Владислав Морозов

ЭКОНО МЕТРИ КА

2015-2016

Владислав Морозов

Материалы к курсу эконометрики - 1

18 июня 2016 г.

Аннотация

Данный файл – записи лекций и семинаров первого курса эконометрики, читавшегося на Экономическом Факультете МГУ в 2015-2016 гг. В него включены ответы на теоретические вопросы для подготовки к письменным работам и решения всех семинарских задач.

Первая часть посвящена пространственным выборкам, вторая – панельным данными, третья – временным рядам.

0

Так как вести цифровые записи оказалось намного легче и приятнее, чем бумажные, постепенно этот файл разрастался и включал все больше материалов.

Благодарность

Спасибо Татьяне, Татьяне, Лидии, Кириллу, Илье, Алине, Ивану, Ивану, Дмитрию, Иосифу, Ирине и другим людям за их бесконечное терпение к опечаткам в первых версиях этого текста. Отдельная благодарность выражается А.В. за его внимательность к тонким вопросам оформления.

Обложка принадлежит тонкому вкусу К.Д., которому высказывается особая благодарность за нее.

Оглавление

Часть I

Теория

Глава 1
Контрольная работа 1

1.1 Вопрос 1. Математическое ожидание

Пусть ξ - дискретная случайная величина, принимающая значения $x_1...x_n$ с вероятностями $p_1...p_n$

Определение 1.1.1. Математическое ожидание дискретной случайной величины ξ есть величина $E\xi = \sum x_i p_i$, при условии, что ряд сходится абсолютно.

Пусть ξ - непрерывная случайная величина с плотностью распределения $p_\xi(x)$.

Определение 1.1.2. Математическое ожидание непрерывной случайной величины ξ есть величина $\mathrm{E}\,\xi = \int_{-\infty}^{+\infty} x_i p_\xi(x) dx$, взятая при условии, что интеграл сходится абсолютно.

Математическое ожидание обладает рядом свойств:

Свойство 1.1.1. $\mathrm{E}(const) = const$

Свойство 1.1.2. $\mathrm{E}(cx) = cE(x), c \in \mathbb{R}$

Свойство 1.1.3. $\mathrm{E}(x + y) = \mathrm{E}(x) + \mathrm{E}(y)$

Свойство 1.1.4. $\mathrm{E}(xy) = \mathrm{E}(x)\,\mathrm{E}(y)$, если x, y - независимые случайные величины

1.2 Вопрос 2. Дисперсия

Определение 1.2.1. Теоретическая дисперсия – величина, характеризующая степень разброса значений величины вокруг матожидания.

$$\mathrm{Var}(x) = \mathrm{E}(x - \mathrm{E}\,x)^2$$

У теоретической дисперсии есть ряд свойств:

Свойство 1.2.1. $\mathrm{Var}(x) = \mathrm{E}(x^2) - \mathrm{E}(x)^2$

Свойство 1.2.2. $\mathrm{Var}(const) = 0$

Свойство 1.2.3. $\mathrm{Var}(cx) = c^2 \mathrm{Var}(x)$

Свойство 1.2.4. $\mathrm{Var}(x + c) = \mathrm{Var}(x)$

Свойство 1.2.5. $\mathrm{Var}(x) = \mathrm{Cov}(x,x)$

Свойство 1.2.6. $\mathrm{Var}(ax \pm by) = a^2 \mathrm{Var}(x) + b^2 \mathrm{Var}(y) \pm 2ab \mathrm{Cov}(x,y)$

1.3 Вопрос 3. Ковариация

Определение 1.3.1. Теоретическая ковариация – величина, показывающая связь случайных величин x и y.

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

Это безразмерная величина, показывающая связь.

Свойство 1.3.1. $\mathrm{Cov}(x,y) = E(xy) - E(x)E(y)$

Свойство 1.3.2. $\mathrm{Cov}(x,const) = 0$

Свойство 1.3.3. $\mathrm{Cov}(cx,y) = c \mathrm{Cov}(x,y)$

Свойство 1.3.4. $\mathrm{Cov}(x + c,y) = \mathrm{Cov}(x,y)$

Свойство 1.3.5. $\mathrm{Var}(x) = \mathrm{Cov}(x,x)$

Свойство 1.3.6. $\mathrm{Cov}(x \pm y, z) = \mathrm{Cov}(x,z) \pm \mathrm{Cov}(y,z)$

Свойство 1.3.7. $\mathrm{Cov}(x,y) = \mathrm{Cov}(y,x)$

Свойство 1.3.8. $-\sqrt{\mathrm{Var}(x) \mathrm{Var}(y)} \le \mathrm{Cov}(x,y) \le \sqrt{\mathrm{Var}(x) \mathrm{Var}(y)}$

Свойство 1.3.9. $\mathrm{Cov}(x,y) = \pm\sqrt{\mathrm{Var}(x) \mathrm{Var}(y)} \iff x,y$ - линейно зависимые случайные величины

1.4 Вопрос 4. Выборочная ковариация

Определение 1.4.1. Выборочная ковариация – величина, показывающая связь случайных величин x и y на основе выборки.

$$\widehat{\mathrm{Cov}}(x,y) = \sum (x_i - \overline{x})(y_i - \overline{y})$$

Это безразмерная величина, показывающая связь.

Свойство 1.4.1. $\widehat{\mathrm{Cov}}(x,y) = \overline{xy} - \overline{x} \cdot \overline{y}$

Свойство 1.4.2. $\widehat{\mathrm{Cov}}(x, const) = 0$

Свойство 1.4.3. $\widehat{\mathrm{Cov}}(cx,y) = c\,\widehat{\mathrm{Cov}}(x,y)$

Свойство 1.4.4. $\widehat{\mathrm{Cov}}(x+c,y) = \widehat{\mathrm{Cov}}(x,y)$

Свойство 1.4.5. $\widehat{\mathrm{Var}}(x) = \widehat{\mathrm{Cov}}(x,x)$

Свойство 1.4.6. $\widehat{\mathrm{Cov}}(x \pm y, z) = \widehat{\mathrm{Cov}}(x,z) \pm \widehat{\mathrm{Cov}}(y,z)$

Свойство 1.4.7. $\widehat{\mathrm{Cov}}(x,y) = \widehat{\mathrm{Cov}}(y,x)$

1.5 Вопрос 5. Выборочная дисперсия

Определение 1.5.1. Выборочная дисперсия – величина, характеризующая степень разброса выборочных значений величины вокруг среднего.

$$\widehat{\mathrm{Var}}(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

У теоретической дисперсии есть ряд свойств:

Свойство 1.5.1. $\widehat{\mathrm{Var}}(x) = \overline{x^2} - \overline{x}^2$

Свойство 1.5.2. $\widehat{\mathrm{Var}}(const) = 0$

Свойство 1.5.3. $\widehat{\mathrm{Var}}(cx) = c^2\,\widehat{\mathrm{Var}}(x)$

Свойство 1.5.4. $\widehat{\mathrm{Var}}(x+c) = \widehat{\mathrm{Var}}(x)$

Свойство 1.5.5. $\widehat{\mathrm{Var}}(x) = \widehat{\mathrm{Cov}}(x,x)$

Свойство 1.5.6. $\widehat{\mathrm{Var}}(ax \pm by) = a^2\,\widehat{\mathrm{Var}}(x) + b^2\,\widehat{\mathrm{Var}}(y) \pm 2ab\,\widehat{\mathrm{Cov}}(x,y)$

Теоретическая и выборочная дисперсии заметно отличаются:

- Теоретическая строится по всей генеральной совокупности, а выборочная только для наблюдений $x_1 ... x_n$
- Выборочная дисперсия - смещенная оценка теоретической. Вместо нее часто вводят S^2
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$
- Выборочная дисперсия имеет тенденцию занижать значение теоретической, так как строится как отклонение от выборочного среднего: среднее заведомо лежит в центре выборки, поэтому отклонения оказываются меньше, чем если бы они оценивались относительно истинного значения матожидания

1.6 Вопрос 6. Коэффициент корреляции

Определение 1.6.1. Коэффициент корреляции случайных величин x и y с ненулевыми дисперсии есть величина:

$$\mathrm{Corr}(x,y) = \frac{\mathrm{Cov}(x,y)}{\sqrt{\mathrm{Var}(x)\,\mathrm{Var}(y)}}.$$

Он также известен как *линейный коэффициент корреляции* или *коэффициент корреляции Пирсона.*

Свойство 1.6.1. $-1 \leq \mathrm{Corr}(x,y) \leq 1$

Свойство 1.6.2. $\rho = \mathrm{Corr}(x,y) = \dfrac{E(x - E(x))(y - E(y))}{\sqrt{E(x - Ex)^2 E(y - Ey)^2}}.$

Свойство 1.6.3. $r = \widehat{\mathrm{Corr}}(x,y) = \dfrac{E(x - \overline{x})(y - \overline{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \cdot \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2}}.$

1.7 Вопрос 7. Несмещенность

Пусть есть параметр θ и его оценка $\hat{\theta}$.

Определение 1.7.1. Оценка $\hat{\theta}$ называется **несмещенной**, если $\mathrm{E}\hat{\theta} = \theta$.

1.8 Вопрос 8. Эффективность.

Пусть есть параметр θ и его оценка $\hat{\theta}$.

Определение 1.8.1. Оценка $\hat{\theta}$ называется **эффективной** в некотором классе оценок, если она имеет минимальную дисперсию в этом классе.

Определение 1.8.2. Информация Фишера - $I(\theta) = E\left(\frac{\partial \ln p(\xi,\theta)}{\partial\theta}\right)^2$

Сведения об эффективности мы получаем из неравенства Рао-Фреше-Крамера. Отсюда можно получить другое определение при выполнении условий теоремы.

Определение 1.8.3. Несмещенная оценка $\hat{\theta}$ называется **эффективной**, если $\mathrm{Var}(\hat{\theta}) = \frac{1}{nI(\theta)}$

1.9 Вопрос 9. Три типа данных

В эконометрике используется три типа данных:

1. Самая простая версия – пространственные выборки.

   Определение 1.9.1. Пространственные выборки (cross-sectional data) - набор значений экономический переменных, полученных в данный момент времени: все наблюдения получаются примерно в неизменных условиях, т. е. представляют собой набор независимых выборочных данных из некоторой генеральной совокупности.

   Пример I.1. Данные о стоимости квартир в г. Москва на конец этого месяца

2. Элемент времени добавляют временные ряды.

Определение 1.9.2. Временные ряды (time-series data) - выборка наблюдений, в которой важны не только сами наблюдаемые значения случайных величин, но и порядок их расположения друг за другом. Как правило, упорядоченность обусловлена тем, что экспериментальные данные представляют собой серию наблюдений одной и той же случайной величины в последовательные моменты времени.

Пример I.2. Ежеквартальные данные по инфляции

3. Характеристики пространственных и временных данных комбинируют панельные данные.

   Определение 1.9.3. Панельные данные (panel data) - прослеженные во времени пространственные экономические выборки: они состоят из наблюдений одних и тех же экономических единиц, которые осуществляются в последовательные периоды времени. Имеют три измерения: признаки - объекты - время.

   Пример I.3. Объекты – коммерческие фирмы; признаки – оборот, прибыль, число сотрудников, отрасль; время - за десять лет.

Есть вторая классификация:

1. Экспериментальные данные — получаются из контролируемых случайных экспериментов. (randomized controlled experiments)
2. Наблюдаемые статистические данные.

1.10 Вопрос 10. Пример с курсами

Исследователь выяснил, что люди, которые ходят на курсы по подготовке к поступлению в магистратуру экономического факультета, в среднем пишут вступительные экзамены в магистратуру хуже, чем люди, которые на эти курсы не ходят. И сделал вывод о том, что, принимая решение идти на курсы, абитуриент снижает свои шансы на поступление.

Однако в его логике есть изъян.

	Ходили на курсы	**Не ходили на курсы**
Выпускники бакалавриата ЭФ	55 баллов (10 человек)	50 баллов (95 человек)
Другие Абитуриенты	40 баллов (40 человек)	20 баллов (5 человек)

Исследователь не учел, что на курсы изначально будут ходить более слабые студенты, тем самым исключив существенные переменные.

Можно предложить эксперимент:

- Нужно собрать достаточно большое число людей.
- Случайным образом (вероятности попадания в каждую равны) их надо разбить на две группы и отправить одну на курсы, а другую не отправить. Можно ожидать, что в среднем знания одинаковы по группам.
- Методологическая яма в том, что это довольно дорогой эксперимент, и в той этической проблеме, что некоторые желают посещать курсы, но не попадают в нужную группу, и наоборот.

1.11 Вопрос 11. Пример с врачами

Используя данные о 10 000 человек, исследователь оценил коэффициент корреляции между количеством обращений к врачу в 2012 году и качеством здоровья в 2013 году. Коэффициент корреляции оказался отрицательным. Иными словами, выяснилось, что люди, которые чаще обращались к врачу впоследствии имели более плохое здоровье, чем те, кто к врачу не обращался. На основе этого результата исследователь сделал вывод о том, что обращаться к врачам вредно для здоровья.

Уважаемый исследователь забыл, что к врачу начнут часто ходить изначально очевидно больные люди[1], чье состояние не улучшается и требует постоянного медицинского наблюдения. Логично ожидать, что к следующему году они не могут сравниться с изначально здоровыми людьми. Вспомним еще, что корреляция не означает причинно-следственную связь.

Уважаемому исследователю можно предложить эксперимент:

- Нужно собрать достаточно большое число людей.
- Случайным образом (вероятности попадания в каждую равны) их надо разбить на две группы и не позволять одной ходить ко врачам вообще. Можно ожидать, что в среднем здоровье в начальный момент времени одинаково по группам.

