, получаем $\Delta Y_i = \beta_1 \cdot \Delta X_i + \varepsilon^*_i = \beta_1 \cdot \Delta X_i + (\rho - 1) \cdot \varepsilon_{i-1} + u_i$, где $i = 2 \dots n$.

Оценив такую регрессию, получим линейную, несмещённую оценку параметра β_1 .