[1] И еще ипохондрики

- Методологическая яма здесь в том, что на этот эксперимент уйдет заметное время, и в том, что довольно сложно будет оправдать этически непозволение страждущим посещать докторов.

- Альтернативно можно разделить группу на несколько подгрупп, в каждой из которых члену необходимо посетить врача некоторое фиксированное число раз.

1.12 Вопрос 12. Эконометрика

Определение 1.12.1. **Эконометрика** – наука, изучающая количественные и качественные экономические взаимосвязи с помощью математических и статистических методов и моделей

Пример I.4. Выявить связь между инфляцией и уровнем жизни населения.

Пример I.5. Выявить связь между исповедуемой религией и потреблением алкоголя.

Пример I.6. Выявить связь между успеваемостью в университете и заработной платой и давать прогнозы.

1.13 Вопрос 13. Парная регрессия. Теорема Гаусса-Маркова

Рассматривается классическая линейная модель парной регрессии с детерминированным регрессором 1.1:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i \tag{1.1}$$

Подбираем параметры так, чтобы сумма квадратов отклонений была минимальна:

$$e_1^2 + ... + e_n^2 = \sum_{i=1}^{n} e_i^2 \to \min$$

$$S = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2 \to \min \tag{1.2}$$

$$\begin{cases} \dfrac{\partial S}{\partial \hat{\beta}_1} = 0 \\ \dfrac{\partial S}{\partial \hat{\beta}_2} = 0 \end{cases}$$

$$\begin{cases} \sum\limits_{i=1}^{n} 2(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0 \\ \sum\limits_{i=1}^{n} 2x_i(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0 \end{cases}$$

$$\begin{cases} \sum\limits_{i=1}^{n} y_i - n\hat{\beta}_1 - \hat{\beta}_2 \sum\limits_{i=1}^{n} x_i = 0 \\ \sum\limits_{i=1}^{n} x_i y_{-} \hat{\beta}_1 \sum\limits_{i=1}^{n} x_i - \hat{\beta}_2 \sum\limits_{i=1}^{n} x_i^2 = 0 \end{cases}$$

$$\begin{cases} \left(\dfrac{1}{n} \sum\limits_{i=1}^{n} y_i\right) - \hat{\beta}_1 - \hat{\beta}_2 \left(\dfrac{1}{n} \sum\limits_{i=1}^{n} x_i\right) = 0 \\ \left(\dfrac{1}{n} \sum\limits_{i=1}^{n} x_i y_i\right) - \hat{\beta}_1 \left(\dfrac{1}{n} \sum\limits_{i=1}^{n} x_i\right) - \hat{\beta}_2 \left(\dfrac{1}{n} \sum\limits_{i=1}^{n} x_i^2\right) = 0 \end{cases}$$

$$\begin{cases} \overline{y} - \hat{\beta}_1 - \hat{\beta}_2 \overline{x} = 0 \\ \overline{xy} - \hat{\beta}_1 \overline{x} - \hat{\beta}_2 \overline{x^2} = 0 \end{cases}$$

$$\hat{\beta}_2 = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{\widehat{\mathrm{Cov}}(x,y)}{\widehat{\mathrm{Var}}(x)}$$

$$\hat{\beta}_1 = \overline{y} - \hat{\beta}_2 \cdot \overline{x}$$

Для всей этой красоты нам нужны предпосылки классической линейной модели парной регрессии:

1. Модель линейна по параметрам и правильно специфицирована: $y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$
2. $x_1 ... x_n$ - детерминированные величины, не все равные друг другу.
3. Математическое ожидание случайных ошибок равно нулю: $\mathrm{E}(\varepsilon)=0$
4. Случайные ошибки имеют постоянную дисперсию: $\mathrm{Var}(\varepsilon)=\sigma^2 = const$
5. Случайные ошибки, соответствующие разным наблюдения не зависят друг от друга: $\mathrm{Cov}(\varepsilon_i,\varepsilon_j) = 0$ при $i \neq j$
6. $\varepsilon_t \in N(0,\sigma^2)$

При таких условиях МНК-оценки весьма хороши:

Теорема I.1 (Теорема Гаусса-Маркова)**.** Если выполнены условия (1)-(5), то оценки $\hat{\beta}_1$,$\hat{\beta}_2$, полученные по МНК,

1. несмещенные
2. эффективные: имеют наименьшую дисперсию в классе всех линейных несмещенных оценок

Доказательство. 1. Несмещенность:

$$\hat{\beta}_2 = \frac{\frac{1}{n}\sum(x_i-\overline{x})(y_i-\overline{y})}{\frac{1}{n}\sum(x_i-\overline{x})^2} =$$

$$= \frac{\sum\left((x_i-\overline{x})(\beta_1+\beta_2\cdot x_i+\varepsilon_i-\beta_1-\beta_2\cdot\overline{x}-\overline{\varepsilon}_i)\right)}{\sum(x_i-\overline{x})^2} =$$

$$= \frac{\sum(x_i-\overline{x})(\beta_2(x_i-\overline{x})+\varepsilon_i-\overline{\varepsilon}_i)}{\sum(x_i-\overline{x})^2} =$$

$$= \frac{\sum\beta_2\sum(x_i-\overline{x})^2+\sum(x_i-\overline{x})(\varepsilon_i-\overline{\varepsilon})}{\sum(x_i-\overline{x})^2} =$$

$$= \beta_2 + \frac{\sum(x_i-\overline{x})\varepsilon_i-\overline{\varepsilon}\sum(x_i-\overline{x})}{\sum(x_i-\overline{x})^2} =$$

$$= \beta_2 + \frac{\sum(x_i-\overline{x})\varepsilon_i-\overline{\varepsilon}\cdot 0}{\sum(x_i-\overline{x})^2} =$$

$$= \beta_2 + \frac{\sum(x_i-\overline{x})\varepsilon_i}{\sum(x_i-\overline{x})^2}$$

$$\mathrm{E}(\hat{\beta}_2) = \mathrm{E}\left(\beta_2 + \frac{\sum(x_i-\overline{x})\varepsilon_i}{\sum(x_i-\overline{x})^2}\right) = \beta_2 + \frac{\sum(x_i-\overline{x})\,\mathrm{E}(\varepsilon_i)}{\sum(x_i-\overline{x})^2} = \beta_2$$

$$\hat{\beta}_1 = \overline{y} - \hat{\beta}_2\overline{x}$$

$$\mathrm{E}(\hat{\beta}_1) = \mathrm{E}(\overline{y}) - \overline{x}\,\mathrm{E}(\hat{\beta}_2)$$

Нам известно, что $\mathrm{E}(y_i) = \beta_1+\beta_2 x_i+\mathrm{E}(\varepsilon_i) = \beta_1+\beta_2 x_i$
Следовательно, $\mathrm{E}(\overline{y}) = \beta_1+\beta_2\overline{x}$
Подставляя назад, получим:

$$\mathrm{E}(\hat{\beta}_1) = \mathrm{E}(\overline{y}) - \overline{x}\,\mathrm{E}(\hat{\beta}_2) = \beta_1 + \beta_2\overline{x} - \beta_2\overline{x} = \beta_1$$

2. Эффективность:

Сначала докажем эффективность $\hat{\beta}_2$:

$$\hat{\beta}_2 = \frac{\frac{1}{n}\sum(x_i - \overline{x})(y_i - \overline{y})}{\frac{1}{n}\sum(x_i - \overline{x})^2} =$$

$$= \frac{\frac{1}{n}(\sum(x_i - \overline{x})y_i - \overline{y}\sum(x_i - \overline{x}))}{\frac{1}{n}\sum(x_i - \overline{x})^2} = \frac{\sum(x_i - \overline{x})y_i}{\sum(x_i - \overline{x})^2} =$$

Пусть $w_i = \dfrac{x_i - \overline{x}}{\sum(x_i - \overline{x})^2}$. Тогда:

$$= \sum w_i y_i$$

Таким образом, мы видим, что оценка линейна по y. Значит, мы находимся в классе линейных несмещенных оценок.

Покажем, что это действительно самая эффективная оценка класса.

Выпишем общий вид оценки и ограничение несмещенности:

$$\begin{cases} \hat{\beta}_2 = \sum c_i y_i \\ \mathrm{E}(\hat{\beta}_2) = E(\sum c_i(\beta_1 + \beta_2 x_i + \varepsilon_i)) = \beta_1 \cdot \sum c_i + \beta_2 \sum c_i x_i \end{cases}$$

Распишем дисперсию оценки:

$$\mathrm{Var}(\hat{\beta}_2) = \mathrm{Var}\left(\sum c_i y_i\right) = \sum c_i^2\,\mathrm{Var}(y_i) = \sigma^2 \sum c_i^2$$

Таким образом, мы можем сформулировать оптимизационную задачу, избавившись от константы σ^2.

$$\begin{cases} \sum c_i^2 \to \min \\ \sum c_i = 0 \\ \sum c_i x_i = 1 \end{cases}$$

Выписываем функцию Лагранжа:

$$\mathcal{L} = \sum c_i^2 - \lambda_1(\sum c_i x_i - 1) - \lambda_2 \sum c_i \to \min$$

Условия первого порядка:

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial c_i} = 2c_i - \lambda_1 x_i - \lambda_2 = 0 \\ \dfrac{\partial \mathcal{L}}{\partial \lambda_1} = 1 - \sum c_i x_i = 0 \\ \dfrac{\partial \mathcal{L}}{\partial \lambda_2} = \sum c_i = 0 \end{cases}$$

Просуммируем производные по c_i:

$$\sum \frac{\partial \mathcal{L}}{\partial c_i} = 2\sum c_i - \lambda_1 \sum x_i - \lambda_2 n = 0$$

$$\lambda_2 = -\lambda_1 \overline{x}$$

Подставляем это назад:

$$2c_i - \lambda_1 x_i + \lambda_1 \overline{x} = 0$$

$$2c_i - \lambda_1 (x_i - \overline{x}) = 0 | \cdot x_i$$

$$2c_i x_i - \lambda_1 x_i (x_i - \overline{x}) = 0$$

Суммируем это:

$$2\sum c_i x_i - \lambda_1 \sum x_i (x_i - \overline{x}) = 0$$

$$\lambda_1 = \frac{2}{\sum (x_i - \overline{x}) \cdot x_i}$$

Подставляем в $2c_i - \lambda_1(x_i - \overline{x}) = 0$.

$$2c_i - \frac{2}{\sum (x_i - \overline{x}) \cdot x_i}(x_i - \overline{x}) = 0$$

$$c_i = \frac{(x_i - \overline{x})}{\sum (x_i - \overline{x}) x_i} = \frac{(x_i - \overline{x})}{\sum (x_i - \overline{x}) \cdot x_i - \overline{x} \sum (x_i - \overline{x})} =$$
$$= \frac{(x_i - \overline{x})}{\sum (x_i - \overline{x})(x_i - \overline{x}))} = \frac{(x_i - \overline{x})}{\sum (x_i - \overline{x})^2}$$

$d^2\mathcal{L} > 0$. Таким образом, мы нашли точку минимума и доказали эффективность оценки.

Как мы понимаем,

$$\hat{\beta}_1 = \overline{y} - \hat{\beta}_2 \overline{x}$$

$$\operatorname{Var}(\hat{\beta}_1) = \frac{\sigma^2}{n} + \overline{x} \operatorname{Var}(\hat{\beta}_2)$$

Таким образом, задача минимизации дисперсии $\hat{\beta}_1$ эквивалентна задаче минимизации дисперсии $\hat{\beta}_2$. Таким образом, $\hat{\beta}_1$ - эффективная оценка β_1. □

Таким образом, чтобы показать несмещенность, нужно воспользоваться тем, что матожидание ε_i равно нулю, детерминированностью иксов, формой модели.

1.14 Вопрос 14. Пять свойств МНК-оценок

Имеем модель, для которой мы делаем МНК-оценки:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i \tag{1.3}$$

Для таких оценок верны пять утверждений:

Утверждение 1.14.1.

$$\sum e_i = 0$$

Доказательство. Для доказательства вернемся к условиям первого порядка:

$$\begin{cases} \sum\limits_{i=1}^{n} 2(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0 \\ \sum\limits_{i=1}^{n} 2x_i(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0 \end{cases}$$

В первом уравнении записано:

$$\sum_{i=1}^{n} 2(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = \sum_{i=1}^{n} 2(y_i - \hat{y}_i) = \sum_{i=1}^{n} 2e_i = 0$$

□

Утверждение 1.14.2.

$$\sum x_i e_i = 0$$

Доказательство. Для доказательства вернемся к условиям первого порядка:

$$\begin{cases} \sum\limits_{i=1}^{n} 2(y_i - \hat\beta_1 - \hat\beta_2 x_i) = 0 \\ \sum\limits_{i=1}^{n} 2x_i(y_i - \hat\beta_1 - \hat\beta_2 x_i) = 0 \end{cases}$$

Второе уравнение дает нам:

$$\sum_{i=1}^{n} 2x_i(y_i - \hat\beta_1 - \hat\beta_2 x_i) = 2\sum x_i e_i = 0$$

Следовательно, $\sum x_i e_i = 0$.

□

Утверждение 1.14.3. $\sum y_i = \sum \hat y_i$

Доказательство. Так как мы знаем, что $\sum e_i = 0$, мы можем записать, что $\sum(y_i - \hat y_i) = 0$, откуда следует, что $\sum y_i = \sum \hat y_i$. □

Утверждение 1.14.4. $\sum(\hat y_i - \overline{y})e_i = \mathrm{Cov}(\hat y, e) = 0$

Доказательство. Будем раскрывать скобки:

$\sum(\hat y_i - \overline{y})e_i = \sum \hat y_i \cdot e_i - \overline{y}\sum e_i = \sum(\hat\beta_1 + \hat\beta_2 x_i)e_i =$
$= \hat\beta_1 \sum e_i + \hat\beta_2 \sum x_i e_i = 0$

Таким образом, видим, что прогнозные значения игрека некоррелированы с ошибками. □

Утверждение 1.14.5. $\sum(y_i - \overline{y})^2 = \sum(\hat y_i - \overline{y})^2 + \sum e_i^2$ или же, другими словами, $TSS = RSS + ESS$

Доказательство. Отметим факт о дисперсиях:

$\widehat{\mathrm{Var}}(y) = \widehat{\mathrm{Var}}(e + \hat y) = \widehat{\mathrm{Var}}(e) + \widehat{\mathrm{Var}}(\hat y) + 2\,\widehat{\mathrm{Cov}}(e, \hat y) =$
$= \widehat{\mathrm{Var}}(e) + \widehat{\mathrm{Var}}(\hat y)$

Запишем, расшифровав:

$$\frac{1}{n}\sum(y_i - \overline{y})^2 = \frac{1}{n}\sum e_i^2 + \frac{1}{n}\sum(\hat y - \overline{y})^2$$

$$\sum(y_i - \overline{y})^2 = \sum e_i^2 + \sum(\hat y - \overline{y})^2$$

$$TSS = ESS + RSS$$

□

1.15 Вопрос 15. Коэффициент детерминации и корреляция

Определение 1.15.1. Коэффициент детерминации R^2 показывает долю дисперсии зависимой переменной, объясненной уравнением регрессии:

$$R^2 = \frac{\widehat{\mathrm{Var}}(\hat{y})}{\widehat{\mathrm{Var}}(y)} = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS} = \frac{\frac{1}{n}\sum(\hat{y}-\overline{y})^2}{\frac{1}{n}\sum(y_i-\overline{y})^2} = 1 - \frac{\sum e_i^2}{\sum(y_i-\overline{y})^2}$$

Утверждение 1.15.1. Пусть дана модель $y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$.

Тогда коэффициент детерминации R^2 равен квадрату коэффициента корреляции между переменными x и y.

Доказательство. Распишем подробно R^2.

$$R^2 = \frac{\widehat{\mathrm{Var}}(\hat{y})}{\widehat{\mathrm{Var}}(y)} = \frac{\widehat{\mathrm{Var}}(\hat{\beta}_1 + \hat{\beta}_2 x)}{\widehat{\mathrm{Var}}(y)} = \hat{\beta}_2^{\,2}\frac{\widehat{\mathrm{Var}}(x)}{\widehat{\mathrm{Var}}(y)}$$

Помним, что $\hat{\beta}_2 = \frac{\widehat{\mathrm{Cov}}(x,y)}{\widehat{\mathrm{Var}}(x)}$. Подставляем:

$$\hat{\beta}_2^{\,2}\frac{\widehat{\mathrm{Var}}(x)}{\widehat{\mathrm{Var}}(y)} = \frac{\widehat{\mathrm{Cov}}^2(x,y)}{\widehat{\mathrm{Var}}^2(x)} \cdot \frac{\widehat{\mathrm{Var}}(x)}{\widehat{\mathrm{Var}}(y)} = \frac{\widehat{\mathrm{Cov}}^2(x,y)}{\widehat{\mathrm{Var}}(x)\,\widehat{\mathrm{Var}}(y)} = \widehat{\mathrm{Corr}}^2(x,y)$$

□

1.16 Вопрос 16. Дисперсия коэффициента при регрессоре

Рассматривается классическая линейная модель парной регрессии с детерминированным регрессором:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i \tag{1.4}$$

Утверждение 1.16.1. $\mathrm{Var}\,\hat{\beta}_2 = \frac{\sigma^2}{\sum(x_i-\overline{x})^2}$

Доказательство. Пребразуем $\hat{\beta}_2$ к полезному виду:

$$\hat{\beta}_2 = \frac{\frac{1}{n}\sum(x_i-\overline{x})(y_i-\overline{y})}{\frac{1}{n}\sum(x_i-\overline{x})^2} =$$

$$= \frac{\sum(x_i-\overline{x})(\beta_1+\beta_2\cdot x_i+\varepsilon_i-\beta_1-\beta_2\cdot\overline{x}-\overline{\varepsilon_i})}{\sum(x_i-\overline{x})^2} =$$

$$= \frac{\sum(x_i-\overline{x})(\beta_2(x_i-\overline{x})+\varepsilon_i-\overline{\varepsilon_i})}{\sum(x_i-\overline{x})^2} =$$

$$= \frac{\sum\beta_2\sum(x_i-\overline{x})^2+\sum(x_i-\overline{x})(\varepsilon_i-\overline{\varepsilon})}{\sum(x_i-\overline{x})^2} =$$

$$= \beta_2 + \frac{\sum(x_i-\overline{x})\varepsilon_i-\overline{\varepsilon}\sum(x_i-\overline{x})}{\sum(x_i-\overline{x})^2} =$$

$$= \beta_2 + \frac{\sum(x_i-\overline{x})\varepsilon_i-\overline{\varepsilon}0}{\sum(x_i-\overline{x})^2} =$$

$$= \beta_2 + \frac{\sum(x_i-\overline{x})\varepsilon_i}{\sum(x_i-\overline{x})^2}$$

Будем искать дисперсию нашей оценки.

$$\operatorname{Var}(\hat{\beta}_2) = \operatorname{Var}\left(\beta_2 + \frac{\sum(x_i-\overline{x})\varepsilon_i}{\sum(x_i-\overline{x})^2}\right) =$$

$$= \operatorname{Var}\left(\frac{\sum(x_i-\overline{x})\varepsilon_i}{\sum(x_i-\overline{x})^2}\right) = \left(\frac{1}{\sum(x_i-\overline{x})^2}\right)^2\cdot\operatorname{Var}\left(\sum(x_i-\overline{x})\varepsilon_i\right) =$$

Здесь мы можем заменить дисперсию суммы на сумму дисперсий, используя предпосылку (5): $\operatorname{Cov}(\varepsilon_i,\varepsilon_j)=0$ при $i\neq j$

$$= \left(\frac{1}{\sum(x_i-\overline{x})^2}\right)^2\cdot\sum\left(\operatorname{Var}((x_i-\overline{x})\varepsilon_i)\right) = \left(\frac{1}{\sum(x_i-\overline{x})^2}\right)^2\cdot\sum(x_i-\overline{x})^2\operatorname{Var}(\varepsilon_i) =$$

$$= \left(\frac{1}{\sum(x_i-\overline{x})^2}\right)^2\cdot\sum(x_i-\overline{x})^2\sigma^2 = \frac{\sum(x_i-\overline{x})^2}{\sum(x_i-\overline{x})^2}\cdot\frac{1}{\sum(x_i-\overline{x})^2}\sigma^2 = \frac{\sigma^2}{\sum(x_i-\overline{x})^2}$$

□

1.17 Вопрос 17. Дисперсия оценки константы

Рассматривается классическая линейная модель парной регрессии с детерминированным регрессором:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i \tag{1.5}$$

Утверждение 1.17.1. $\operatorname{Var} \hat{\beta}_1 = \dfrac{\sigma^2}{n} + \dfrac{\sigma^2 \overline{x}^2}{\sum (x_i - \overline{x})^2}$

Доказательство. Разложим дисперсию $\hat{\beta}_1$ на три компонента:

Используем, что $\hat{\beta}_1 = \overline{y} - \overline{x}\hat{\beta}_2 = \beta_1 + \beta_2\overline{x} + \overline{\varepsilon} - \hat{\beta}_2\overline{x}$, откуда видно, что $\hat{\beta}_1$ – несмещенная оценка.

$$\begin{aligned}\operatorname{Var}(\hat{\beta}_1) = \operatorname{Var}(\beta_1 + \beta_2\overline{x} + \overline{\varepsilon} - \hat{\beta}_2\overline{x}) = \operatorname{Var}(\overline{\varepsilon} - \hat{\beta}_2\overline{x}) = \\ = \operatorname{Var}(\overline{\varepsilon}) + \operatorname{Var}(\hat{\beta}_2\overline{x}) - 2\operatorname{Cov}(\overline{\varepsilon}, \hat{\beta}_2\overline{x})\end{aligned}$$

Рассмотрим каждый из этих пунктов отдельно:

1. Дисперсия $\overline{\varepsilon}$

$$\operatorname{Var}(\overline{\varepsilon}) = \operatorname{Var}\left(\frac{\varepsilon_1 + \dots \varepsilon_n}{n}\right) = \left(\frac{1}{n}\right)^2 \operatorname{Var}(\sum \varepsilon_i) = \frac{\sigma^2}{n}$$

2. $\operatorname{Var}(\hat{\beta}_2\overline{x}) = \overline{x}^2 \cdot \dfrac{\sigma^2}{\sum (x_i - \overline{x})^2}$

3. $\operatorname{Cov}(\overline{\varepsilon}, \hat{\beta}_2\overline{x}) = \overline{x}\operatorname{Cov}(\overline{\varepsilon}, \beta_2 + \dfrac{\sum (x_i - \overline{x})\varepsilon_i}{\sum (x_i - \overline{x})^2}) = \overline{x}\dfrac{\operatorname{Cov}(\overline{\varepsilon}, \sum (x_i - \overline{x})\varepsilon_i)}{\sum (x_i - \overline{x})^2} =$
   $= \overline{x}\dfrac{\sum (x_i - \overline{x})\operatorname{Cov}(\overline{\varepsilon}, \varepsilon_i)}{\sum (x_i - \overline{x})^2} = \dfrac{\sum (x_i - \overline{x})\sigma^2}{\sum (x_i - \overline{x})^2} = \dfrac{0}{\sum (x_i - \overline{x})^2}$

   Согласно предпосылке (5), $\operatorname{Cov}(\varepsilon_i, \varepsilon_j) = 0$ при $i \neq j$, поэтому все слагаемые, где $i \neq j$, обнуляются. Остаются только с совпадающими индексами.

   Используем при этом также факт, что $\sum (x_i - \overline{x}) = \sum x_i - n\overline{x} = \sum x_i - n\frac{1}{n}\sum x_i = 0$

Следовательно,

$$\operatorname{Var} \hat{\beta}_1 = \frac{\sigma^2}{n} + \frac{\sigma^2 \overline{x}^2}{\sum (x_i - \overline{x})^2}$$

$\square$

1.18 Вопрос 18. Ковариация оценок коэффициентов

Рассматривается классическая линейная модель парной регрессии с детерминированным регрессором:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i \tag{1.6}$$

Утверждение 1.18.1. $\mathrm{Cov}(\hat{\beta}_1; \hat{\beta}_2) = -\overline{x}\dfrac{\sigma^2}{\sum(x_i - \overline{x})^2}$

Доказательство. Заменим под знаком ковариации $\hat{\beta}_1$:

$$\hat{\beta}_1 = \overline{y} - \hat{\beta}_2 \cdot \overline{x}$$

$$\mathrm{Cov}(\hat{\beta}_1; \hat{\beta}_2) = \mathrm{Cov}(\overline{y} - \hat{\beta}_2\overline{x}; \hat{\beta}_2) = \mathrm{Cov}(\overline{y}; \hat{\beta}_2) - \overline{x}\,\mathrm{Cov}(\hat{\beta}_2; \hat{\beta}_2)$$

Помним, что $\mathrm{Cov}(\hat{\beta}_2,\hat{\beta}_2) = \mathrm{Var}(\hat{\beta}_2) = \dfrac{\sigma^2}{\sum(x_i - \overline{x})^2}$

Подставляем назад:

$$\mathrm{Cov}(\overline{y}; \hat{\beta}_2) - \overline{x}\frac{\sigma^2}{\sum(x_i - \overline{x})^2} = \mathrm{Cov}(\overline{\varepsilon},\hat{\beta}_2) - \overline{x}\frac{\sigma^2}{\sum(x_i - \overline{x})^2}$$

Докажем, что первое слагаемое равно нулю:

$$\mathrm{Cov}(\overline{\varepsilon},\hat{\beta}_2) = \mathrm{Cov}(\overline{\varepsilon},\hat{\beta}_2) = \mathrm{Cov}(\overline{\varepsilon}, \beta_2 + \frac{\sum(x_i - \overline{x})\varepsilon_i}{\sum(x_i - \overline{x})^2}) = \frac{\mathrm{Cov}(\overline{\varepsilon}, \sum(x_i - \overline{x})\varepsilon_i)}{\sum(x_i - \overline{x})^2} =$$

$$= \frac{\sum(x_i - \overline{x})\,\mathrm{Cov}(\overline{\varepsilon},\varepsilon_i)}{\sum(x_i - \overline{x})^2} = \frac{\sum(x_i - \overline{x})\sigma^2}{\sum(x_i - \overline{x})^2} = \frac{0}{\sum(x_i - \overline{x})^2}$$

Согласно предпосылке (5), $\mathrm{Cov}(\varepsilon_i,\varepsilon_j) = 0$ при $i \neq j$, поэтому все слагаемые, где $i \neq j$, обнуляются. Остаются только с совпадающими индексами.

Используем при этом также факт, что $\sum(x_i - \overline{x}) = \sum x_i - n\overline{x} = \sum x_i - n\dfrac{1}{n}\sum x_i = 0$.

Следовательно,

$$\mathrm{Cov}(\hat{\beta}_1; \hat{\beta}_2) = -\overline{x}\frac{\sigma^2}{\sum(x_i - \overline{x})^2}$$

□

1.19 Вопрос 19. Регрессия без константы

Рассматривается альтернативная версии регрессии:

$$y_i = \theta x_i + \varepsilon_i \tag{1.7}$$

Необходимо вычислить МНК-оценку коэффициента и показать ее несмещенность.

Утверждение 1.19.1. МНК-оценка θ $\quad \hat{\theta} = \dfrac{\overline{xy}}{\overline{x^2}}$

Доказательство. Выпишем задачу и условие первого порядка.

Подбираем параметры так, чтобы сумма квадратов отклонений была минимальна:

$$e_1^2 + ... + e_n^2 = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{\theta} x_i)^2 \to \min$$

$$\frac{\partial S}{\partial \hat{\theta}} = 0$$

Вау, как все классно, да?

$$\frac{\partial S}{\partial \hat{\theta}} = \sum_{i=1}^{n} -2x_i(y_i - \hat{\theta} x_i) = 0$$

$$\sum x_i y_i - \hat{\theta} \sum x_i^2 = 0$$

$$\hat{\theta} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\overline{xy}}{\overline{x^2}}$$

□

Утверждение 1.19.2. $\hat{\theta}$ – несмещенная оценка параметра θ.

Доказательство. Преобразуем:

$$\hat{\theta} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i(\theta x_i + \varepsilon_i)}{\sum x_i^2} = \frac{\theta \sum x_i^2 + \sum x_i \varepsilon_i}{\sum x_i^2} =$$

$$= \theta + \frac{\sum x_i \varepsilon_i}{\sum x_i^2}.$$

Проверяем несмещенность:

$$E(\hat{\theta}) = E(\theta + \frac{\sum x_i \varepsilon_i}{\sum x_i^2}) = \theta + E(\frac{\sum x_i \varepsilon_i}{\sum x_i^2}) = \theta + 0 = \theta$$

□

Утверждение 1.19.3. $\mathrm{Var}(\hat{\theta}) = \dfrac{\sigma^2}{\sum x_i^2}$.

Доказательство.

$$\mathrm{Var}(\theta + \frac{\sum x_i \varepsilon_i}{\sum x_i^2}) = Var(\frac{\sum x_i \varepsilon_i}{\sum x_i^2}) = \frac{\mathrm{Var}(\sum x_i \varepsilon_i)}{(\sum x_i^2)^2} =$$

Используем предпосылки модели о независимости и дисперсии ошибок:

$$= \frac{\sum x_i^2 \,\mathrm{Var}(\varepsilon_i)}{(\sum x_i^2)^2} = \frac{\sigma^2}{\sum x_i^2}.$$

□

1.20 Вопрос 20. Регрессия на константу

Рассматривается альтернативная версии регрессии:

$$y_i = \theta + \varepsilon_i \tag{1.8}$$

Необходимо вычислить МНК-оценку коэффициента и показать ее несмещенность.

Утверждение 1.20.1. МНК-оценка θ $\quad \hat{\theta} = \overline{y}$.

Доказательство. Выпишем задачу и условие первого порядка.

Подбираем параметры так, чтобы сумма квадратов отклонений была минимальна:

$$e_1^2 + ... + e_n^2 = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{\theta})^2 \to \min$$

$$\frac{\partial S}{\partial \hat{\theta}} = 0$$

Очевидно, что

$$\frac{\partial S}{\partial \hat{\theta}} = \sum_{i=1}^{n} -2(y_i - \hat{\theta}) = 0$$

$$-2 \sum y_i = n\hat{\theta} = 0$$

$$\hat{\theta} = \overline{y}$$

□

Утверждение 1.20.2. $\hat{\theta}$ – несмещенная оценка параметра θ.

Доказательство. Проверяем несмещенность:

$$E(\hat{\theta}) = E(\overline{y}) = \frac{1}{n}\sum E(y) = \frac{1}{n}E(n\theta + n\varepsilon_i) = \theta + E(\varepsilon_i) = \theta$$

□

Утверждение 1.20.3. $\mathrm{Var}(\hat{\theta}) = \dfrac{\sigma^2}{n}$.

Доказательство.

$$\mathrm{Var}(\hat{\theta}) = \mathrm{Var}(\overline{y}) = \left(\frac{1}{n}\right)^2 \mathrm{Var}(\sum y_i) = \left(\frac{1}{n}\right)^2 \mathrm{Var}(n\theta + \sum \varepsilon_i) =$$

$$= \left(\frac{1}{n}\right)^2 \mathrm{Var}(\sum \varepsilon_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

□

1.21 Вопрос 21. Коэффициент детерминации

Определение 1.21.1. Коэффициент детерминации R^2 показывает долю дисперсии зависимой переменной, объясненной уравнением регрессии:

$$R^2 = \frac{\widehat{\mathrm{Var}}(\hat{y})}{\widehat{\mathrm{Var}}(y)} = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS} = \frac{\frac{1}{n}\sum(\hat{y} - \overline{y})^2}{\frac{1}{n}\sum(y_i - \overline{y})^2} = 1 - \frac{\sum e_i^2}{\sum(y_i - \overline{y})^2}$$

Пример I.7. Пусть дано, что R^2=0,79.

Это означает, что 79% изменений y объяснены изменениями x в нашей модели, а остальные 21% описываются чем-то иным. Точность модели достаточно высока, а связь весьма тесна. Конечно, всегда остается риск ложной регрессии.

1.22 Вопрос 22. Предпосылки КЛМПР

Предпосылки классической линейной модели парной регрессии:

1. Модель линейна по параметрам и правильно специфицирована: $y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$

2. $x_1 ... x_n$ - детерминированные величины, не все равные друг другу.
3. Математическое ожидание случайных ошибок равно нулю: E(ε)=0
4. Случайные ошибки имеют постоянную дисперсию: Var(ε)=$\sigma^2 = const$
5. Случайные ошибки, соответствующие разным наблюдения не зависят друг от друга: Cov(ε_i,ε_j) $= 0$ при $i \neq j$
6. $\varepsilon_t \in N(0,\sigma^2)$

1.23 Вопрос 23. Теорема Гаусса-Маркова

Теорема I.2 (Теорема Гаусса-Маркова)**.** Если выполнены условия:

1. Модель линейна по параметрам и правильно специфицирована: $y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$
2. $x_1 ... x_n$ - детерминированные величины, не все равные друг другу.
3. Математическое ожидание случайных ошибок равно нулю: E(ε)=0
4. Случайные ошибки имеют постоянную дисперсию: Var(ε)=$\sigma^2 = const$
5. Случайные ошибки, соответствующие разным наблюдения не зависят друг от друга: Cov(ε_i,ε_j) $= 0$ при $i \neq j$,

то оценки $\hat{\beta}_1$,$\hat{\beta}_2$, полученные по МНК,

1. несмещенные: $E(\hat{\beta}_1) = \beta_1$, $E(\hat{\beta}_2) = \beta_2$
2. эффективные: имеют наименьшую дисперсию в классе всех линейных несмещенных оценок.

1.24 Вопрос 24. Случайные ошибки и остатки регрессии

В модели парной регрессии существует примерная взаимосвязь вида:

$$y \approx \beta_1 + \beta_2 x_i \tag{1.9}$$

Из этого мы получаем $y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$. При этом y имеет ту же функцию распределения, что и ε при предпосылке детерминированности регрессоров.

Определение 1.24.1. Случайная ошибка ε_i - ошибки модели, возникающая из-за

- Невключения объясняющих переменных;
- Агрегирования объясняющих переменных;
- Неправильной структуры модели;
- Ошибок измерения.

С помощью оценки мы можем получить прогнозные значения $\hat{y}$. Тогда мы можем ввести новое понятие:

Определение 1.24.2. Остатки регрессии $e_i \overset{\text{def}}{=} y_i - \hat{y}_i$ - разность между наблюдаемыми значениями и прогнозными, полученными по регрессии.

1.25 Вопрос 25. Значимость коэффициента

Пусть имеется модель вида

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i. \tag{1.10}$$

Нас интересует вопрос о значимости коэффициента при переменной x.

Для получения ответа на этот вопрос мы формулируем две конкурирующие гипотезы H_0 и H_1.

$H_0 : \beta_2 = 0$
$H_1 : \beta_2 \neq 0$

Как мы помним:

$$\hat{\beta}_2 = \frac{\frac{1}{n}\sum(x_i - \overline{x})(y_i - \overline{y})}{\frac{1}{n}\sum(x_i - \overline{x})^2} =$$

$$= \frac{\sum(x_i - \overline{x})(\beta_1 + \beta_2 \cdot x_i + \varepsilon_i - \beta_1 - \beta_2 \cdot \overline{x} - \overline{\varepsilon_i})}{\sum(x_i - \overline{x})^2} =$$

$$= \frac{\sum(x_i - \overline{x})(\beta_2(x_i - \overline{x}) + \varepsilon_i - \overline{\varepsilon_i})}{\sum(x_i - \overline{x})^2} =$$

$$= \frac{\sum \beta_2 \sum(x_i - \overline{x})^2 + \sum(x_i - \overline{x})(\varepsilon_i - \overline{\varepsilon})}{\sum(x_i - \overline{x})^2} =$$

$$= \beta_2 + \frac{\sum(x_i - \overline{x})\varepsilon_i - \overline{\varepsilon}\sum(x_i - \overline{x})}{\sum(x_i - \overline{x})^2} =$$

$$= \beta_2 + \frac{\sum(x_i - \overline{x})\varepsilon_i - \overline{\varepsilon}0}{\sum(x_i - \overline{x})^2} =$$

$$= \beta_2 + \frac{\sum(x_i - \overline{x})\varepsilon_i}{\sum(x_i - \overline{x})^2}$$

Если выполнены предпосылки

- x_i - детерминированные величины;
- $\varepsilon_i \sim N(0;1)$,

то $\hat{\beta}_2$ - ЛК нормальных случайных величин.

Таким образом, $\hat{\beta}_2 \sim N\left(\beta_2; \frac{\sigma^2}{\sum(x_i - \overline{x})^2}\right)$

Это значит, что величина $\frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} \in t_{n-2}$

В рамках гипотезы H_0 предполагается, что $\beta_2 = 0$. Таким образом, $t_{\text{расч}} = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)}$.

Мы выбираем необходимый нам уровень значимости, после чего из таблицы распределения Стьюдента мы выбираем критическое значение статистики t для двустороннего интервала.

Мы сравниваем $|t_{\text{расч}}|$ и $t_{\text{кр}}$. Если $|t_{\text{расч}}| > t_{\text{кр}}$, то мы отклоняем нулевую гипотезу: с уверенностью $100 - \alpha\%$ мы можем утверждать, что переменная x значимо влияет на y. В альтернативном случае мы принимаем нулевую гипотезу: переменная x не оказывает значимого влияния на y.

1.26 Вопрос 26. Уровень значимости

Определение 1.26.1. Уровень значимости - вероятность ошибки первого рода в модели.

Определение 1.26.2. Вероятность ошибки первого рода - вероятность отклонить гипотезу, если она верна.

1.27 Вопрос 27. Корреляция регрессора и предсказанного значения

Рассматривается классическая линейная модель парной регрессии с детерминированным регрессором:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i. \tag{1.11}$$

Коэффициенты регрессии были оценены при помощи МНК.

Утверждение 1.27.1. $\widehat{\mathrm{Corr}}(x,\hat{y}) = \begin{cases} 1 \\ 0 \\ -1 \end{cases}.$

Доказательство. По свойству 1.3.9 ковариации $\mathrm{Cov}(x,y) = \pm\sqrt{\mathrm{Var}(x)\,\mathrm{Var}(y)} \iff x,y$ - линейно зависимые случайные величины. Это значит, что $\mathrm{Corr}(x,y) = \pm 1 \iff x,y$ - линейно зависимые случайные величины.

$$\widehat{\mathrm{Corr}}(x,\hat{y}) = \widehat{\mathrm{Corr}}(x,\hat{\beta}_1 + \hat{\beta}_2 x)$$

.

Таким образом, мы видим, что при $\hat{\beta}_2 \neq 0$ существует линейная связь между x и $\hat{y}$. Значит, при $\hat{\beta}_2 \neq 0$ $\widehat{\mathrm{Corr}}(x,\hat{y}) = \pm 1$.

Если же $\hat{y} = const$, то $\hat{\beta}_2 = 0$. Тогда, согласно свойству 1.3.2, мы получаем, что $\widehat{\mathrm{Corr}}(x,\hat{y}) = 0$. □

1.28 Вопрос 28. Альтернативная оценка

Исследователь правильно предполагает, что зависимость между переменными x и y описывается уравнением:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i \tag{1.12}$$

При этом, имея выборку из переменных x, y и некоторой третьей детерминированной переменной z, исследователь вместо МНК-оценки использует альтернативную оценку для коэффициента β_2

$$\tilde{\beta}_2 = \frac{\sum(z_i - \overline{z})(y_i - \overline{y})}{\sum(z_i - \overline{z})(x_i - \overline{x})}$$

Утверждение 1.28.1. $\tilde{\beta}_2$ - несмещенная оценка β_2.

Доказательство. Преобразуем числитель:

$$\sum(z_i - \overline{z})(y_i - \overline{y}) = \sum(z_i - \overline{z})(\beta_1 + \beta_2 x_i + \varepsilon_i - \beta_1 - \beta_2\overline{x} - \overline{\varepsilon}) =$$

$$= \sum(z_i - \overline{z})(\beta_2 \cdot (x_i - \overline{x}) + \varepsilon_i - \overline{\varepsilon}) = \beta_2 \sum(z_i - \overline{z})(x_i - \overline{x}) + \sum(z_i - \overline{z})(\varepsilon_i - \overline{\varepsilon})$$

Таким образом,

$$E(\tilde{\beta}_2) = E\left(\frac{\beta_2 \sum(z_i - \overline{z})(x_i - \overline{x}) + \sum(z_i - \overline{z})(\varepsilon_i - \overline{\varepsilon})}{\sum(z_i - \overline{z})(x_i - \overline{x})}\right) =$$

$$= \beta_2 + \frac{\sum(z_i - \overline{z})E(\varepsilon_i - \overline{\varepsilon})}{\sum(z_i - \overline{z})(x_i - \overline{x})} = \beta_2$$

□

1.29 Вопрос 29. Свойства прогноза

Рассматривается классическая линейная модель парной регрессии:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i. \tag{1.13}$$

На основе n наблюдений оценено уравнение регрессии:

$$\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x_i. \tag{1.14}$$

Пусть известно (n+1)-ое значение регрессора: x_{n+1}. Используя его, предскажем y_{n+1}: $\hat{y}_{n+1} = \hat{\beta}_1 + \hat{\beta}_2 x_i$.

Утверждение 1.29.1. $\hat{y}_{n+1}$ - несмещенная оценка y_{n+1}.

Доказательство. Найдем математическое ожидание прогноза:

$$E(\hat{y}_{n+1}) = E(\hat{\beta}_1 + \hat{\beta}_2 x_{n+1}) = E(\hat{\beta}_1) + E(\hat{\beta}_2)x_{n+1} = \beta_1 + \beta_2 x_{n+1}$$

Найдем теперь математическое ожидание истинного значения:

$$E(y_{n+1}) = E(\beta_1 + \beta_2 x_{n+1} + \varepsilon_{n+1}) = \beta_1 + \beta_2 x_{n+1} + E(\varepsilon_{n+1}) = \beta_1 + \beta_2 x_{n+1}$$

Таким образом,

$$E(\hat{y}_{n+1}) = E(y_{n+1})$$

□

Утверждение 1.29.2. $\hat{y}_{n+1}$ - эффективная несмещенная оценка y_{n+1}.

Утверждение 1.29.3. $\operatorname{Var}(\hat{y}_{n+1} - y_{n+1}) = \sigma^2 \left(1 + \frac{1}{n} + \frac{(x_{n+1} - \overline{x})^2}{\sum(x_i - \overline{x})^2}\right)$.

Доказательство. Преобразуем выражения под знаком дисперсии и разложим ее на отдельные слагаемые.

$$\operatorname{Var}(\hat{y}_{n+1}-y_{n+1}) = \operatorname{Var}(\hat{\beta}_1+\hat{\beta}_2x_{n+1}-\beta_1-\beta_2x_{n+1}-\varepsilon_{n+1}) = \operatorname{Var}(\hat{\beta}_1+\hat{\beta}_2x_{n+1}-\varepsilon_{n+1}) =$$

$$= \operatorname{Var}(\hat{\beta}_1) + (x_{n+1})^2 \operatorname{Var}(\hat{\beta}_2) + \operatorname{Var}(\varepsilon_{n+1}) + 2x_{n+1} \operatorname{Cov}(\hat{\beta}_1,\hat{\beta}_2) - \\ -2 \operatorname{Cov}(\hat{\beta}_1,\varepsilon_{n+1}) - 2 \operatorname{Cov}(\hat{\beta}_2,\varepsilon_{n+1})$$

.

Рассмотрим каждое из слагаемых:

- Согласно утверждению 1.17.1 $\operatorname{Var}\hat{\beta}_1 = \frac{\sigma^2}{n} + \frac{\sigma^2\overline{x}^2}{\sum(x_i - \overline{x})^2}$;
- Согласно утверждению 1.16.1 $\operatorname{Var}\hat{\beta}_2 = \frac{\sigma^2}{\sum(x_i - \overline{x})^2}$;
- Согласно утверждению 1.18.1 $\operatorname{Cov}(\hat{\beta}_1; \hat{\beta}_2) = -\overline{x}\frac{\sigma^2}{\sum(x_i - \overline{x})^2}$;
- $\operatorname{Cov}(\hat{\beta}_1,\varepsilon_{n+1}) = \operatorname{Cov}(\hat{\beta}_2,\varepsilon_{n+1}) = 0$, так как β - случайный вектор, зависящий от $(\varepsilon_1 ... \varepsilon_n)$, а, согласно нашем предпосылкам, $\forall i \operatorname{Cov}(\varepsilon_i, \varepsilon_{n+1}) = 0$.

  Значит,

$$\operatorname{Var}(e_{n+1}) = \sigma^2+\frac{\sigma^2}{n}+\frac{\sigma^2\overline{x}^2}{\sum(x_i - \overline{x})^2}+(x_{n+1})^2\frac{\sigma^2}{\sum(x_i - \overline{x})^2}-2x_{n+1}\overline{x}\frac{\sigma^2}{\sum(x_i - \overline{x})^2} =$$

$$= \sigma^2 \left(1 + \frac{1}{n} + \frac{(x_{n+1} - \overline{x})^2}{\sum(x_i - \overline{x})^2}\right)$$

□

Следствие 1. $se(\hat{y}_{n+1} - y_{n+1}) = \sqrt{S^2 \left(1 + \frac{1}{n} + \frac{(x_{n+1} - \overline{x})^2}{\sum(x_i - \overline{x})^2}\right)}$, где S^2 - исправленная оценка дисперсии.

1.30 Вопрос 30. Предпосылки КЛММР

Предпосылки классической линейной модели множественной регрессии:

1. Модель линейна по параметрам и правильно специфицирована: $y_i = \beta_1 + \beta_2 \cdot x_i^{(2)} + ... + \beta_k \cdot x_i^{(k)} + \varepsilon_i$
2. $\boldsymbol{x_1 ... x_n}$ - детерминированные ЛНЗ вектора.
3. Математическое ожидание случайных ошибок равно нулю: $\mathrm{E}(\boldsymbol{\varepsilon}) = \overline{0}$
4. Случайные ошибки имеют постоянную дисперсию: $\mathrm{Var}(\varepsilon)=\sigma^2 = const$
5. Случайные ошибки, соответствующие разным наблюдения не зависят друг от друга: $\mathrm{Cov}(\varepsilon_i,\varepsilon_j) = 0$ при $i \neq j$
6. $\varepsilon_t \in N(0,\sigma^2)$

Альтернативно можно сформулировать предпосылки (4) и (5) через матрицу ковариаций:

$$V(\boldsymbol{\varepsilon}) = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix}$$

1.31 Вопрос 31. Теорема Гаусса-Маркова

Пусть X – матрица регрессоров, а y – вектор значенией объясняемой переменной, пусть $k < n$.

$$X = \begin{pmatrix} x_1^{(1)} & \cdots & x_1^{(k)} \\ \vdots & \ddots & \vdots \\ x_n^{(1)} & \cdots & x_n^{(k)} \end{pmatrix}; \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

.

В модели с константой первый столбец состоит из единицы: $x_i^{(1)} = 1$.

Пусть мы имеем модель и ее оценку:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}. \tag{1.15}$$

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}. \tag{1.16}$$

Теорема I.3 (Теорема Гаусса-Маркова)**.** Если выполнены условия:

1. Модель линейна по параметрам и правильно специфицирована: $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \varepsilon$;
2. $\mathbf{X}$ - детерминированная матрица имеющая максимальный ранг k;
3. Математическое ожидание случайных ошибок равно нулю: $\mathrm{E}(\boldsymbol{\varepsilon}) = \overline{0}$;
4. $V(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I_n}$;

то вектор оценок $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'y$, полученный по МНК,

1. несмещенный;
2. эффективный.

1.32 Вопрос 32. Стандартная ошибка регрессии

Определение 1.32.1. Стандартная ошибка регрессии

$$SEE \stackrel{\text{def}}{=} \sqrt{S^2} = \sqrt{\frac{\sum e_i^2}{n-k}}$$

.

Стандартная ошибка регрессии

- Измеряет среднюю величину ошибки модели;
- Используется для оценки качества подгонки модели: чем меньше SEE, тем точнее модель;
- Можно использовать для сравнения нескольких однотипных уравнений регрессии

1.33 Вопрос 33. Исправленный R^2

Как мы помним из определения 1.21.1, коэффициент детерминации R^2 показывает долю дисперсии зависимой переменной, объясненной уравнением регрессии:

$$R^2 = \frac{\widehat{\mathrm{Var}}(\hat{y})}{\widehat{\mathrm{Var}}(y)} = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS} = \frac{\frac{1}{n}\sum(\hat{y}-\overline{y})^2}{\frac{1}{n}\sum(y_i-\overline{y})^2} = 1 - \frac{\sum e_i^2}{\sum(y_i-\overline{y})^2}$$

Однако при добавлении в модель новых переменных R^2 не уменьшается, поэтому сравнение моделей с разным числом переменных на основе этого показателя некорректно. Таким образом, мы вводим исправленный коэффициент R^2_{adj}.

Определение 1.33.1. $R^2_{adj} \overset{\text{def}}{=} R^2 - \dfrac{k-1}{n-k}(1-R^2)$.

Во многих ситуациях уместнее использовать исправленных коэффициент:

Пример I.8. Пусть есть две модели с одной зависимой переменной, но с разным числом регрессоров:

$$\hat{y}_i = \beta_1 + \beta_2 \cdot x_i^{(2)} + \beta_3 \cdot x_i^{(3)} \tag{1.17}$$

$$\hat{y}_i = \beta_1 + \beta_2 \cdot x_i^{(2)} + \beta_3 \cdot x_i^{(3)} + ... + \beta_{3003} \cdot x_i^{(3003)} \tag{1.18}$$

Всего было сделано 4003 наблюдений, а для первой модели TSS = 1000, ESS=300. Пусть для второй модели ESS=200.

Тогда для первой $R^2 = 0.7$, а для второй - $R^2 = 0.8$. Кажется, что вторая модель лучше. Однако, если мы найдем исправленные коэффициенты, то мы увидим преимущество первой модели:

1. $R^2_{adj} = 0.7 - \dfrac{3-1}{4003-3}(1-0{,}7) = 0{,}69985$;
2. $R^2_{adj} = 0.8 - \dfrac{3003-1}{4003-3003}(1-0.8) = 0{,}1996$.

Таким образом, первая модель лучше.

1.34 Вопрос 34. Значимость коэффициента

Пусть мы имеем модель и ее оценку:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}. \tag{1.19}$$

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}. \tag{1.20}$$

Нас интересует вопрос о значимости коэффициента при переменной $x^{(k)}$.

Для получения ответа на этот вопрос мы формулируем две конкурирующие гипотезы H_0 и H_1.

$H_0 : \beta_k = 0$
$H_1 : \beta_k \neq 0$

Как мы помним: вектор оценок β, полученный по МНК, имеет вид

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'y$$

.

Подставим назад, чтобы получить альтернативное выражение для нашего вектора:

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\boldsymbol{X'X})^{-1}\boldsymbol{X'\varepsilon}$$

Если выполнены предпосылки

- x_i - детерминированные величины;
- $\varepsilon_i \sim N(0; 1)$,

то $\hat{\boldsymbol{\beta}}$ - вектор ЛК нормальных случайных величин.

Таким образом, $\hat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta}; \sigma^2 (\boldsymbol{X'X})^{-1}\right)$

Это значит, что величина $\dfrac{\hat{\beta}_k - \beta_k}{se(\hat{\beta}_k)} \in t_{n-k}$

В рамках гипотезы H_0 предполагается, что $\beta_k = 0$. Таким образом, $t_{\text{расч}} = \dfrac{\hat{\beta}_k}{se(\hat{\beta}_k)}$.

Мы выбираем необходимый нам уровень значимости, после чего из таблицы распределения Стьюдента мы выбираем критическое значение статистики t для двустороннего интервала.

Мы сравниваем $|t_{\text{расч}}|$ и $t_{\text{кр}}$. Если $|t_{\text{расч}}| > t_{\text{кр}}$, то мы отклоняем нулевую гипотезу: с уверенностью $100-\alpha\%$ мы можем утверждать, что переменная $x^{(k)}$ значимо влияет на y. В альтернативном случае мы принимаем нулевую гипотезу: переменная $x^{(k)}$ не оказывает значимого влияния на y.

1.35 Вопрос 35. Значимость уравнения

Пусть мы имеем модель и ее оценку:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}. \tag{1.21}$$

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}. \tag{1.22}$$

Нас интересует вопрос о значимости всего уравнения в целом. Формулируем наши гипотезы:

$H_0 : \boldsymbol{\beta} = \overline{0}$
$H_1 : \boldsymbol{\beta} \neq \overline{0}$

Строим F-статистику:

$$F_{\text{расч}} = \frac{R^2}{1-R^2}\frac{n-k}{k-1} = \frac{RSS}{ESS}\frac{n-k}{k-1} = \frac{\sum(\hat{y}_i - \overline{y})^2/(k-1)}{\sum e_i^2/(n-k)} =$$

$$= \frac{\dfrac{\hat{\boldsymbol{y}}'_*\hat{\boldsymbol{y}}_*}{\sigma^2}\dfrac{1}{k-1}}{\dfrac{\boldsymbol{e}'\boldsymbol{e}}{\sigma^2}\dfrac{1}{n-k}}$$

- Числитель $\dfrac{\hat{\boldsymbol{y}}'_*\hat{\boldsymbol{y}}_*}{\sigma^2}\dfrac{1}{k-1}$ имеет распределение $\dfrac{1}{k-1}\chi^2(k-1)$.
- Числитель $\dfrac{\boldsymbol{e}'\boldsymbol{e}}{\sigma^2}\dfrac{1}{n-k}$ имеет распределение $\dfrac{1}{n-k}\chi^2(n-k)$

Если выполняются преподсылки (1)-(6), то $\hat{\boldsymbol{\beta}}$ и $\mathbf{e}$ независимы, поэтому статистика F имеет распределение Фишера:

$$F_{\text{расч}} = \frac{R^2}{1-R^2}\frac{n-k}{k-1} \sim F(k-1,n-k).$$

Используем ее для проверки нашей гипотезы.

Мы выбираем необходимый нам уровень значимости, после чего из таблицы распределения Фишера мы выбираем критическое значение статистики F, соответствующее нашему уровню значимости и $(k-1)$ и $(n-k)$ степеням свободы .

Мы сравниваем $F_{\text{расч}}$ и $F_{\text{кр}}$. Если $F_{\text{расч}} > F_{\text{кр}}$, то мы отклоняем нулевую гипотезу: с уверенностью $100-\alpha\%$ мы можем утверждать, что уравнение значимо и хоть одна переменная $x^{(k)}$ значимо влияет на y. В альтернативном случае мы принимаем нулевую гипотезу: уравнение незначимо, а все переменные $x^{(k)}$ не оказывают значимого влияния на y.

1.36 Вопрос 36. Тест на короткую и длинную регрессию

Пусть мы имеем две модели: "короткую" и "длинную" регрессии:

$$y_i = \beta_1 + \beta_2 \cdot x_i^{(2)} + ... + \beta_k \cdot x_i^{(k)} + \varepsilon_i \tag{1.23}$$

$$y_i = \beta_1 + \beta_2 \cdot x_i^{(2)} + ... + \beta_k \cdot x_i^{(k)} + ... + \beta_{k+q} \cdot x_i^{(k+q)} + \varepsilon_i \tag{1.24}$$

Нас интересует, значимы ли коэффициенты при добавленных в длинную регрессию переменных. Чтобы получить ответ, формулируем конкурирующие гипотезы:

$H_0 : \beta_{k+1} = ... = \beta_{k+q} = 0;$
$H_1 : \exists j \in (k+1, ..., k+q) : \beta_j \neq 0.$

Введем новые обозначения и построим статистику: пусть R_R^2 – коэффициент детерминации R^2 короткой регрессии, а R_{UR}^2 – коэффициент детерминации R^2 длинной регрессии.

Тогда $F_{\text{расч}} = \dfrac{(R_{UR}^2 - R_R^2)/q}{(1-R_{UR}^2)/(n-(k+q))}$.

- Числитель $(R_{UR}^2 - R_R^2)/q$ имеет распределение $\dfrac{1}{q}\chi^2(q)$.
- Знаменатель $(1-R_{UR}^2)/(n-(k+q))$ имеет распределение $\dfrac{1}{n-(k+q)}\chi^2(n-(k+q))$

Если выполняются препосылки (1)-(6), то статистика F имеет распределение Фишера:

$$F_{\text{расч}} = \frac{(R^2_{UR} - R^2_R)/q}{(1 - R^2_{UR})/(n - (k + q))} \sim F(q, n - k - q).$$

Используем ее для проверки нашей гипотезы.

Мы выбираем необходимый нам уровень значимости, после чего из таблицы распределения Фишера мы выбираем критическое значение статистики F, соответствующее нашему уровню значимости и (q) и $(n - k - q)$ степеням свободы .

Мы сравниваем $F_{\text{расч}}$ и $F_{\text{кр}}$. Если $F_{\text{расч}} > F_{\text{кр}}$, то мы отклоняем нулевую гипотезу: с уверенностью $100 - \alpha\%$ мы можем утверждать, что длинная регрессия лучше: хоть одна переменная $x^{(j)}, j \in (k + 1, ..., k + q)$, значимо влияет на у. В альтернативном случае мы принимаем нулевую гипотезу: все добавленные в уравнение переменные незначимы.

1.37 Вопрос 37. Связь $F_{\text{расч}}$ и $t_{\text{расч}}$

Рассматривается классическая линейная модель парной регрессии с детерминированным регрессором:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i \tag{1.25}$$

Утверждение 1.37.1. Расчетное значение статистики для теста на значимость уравнения в целом (F) и расчетное значение статистики для теста на значимость коэффициента при переменной (t) соотносятся следующим образом:

$$F_{\text{расч}} = t^2_{\text{расч}}.$$

Доказательство. Как мы помним, для проверки значимости коэффициента при переменной мы имели гипотезы:

$H_0 : \beta_2 = 0$
$H_1 : \beta_2 \neq 0$

Мы помним, что $\text{Var}(\hat{\beta}_2) = \dfrac{\sigma^2}{\sum(x_i - \overline{x})^2} \approx \dfrac{S^2}{\sum(x_i - \overline{x})^2}$

При этом t-статистика имела вид $t_{\text{расч}} = \dfrac{\hat{\beta}_2}{se(\hat{\beta}_2)} = \dfrac{\hat{\beta}_2\sqrt{\sum(x_i - \overline{x})^2}}{\sqrt{S^2}}$.

Для проверки значимости уравнения в целом мы формулировали гипотезы:

$H_0 : \hat{\boldsymbol{\beta}} = \overline{0}$
$H_1 : \hat{\boldsymbol{\beta}} \neq \overline{0}$.

F-статистика имела вид: $F_{\text{расч}} = \dfrac{R^2}{1-R^2}\dfrac{n-k}{k-1} = \dfrac{RSS}{ESS}\dfrac{n-k}{k-1}$.
Так мы работаем с моделью парной регрессии, k=2.
Преобразуем нашу F-статистику:

$$F_{\text{расч}} = \frac{R^2}{1-R^2}\frac{n-k}{k-1} = \frac{RSS}{ESS}\frac{n-k}{k-1} = \frac{RSS}{ESS}\frac{n-2}{1} =$$

$$RSS\frac{1}{\dfrac{1}{n-2}ESS} = \frac{RSS}{S^2} = \frac{\sum(\hat{y}_i - \overline{y})^2}{S^2}.$$

Сделаем некоторые преобразования:

$$RSS = \sum(\hat{y}_i - \overline{y})^2 = n\frac{\sum(\hat{y}_i - \overline{y})^2}{n} =$$
$$= n\,\widehat{\mathrm{Var}}(\hat{y}) = n\,\widehat{\mathrm{Var}}(\hat{\beta}_1 + \hat{\beta}_2 x) = \hat{\beta}_2^{\,2} n\,\widehat{\mathrm{Var}}(x) =$$
$$= \hat{\beta}_2^{\,2} n\frac{\sum(x_i - \overline{x})^2}{n} = \hat{\beta}_2^{\,2}\sum(x_i - \overline{x})^2.$$

Это значит, что

$$F_{\text{расч}} = \frac{RSS}{S^2} = \frac{\hat{\beta}_2^{\,2}\sum(x_i - \overline{x})^2}{S^2}.$$

Тем самым,

$$F_{\text{расч}} = t^2_{\text{расч}}.$$

□

1.38 Вопрос 38. Связь SEE и R^2

Исследователь сравнивает две модели множественной регрессии, оцененные на одинаковых данных и с одинаковой зависимой переменной. При этом наборы регрессоров для двух моделей различаются.

Утверждение 1.38.1. Верно, что не имеет значения, на основе какого из показателей делать выбор: скорректированного коэффициента R^2_{adj} или стандартной ошибки регрессии.

Доказательство. Пусть мы имеем модели A и B, построенные на одинаковых данных. Пусть в модели A k регрессоров, а в модели B m регрессоров.

Предположим, что $SEE_A < SEE_B$.

$$\sqrt{\frac{ESS_A}{n-k}} < \sqrt{\frac{ESS_B}{n-m}}$$

Преобразуем выражение для R^2_{adj}:

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

$$= \frac{TSS(n-k) - ESS(n-k) - ESS(k-1)}{(n-k)TSS} = 1 - \frac{(n-1)}{TSS}\frac{ESS}{n-k}.$$

TSS и число наблюдений равны для обеих моделей. Таким образом, R^2_{adj} представляет собой линейную функцию от $\frac{ESS}{n-k}$. Мы можем заметить, что это в точности выражение под корнем в SEE.

Мы видим, что $R^2_{adj;A} > R^2_{adj;B}$.

Аналогично показывается, что, если $R^2_{adj;A} > R^2_{adj;B}$, то $SEE_A < SEE_B$.. Значит, не имеет значения, на основе какого из показателей делать выбор.

□

1.39 Вопрос 39. Мультиколлинеарность

Определение 1.39.1. Строгая мультиколлинеарность - наличие линейной функциональной связи между объясняющими переменными.

В случае строгой мультиколлинеарности оценка коэффициентов модели при помощи метода наименьших квадратов невозможна.

Определение 1.39.2. Нестрогая мультиколлинеарность - наличие сильной линейной корреляционной связи между регрессорами.

В случае нестрогой мультиколлинеарности оценка возможна, но имеет ряд проблем:

1. Стандартные ошибки оценок коэффициентов оказываются высокими. Точность оценивания оказывается низкой;

2. Небольшое изменении исходных данных приводит к существенному изменению оценок коэффициентов;

3. Каждая переменная в отдельности является незначимой, а уравнение в целом имеет высокий R^2 и является значимым;

4. Оценки коэффициентов имеют неправильные с точки зрения экономической теории знаки или неоправданно большие значения.

1.40 Вопрос 40. Признаки мультиколлинеарности

Эти негативные последствия и можно использовать как признаки наличия мультиколлинеарности:

1. Небольшое изменении исходных данных приводит к существенному изменению оценок коэффициентов;

2. Каждая переменная в отдельности является незначимой, а уравнение в целом имеет высокий R^2 и является значимым;

3. Оценки коэффициентов имеют неправильные с точки зрения экономической теории знаки или неоправданно большие значения.

Существует способ выявления мультиколлинеарности: коэффициент **VIF** (variance inflation factor) характеризует силу мультиколлинеарности.

Он вычисляется на основе значений R^2 во вспомогательных регрессиях одного регрессора на другие:

$$x_i^{(k)} = \beta_1 + \beta_2 x_i^{(2)} + ... + \beta_{k-1} x_i^{(k-1)} + \varepsilon_i$$

Тогда:

Определение 1.40.1. VIF$=\dfrac{1}{1-R^2}$

Если он больше 10, то мы можем заключить, что мультиколлинеарность присутствует.

Можно предложить ряд способов устранения мультиколлинеарности:

1. Увеличить число наблюдений;

2. Исключить проблемные переменные;

3. Использовать альтернативные формы модели;

4. Агрегировать переменные.

1.41 Вопрос 41. Фиктивные переменные

Определение 1.41.1. Фиктивная переменная – бинарная переменная, включаемая в модель для отражения влияния событий и качественных признаков на объясняемую переменную.

Определение 1.41.2. Фиктивная переменная **сдвига** – фиктивная переменная, меняющая точку пересечения линии регрессии с осью ординат при наличии признака.

Определение 1.41.3. Фиктивная переменная **наклона** – фиктивная переменная, меняющая наклон линии регрессии при наличии признака.

Пример I.9. Пусть есть модель зависимости потребления от дохода:

$$C_i = C_a + mpcY_i + \varepsilon_i \tag{1.26}$$

Добавим в модель фиктивную переменную incomehi, равную единице, если семья располагает высоким доходом, и нулю в прочих случаях.

Если мы считаем, что домохозяйства с высоким доходом имеют более высокое автономное потребление, то мы имеет фиктивную переменную сдвига в модели:

$$C_i = C_a + \beta \cdot incomehi_i + mpcY_i + \varepsilon_i. \tag{1.27}$$

Если мы считаем, что домохозяйства с высоким доходом имеют другую предельную склонность к потреблению, то мы имеет фиктивную переменную наклона в модели:

$$C_i = C_a + \beta \cdot incomehi_i Y_i + mpcY_i + \varepsilon_i. \tag{1.28}$$

1.42 Вопрос 42. Существенные и несущественные переменные

Определение 1.42.1. Переменная **существенная**, если она должна быть включена в уравнение.

Определение 1.42.2. Переменная **несущественная**, если она не должна быть включена в уравнение.

Пример I.10 (Невключение существенной переменной)**.** Невключение существенной переменной имеет несколько отрицательных последствий:

1. Коэффициенты при оставшихся переменных могут оказаться смещенными – *omitted variable bias*;
2. Стандартные ошибки оставшихся переменных и другие показатели качества становятся некорректными и не могут быть использованы для суждения о качестве уравнения.

Рассмотрим конкретный пример: пусть истинная связь имеет вид:

$$y_i = \beta_1 + \beta_2 x_i^{(1)} + \beta_3 x_i^{(2)} + \varepsilon_i$$

При этом мы ошибочно не включили вторую переменную в нашу модель и получили оценку:

$$\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x_i.$$

Покажем смещенность коэффициента $\hat{\beta}_2$:

$$E(\hat{\beta}_2) = E\left(\frac{\operatorname{Cov}(x^{(1)}, y)}{\operatorname{Var}(x^{(1)})}\right) = E\left(\frac{\operatorname{Cov}(x^{(1)}, \beta_1 + \beta_2 x^{(1)} + \beta_3 x^{(2)} + \varepsilon)}{\operatorname{Var}(x^{(1)})}\right) =$$
$$= \frac{\beta_2 \operatorname{Cov}(x^{(1)}, x^{(1)}) + \beta_3 \operatorname{Cov}(x^{(1)}, x^{(2)})}{\operatorname{Var}(x^{(1)})} = \beta_2 + \frac{\beta_3 \operatorname{Cov}(x^{(1)}, x^{(2)})}{\operatorname{Var}(x^{(1)})}.$$

Пример I.11 (Включение несущественной переменной)**.** Включение несущественной переменной имеет более мягкие последствия:

1. Не теряется возможность правильной оценки и интерпретации уравнения;
2. Коэффициенты при переменных остаются несмещенными;
3. Стандартные ошибки растут, следовательно точность оценок падает;
4. Увеличивается риск мультиколлинеарности.

1.43 Вопрос 43. Контрольная переменная

Определение 1.43.1. Variable of interest – переменная, влияние которой на зависимую переменную нас интересует.

Определение 1.43.2. Контрольная переменная – переменная, которые мы включаем в модель для того, чтобы избежать смещения коэффициента при интересующей нас переменной.

Пример I.12 (Разница между variable of interest и control variable)**.** Пусть нас интересует вопрос наличия дискриминация на рынке труда.

Мы строим модель вида:

$$Wage_i = \beta_1 + \beta_2 G_i + \varepsilon_i \tag{1.29}$$

Здесь $G_i = \begin{cases} 0, & \text{если респондент i - мужчина;} \\ 1, & \text{если респондент i - женщина.} \end{cases}$ - variable of interest.

Но оценка $\hat{\beta}_2$ будет смещенной из-за omitted variable bias, так как на зарплату влияют многие факторы,

Чтобы избежать смещения оценок из-за пропуска существенных переменных, добавим в модель остальные факторы, которые влияют на зарплату :образование, возраст и так далее. Эти переменные - control variables.

$$Wage_i = \beta_1 + \beta_2 G_i + \beta_3 Education_i + \beta_4 Age_i + \varepsilon_i. \tag{1.30}$$

1.44 Вопрос 44. Замещающая переменная

Пусть истинная модель имеет вид:

$$y_i = \beta_1 + \beta_2 x_i^{(2)} + \beta_3 x_i^{(3)} + \varepsilon_i \tag{1.31}$$

Мы располагаем данными о $x^{(2)}, y, z$, но не о $x^{(3)}$. Однако мы знаем, что $\exists \mu \neq 0, \theta$:

$$x^{(3)} = \theta + \mu z. \tag{1.32}$$

Таким образом, z - *замещающая переменная.*

Утверждение 1.44.1. В этом случае можно получить несмещенную оценку для параметра β_2, но нельзя для β_3.

Доказательство. Согласно нашему условию:

$$y_i = \beta_1 + \beta_2 x_i^{(2)} + \beta_3 x_i^{(3)} + \varepsilon_i$$

$$x^{(3)} = \theta + \mu z.$$

Подставим выражения для $x^{(3)}$ в нашу модель:

$$y_i = \beta_1 + \beta_2 x_i^{(2)} + \beta_3(\theta + \mu z_i) + \varepsilon_i$$

$$y_i = (\beta_1 + \beta_3 \cdot \theta) + \beta_2 x_i^{(2)} + \beta_3 \cdot \mu z_i + \varepsilon_i$$

Сделаем замену обозначений:

$$y_i = \alpha_1 + \beta_2 x_i^{(2)} + \alpha_3 z_i + \varepsilon_i$$

Дальше мы можем получить МНК-оценку для $\alpha_1, \beta_2, \alpha_3$. Согласно теореме Гаусса-Маркова, она будет несмещенной.

Однако, так как мы не знаем μ, мы не можем получить оценку для β_3. □

1.45 Вопрос 45. Включение регрессора

Можно привести некоторые соображения, которые могут помочь принять решение о том, следует ли включать ту или иную переменную в уравнение регрессии.

1. Роль переменной должна обосновываться теоретически или исходя из сильной эмпирической интуиции и здравого смысла;

2. Переменная статистически значима;

3. Оценки других коэффициентов сильно меняются при включении новой переменной в модели: часто это значит, что до ее включения они страдали от omitted variable bias;

4. R^2_{adj} существенно увеличивается в результате включения переменной в модель.

1.46 Вопрос 46. Тест Рамсея

Самый простой способ проверить справедливость линейной модели - добавить в правую часть нелинейные члены и тестировать их значимость с помощью обычного F-теста.

$$y_i = \beta_1 + \beta_2 \cdot x_i^{(2)} + ... + \beta_k \cdot x_i^{(k)} + \varepsilon_i \tag{1.33}$$

Тест Рамсея (RESET – Regression Equation Specification Error Test) опирается на идею, что в верной модели добавление функций вида $\hat{y}_i = \sum \hat{\beta}_k x_i^{(k)}$ не должно помогать объяснять y. Можно добавлять степени для простоты: (обычно $m \in (2,3,4)$)

$$y_i = \sum \beta_k x_i^{(k)} + \sum_{m=2}^{p} \alpha_m \hat{y_i}^m + \varepsilon_i \tag{1.34}$$

Теперь мы можем сформулировать гипотезы:

$H_0 : \boldsymbol{\alpha} = \overline{0}$
$H_1 : \boldsymbol{\alpha} \neq \overline{0}$

Проверить эти гипотезы можно обыкновенным F-тестом, аналогичным тесту на "короткую" и "длинную" регрессию.

Если принимается нулевая гипотеза, то мы можем сделать вывод, что модель верна специфицирована. Нулевая гипотеза может быть отвергнута не только потому, что в истинной модели есть нелинейные члены, но и в силу того, что в исходном уравнении пропущена переменная, влияние которого частично учтено нелинейными членами $\hat{y}_i$.

1.47 Вопрос 47. J-тест

Пусть мы имеем две разных невложенных модели:

$$A : y = \beta_1 + \sum \beta_k x^{(k)} + \varepsilon \tag{1.35}$$

$$B : y = \alpha_1 + \sum \alpha_m z^{(m)} + \varepsilon \tag{1.36}$$

Чтобы выбрать одну из них, можно использовать J-тест.
Строится модель, включающая как частный случай модели A и B:

$$y = (1-\theta)(\beta_1 + \sum \beta_k x^{(k)}) + \theta(\alpha_1 + \sum \alpha_m z^{(m)}) + \varepsilon \tag{1.37}$$

При $\theta = 0$ модель совпадает с моделью A, а при $\theta = 1$ – с моделью B. Однако невозможно однозначно оценить параметры этого уравнения сразу. Мы будем использовать альтернативный подход.

На первом этапе:

1. Отдельно оценивается модель A;
2. По ней получаются прогнозные значения $\hat{y}_A$;
3. Оценивается модель:

$$y = \alpha_1 + \sum \alpha_m z^{(m)} + \theta \hat{y}_A + \varepsilon$$

4. Проверяем гипотезы:

$H_0^B : \theta = 0;$
$H_1^B : \theta \neq 0.$

На втором этапе:

1. Отдельно оценивается модель B;
2. По ней получаются прогнозные значения $\hat{y}_B$;
3. Оценивается модель:

$$y = \beta_1 + \sum \beta_k x^{(k)} + \theta \hat{y}_B + \varepsilon$$

4. Проверяем гипотезы:

$H_0^A : \theta = 0;$
$H_1^A : \theta \neq 0.$

Дальше возможны три ситуации:

- Если принята H_0^A, а H_0^B отклонена, то модель A лучше;

- Если принята H_0^B, а H_0^A отклонена, то модель B лучше;
- Если исходы тестирования совпадают, то ситуация остается неопределенной.

1.48 Вопрос 48. Тест Чоу

Пусть у нас есть некая модель. Мы предполагаем, что в данных происходит структурный сдвиг (при наличии какого-то признака). Чтобы проверить нашу гипотезу о наличии сдвига, мы проводим **тест Чоу**.

Рассмотрим на простом примере: пусть в оригинальной постановке мы рассматривали парную регрессию:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i \tag{1.38}$$

Для проверки мы разделяем наши данные на две части: до предполагаемого сдвига и после него (при наличии признака и без него). Вводим фиктивную переменную:

$$D_i = \begin{cases} 1, \text{если данные после предполагамого сдвига;} \\ 0, \text{если данные до предполагамого сдвига;} \end{cases}$$

Наша модель преображается:

$$\hat{y}_i = \beta_1 + \beta_2 x_i + \beta_3 D_i + \beta_4 D_i x_i \tag{1.39}$$

Чтобы проверить оправданность включения фиктивных переменных в нашу модель, проводим обычный тест на "короткую" и "длинную" регрессии. Формулируем конкурирующие гипотезы:

$H_0 : \beta_3 = \beta_4 = 0;$
$H_1 : \exists j \in (3{,}4) : \beta_j \neq 0.$

Подробно алгоритм описан в вопросе 1.36.

Альтернативно тест Чоу можно провести с помощью двух моделей.
Разбиваем нашу выборку на две подвыборки: до сдвига/после сдвига (с признаком/без признака).
При этом наши модели имеют один набор регрессоров.

$$A : y = \beta_1 + \sum \beta_k x^{(k)} + \varepsilon \tag{1.40}$$

$$B : y = \alpha_1 + \sum \alpha_k x^{(k)} + \varepsilon \tag{1.41}$$

Формулируем конкурирующие гипотезы:

$H_0 : \forall i \ \alpha_i = \beta_i$;
$H_1 : \exists j \in (1,2...k) : \alpha_j \neq \beta_j$.

Для такого теста мы вводим F-статистику:

$$F_{\text{расч}} = \frac{(ESS_{\sum} - ESS_A - ESS_B)/k}{(ESS_A + ESS_B)/(n-2k)} \sim F(k, n-2k)$$

При этом $ESS_{\sum}$ – ESS, вычисленная по общей модели без разделения на подвыборки, а ESS_A, ESS_B – на основе оцененных по подвыборкам моделям.

Мы сравниваем $F_{\text{расч}}$ и $F_{\text{кр}}(\alpha,k,n-2k)$. Если $F_{\text{расч}} > F_{\text{кр}}$, то мы отклоняем нулевую гипотезу: с уверенностью $(100-\alpha)\%$ мы можем утверждать, что в данных есть структурный сдвиг. В альтернативном случае мы принимаем нулевую гипотезу: Данные однородны. Структурного сдвига в данных нет.

1.49 Вопрос 49. Сравнение Бокса-Кокса

Иногда перед нами стоит необходимость сравнения формы моделей. Рассмотрим две проблемы здесь.

I. Нам нужно выбрать одну из двух форм задания регрессора:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$$

Или же:

$$y_i = \beta_1 + \beta_2 \ln x_i + \varepsilon_i$$

Это сравнение можно проводить, напрямую изучая стандартную ошибку или R^2, так как в зависимая переменная в них одна и та же.

II. Нам нужно выбрать форму задания зависимой переменной:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$$

Или же:

$$\ln y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

Сглаживающие свойства логарифма приводят к тому, что при больших значениях y SEE во второй модели всегда будет меньше, даже если она хуже подходит для описания связи.

Таким образом, мы используем **метод Бокса-Кокса**.

Опишем его алгоритм:

1. На первом этапе мы преобразуем истинные значения зависимой переменной:

$$y_i \to y_i^* = \frac{y_i}{(y_1 y_2 ... y_n)^{\frac{1}{n}}}$$

2. На втором этапе мы задаем две конкурирующие модели:

$$y_i^* = \beta_1 + \beta_2 x_i + \varepsilon_i$$

$$\ln y_i^* = \beta_1 + \beta_2 x_i + \varepsilon_i$$

3. Мы сравниваем ESS у полученных моделей и выбираем модель с меньшей суммой. После этого оцениваем избранную модель с исходными переменными.

Глава 2
Зачет

2.1 Вопрос 1. Предпосылки КЛММР. Теорема Гаусса-Маркова

Пусть X – матрица регрессоров, а y – вектор значенией объясняемой переменной, пусть $k < n$, пусть $\boldsymbol{\varepsilon}$ – вектор случайных ошибок.

$$X = \begin{pmatrix} x_1^{(1)} & \cdots & x_1^{(k)} \\ \vdots & \ddots & \vdots \\ x_n^{(1)} & \cdots & x_n^{(k)} \end{pmatrix}; \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

.

В модели с константой первый столбец состоит из единицы: $x_i^{(1)} = 1$.

Сформулируем еще раз предпосылки классической линейной модели множественной регрессии.

1. Модель линейна по параметрам и правильно специфицирована: $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$;

2. Матрица $\boldsymbol{X}$ — детерминированная матрица, имеющая максимальный ранг k;

3. $E\boldsymbol{\varepsilon} = \overline{0}$;

4. Ковариационная матрица случайных ошибок имеет следующий вид:

$$\operatorname{Var}(\boldsymbol{\varepsilon}) = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix}$$

5. $\boldsymbol{\varepsilon} \sim N(\overline{0}; \boldsymbol{I_n}\sigma^2)$

Теперь мы снова можем сформулировать теорему Гаусса-Маркова.

Теорема I.4 (Теорема Гаусса-Маркова)**.** Если выполнены условия:

1. Модель линейна по параметрам и правильно специфицирована: $\boldsymbol{y} = \boldsymbol{X\beta} + \varepsilon$;
2. **X** - детерминированная матрица имеющая максимальный ранг k;
3. Математическое ожидание случайных ошибок равно нулю: $\mathrm{E}(\boldsymbol{\varepsilon}) = \overline{0}$;
4. $\mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I_n}$;

то вектор оценок $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'y$, полученный по МНК,

1. несмещенный;
2. эффективный.

2.2 Вопрос 2. Вывод МНК-оценок

Пусть мы имеем модель $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$. Мы хотим получить ее оценку $\boldsymbol{y} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ с помощью метода наименьших квадратов.

$$ESS = \sum e^2 = \sum (y - \hat{y})^2 = \boldsymbol{e'e} = \longrightarrow \min$$

Распишем это несколько подробнее в матричном виде:

$$ESS = (\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta}) = \boldsymbol{y'y} - \boldsymbol{2y'X\beta} + \boldsymbol{\beta'X'X\beta} \rightarrow \min$$

Дифференцируем по коэффициентам:

$$\frac{\partial \boldsymbol{ESS}}{\partial \boldsymbol{\beta}} = \boldsymbol{-2X'y + 2X'X\beta = 0}$$

Отсюда мы получаем, что:

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'y$$

Очевидно, что мы нашли глобальный минимум суммы квадратов ошибок.

2.3 Вопрос 3. Несмещенность МНК-оценок

Покажем теперь несмещенность полученных оценок нашего вектора. Как мы помним, для модели $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ получается оценка:

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'y$$

Подставляем в нее саму модель и получаем, что:

$$\hat{\boldsymbol{\beta}} = \beta + (X'X)^{-1}X'\boldsymbol{\varepsilon}$$

Утверждение 2.3.1. $\hat{\boldsymbol{\beta}}$ – несмещенная оценка вектора коэффициентов

Доказательство. Как мы уже получили,

$$\hat{\boldsymbol{\beta}} = \beta + (X'X)^{-1}X'\boldsymbol{\varepsilon}$$

Используя то, что у нас пока детерминированные регрессоры, возьмем матожидание этого выражения:

$$E(\hat{\boldsymbol{\beta}}) = E(\beta + (X'X)^{-1}X'\boldsymbol{\varepsilon}) =$$
$$= E(\boldsymbol{\beta}) + (X'X)^{-1}X'E(\boldsymbol{\varepsilon}) = \boldsymbol{\beta}$$

□

2.4 Вопрос 4. Ковариационная матрица МНК-оценок

Рассмотрим теперь вопрос о виде ковариационной матрицы вектора оценок коэффициентов в модели:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Утверждение 2.4.1. В гомоскедастичном случае с детерминированными регрессорами ковариационная матрица вектора оценок коэффициентов имеет вид $\mathbf{Var}\,(\hat{\boldsymbol{\beta}}) = (X'X)^{-1}\sigma^2$.

Доказательство. Как мы помним,

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'y$$

Подставляем в нее саму модель и получаем, что:

$$\hat{\boldsymbol{\beta}} = \beta + (X'X)^{-1}X'\boldsymbol{\varepsilon}$$

Теперь найдем матрицу:

$$\operatorname{Var}(\hat{\boldsymbol{\beta}}) = \operatorname{Var}(\beta + (X'X)^{-1}X'\boldsymbol{\varepsilon}) = \operatorname{Var}((X'X)^{-1}X'\boldsymbol{\varepsilon}) =$$

Убирая константу под дисперсией, получаем последнее выражение сверху. Выносим константу из-под знака дисперсии в квадрате. В данном случае выносится матрица:

$$= ((X'X)^{-1}X') \operatorname{Var}(\boldsymbol{\varepsilon})((X'X)^{-1}X')' =$$

Согласно предпосылкам, вставляем ковариационную матрицу вектора стандартных ошибок:

$$= ((X'X)^{-1}X')\sigma^2 I_n((X'X)^{-1}X')' =$$

Наконец, сокращаем лишние матрицы:

$$= (X'X)^{-1}X'X(X'X)^{-1}\sigma^2 = (X'X)^{-1}\sigma^2$$

□

2.5 Вопрос 5. Ковариация случайных ошибок

Рассмотрим теперь вопрос о ковариационной матрице вектора случайных ошибок:

$$\operatorname{Var}(\boldsymbol{\varepsilon}) = \begin{pmatrix} \operatorname{Cov}(\varepsilon_1, \varepsilon_1) & \operatorname{Cov}(\varepsilon_1, \varepsilon_2) & \cdots & \operatorname{Cov}(\varepsilon_1, \varepsilon_n) \\ \operatorname{Cov}(\varepsilon_2, \varepsilon_1) & \operatorname{Cov}(\varepsilon_2, \varepsilon_2) & \cdots & \operatorname{Cov}(\varepsilon_2, \varepsilon_n) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{Cov}(\varepsilon_n, \varepsilon_1) & \operatorname{Cov}(\varepsilon_n, \varepsilon_2) & \cdots & \operatorname{Cov}(\varepsilon_n, \varepsilon_n) \end{pmatrix}$$

Альтернативно, с опорой на предпосылку о нулевом матожидании случайных ошибок, мы можем записать матрицу через матожидание:

$$\operatorname{Var}(\boldsymbol{\varepsilon}) = \begin{pmatrix} E(\varepsilon_1\varepsilon_1) & \cdots & E(\varepsilon_1\varepsilon_n) \\ \vdots & \ddots & \vdots \\ E(\varepsilon_1\varepsilon_n) & \cdot & E(\varepsilon_n\varepsilon_n) \end{pmatrix}$$

Ковариационные матрицы в целом имеют ряд чудесных свойств. Ограничимся здесь теми, которые присущи матрице от случайных ошибок:

Свойство 2.5.1. Она симметрична

Свойство 2.5.2. Она положительно определена[1]

Свойство 2.5.3. $\operatorname{Var}(A\boldsymbol{\varepsilon} + a) = A \operatorname{Var}(\boldsymbol{\varepsilon})A'$

[1] Вообще говоря, ковариационные матрицы положительно *полуопределены*. Вопрос строгости определения сводится к наличию нулевого собственного значения. Как известна, матрица обратима iff нет собственного значения 0.

2.6 Вопрос 6. Ковариация оценок коэффициентов

Аналогичным образом мы получаем и ковариационную матрицу для наших оценок коэффициентов:

$$\operatorname{Var}(\hat{\boldsymbol{\beta}}) = \begin{pmatrix} \operatorname{Cov}(\hat{\beta}_1,\hat{\beta}_1) & \operatorname{Cov}(\hat{\beta}_1,\hat{\beta}_2) & \cdots & \operatorname{Cov}(\hat{\beta}_1,\hat{\beta}_n) \\ \operatorname{Cov}(\hat{\beta}_2,\hat{\beta}_1) & \operatorname{Cov}(\hat{\beta}_2,\hat{\beta}_2) & \cdots & \operatorname{Cov}(\hat{\beta}_2,\hat{\beta}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{Cov}(\hat{\beta}_n,\hat{\beta}_1) & \operatorname{Cov}(\hat{\beta}_n,\hat{\beta}_2) & \cdots & \operatorname{Cov}(\hat{\beta}_n,\hat{\beta}_n) \end{pmatrix}$$

Для нее мы получаем аналогичные свойства:

Свойство 2.6.1. Она симметрична

Свойство 2.6.2. Она положительно полуопределена.

Свойство 2.6.3. $\operatorname{Var}(A\hat{\boldsymbol{\beta}} + a) = A\operatorname{Var}(\hat{\boldsymbol{\beta}})A'$

2.7 Вопрос 7. Оценка при линейной гетероскедастичности

Рассматривается модель:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$$

Для нее выполнены все предположения классической линейной модели парной регрессии, за одним исключением: дисперсии ошибок удовлетворяют условию $\sigma_i^2 = \alpha_1 + \alpha_2 x_i$. В два шага мы хотим получить асимптотически эффективные МНК-оценки для $\boldsymbol{\beta}$.

$\boxed{\text{I}}$ МНК-оцениваем нашу модель. Получаем остатки регрессии e_i.
На их основе оцениваем модель:

$$e_i^2 = \alpha_1 + \alpha_2 x_i + u_i$$

Прогнозные значения обозначаем $\hat{\sigma}_i^2$.

$\boxed{\text{II}}$ Взвешиваем нашу модель по полученным оценкам:

$$\frac{y_i}{\hat{\sigma}_i} = \beta_1 \frac{1}{\hat{\sigma}_i} + \beta_2 \frac{x_i}{\hat{\sigma}_i} + \frac{\varepsilon_i}{\hat{\sigma}_i}$$

2.8 Вопрос 8. Тест Уайта

Пусть мы рассматриваем некоторую модель:

$$y_i = \beta_1 + \beta_2 \cdot x_i^{(2)} + ... + \beta_k \cdot x_i^{(k)} + \varepsilon_i$$

Мы подозреваем в ней наличии гетероскедастичности. Для проверки этого подозрения мы используем **тест Уайта**. Он хорош тем, что не требует от нас знания формы гетероскедастичности.

Сформулируем его гипотезы:

$$H_0 : \sigma_1^2 = ... = \sigma_n^2$$

$$H_1 : \exists i,j, i \neq j : \sigma_i^2 \neq \sigma_j^2$$

Опишем его алгоритм:

1. МНК-оцениваем нашу регрессию. Из этой оценки получаем остатки регрессии e_i^2.

2. После этого мы строим регрессию на константу, все регрессоры, их квадраты[2] и их попарные произведения:

$$e_i^2 = a_0 + \sum a_i x_i + \sum a_j x_j^2 + \sum_{k \neq l} \sum x_k x_j.$$

   Всего в этом процессе имеем p регрессоров, не включая константу. Оценив регрессию, берем ее R^2. Статистика теста – $\chi^2_{\text{расч}} = nR^2$, где n – число наблюдений.

3. Сравниваем расчетное значение с критическим – $\chi^2_{\text{кр}}(p, \alpha)$. Если расчетное значение больше, то отвергаем нулевую гипотезу и делаем вывод о наличии гетероскедастичности.

2.9 Вопрос 9. Тест Голдфелда-Квандта

Пусть у нас есть некоторая модель:

$$y_i = \beta_1 + \beta_2 \cdot x_i^{(2)} + ... + \beta_k \cdot x_i^{(k)} + \varepsilon_i$$

[2]Кроме фиктивных переменных.

Мы подозреваем, что у нас есть гетероскедастичность специфичного вида: дисперсия ошибки пропорциональна одной из независимых переменных.

Чтобы проверить такие подозрения, мы используем **тест Голдфелда-Квандта**. Он весьма ограничен в своей полезности. Сформулируем его гипотезы и опишем алгоритм.

$$H_0 : \sigma_1^2 = ... = \sigma_n^2$$

H_1 : Дисперсия ошибок пропорциональна одной из независимых переменных

1. Для начала мы сортируем наши данные, располагая их по возрастанию подозреваемой переменной. Делим данные на три равные[3] или почти равные части так, чтобы число наблюдений в первой и третьей были равны.

2. Оцениваем нашу модель на первой трети данных. Получаем ESS_1.

3. Оцениваем нашу модель на третьей трети данных. Получаем ESS_3.

4. Конструируем расчетную статистику – $F_{\text{расч}} = \frac{ESS_{\text{больший}}}{ESS_{\text{меньший}}}$. В наших предпосылках он имеет распеделение Фишера.

5. Сравниваем с критическим значением $F_{\text{кр}}$(число наблюдений в крайних кусках−число регрессоров, число наблюдений в крайних кусках−число регрессоров).

2.10 Вопрос 10. Тест Бреуша-Пагана

Пусть у нас есть какая-то модель:

$$y_i = \beta_1 + \beta_2 \cdot x_i^{(2)} + ... + \beta_k \cdot x_i^{(k)} + \varepsilon_i$$

Мы имеем подозрения, что в эту модель закралась гетероскедастичность. При этом нам кажется, что она линейно зависит от некоторого набора переменных. Для проверки этого подозрения используем **тест Бреуша-Пагана**. С ним следует быть аккуратным при нарушении предпосылки нормальности.

Его гипотезы:

$$H_0 : \sigma_1^2 = ... = \sigma_n^2$$

[3] Можно и чуть либеральнее.

$$H_1 : \sigma_i^2 = a_0 + a_1 z_i^{(1)} + ... + a_p z_i^{(p)}$$

Опишем его процесс:

1. МНК-оцениваем нашу исходную модель. На ее основе мы получаем остатки регрессии e_i. Возводим их в квадрат.

2. Вычисляем оценку дисперсии: $\tilde{\sigma}^2 = \dfrac{1}{n}\sum e_i^2$.

3. Оцениваем вспомогательную регрессию:

$$\frac{e_i^2}{\tilde{\sigma}^2} = a_0 + a_1 z_i^{(1)} + ... + a_p z_i^{(p)}$$

4. На основе этой регрессии мы получаем RSS. $\chi^2_{\text{расч}} = \dfrac{RSS}{2}$.

5. Сравниваем с критическим значением $\chi^2_{\text{кр}}(p, \alpha)$.

2.11 Вопрос 11. Ложная гетероскедастичность

Истинная гетероскедастичность – непостоянность дисперсий случайных ошибок. Однако иногда в модели может не быть гетероскедастичности, а формальные тесты будут показывать ее наличие.

Такая ситуация может возникнуть при неверной спецификации уравнения регрессии.

Есть несколько отличий:

- Истинная гетероскедастичность делает тесты некорректными из-за проблем стандартных ошибок, но оставляет коэффициенты несмещенными и асимптотически нормальными.

- Ложная гетероскедастичность не вызывает таких проблем, так как они уже вызваны неверной спецификацией. При этом нарушается предпосылка верной спецификации, что ведет к еще большим проблемам.

2.12 Вопрос 12. Обобщенная модель ЛМР

Пусть X – матрица регрессоров, а y – вектор значенией объясняемой переменной, пусть $k < n$, пусть $\boldsymbol{\varepsilon}$ – вектор случайных ошибок.

$$X = \begin{pmatrix} x_1^{(1)} & \cdots & x_1^{(k)} \\ \vdots & \ddots & \vdots \\ x_n^{(1)} & \cdots & x_n^{(k)} \end{pmatrix}; \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

.

В модели с константой первый столбец состоит из единицы: $x_i^{(1)} = 1$.

Обобщим нашу модель множественной регрессии. Ее предпосылки:

1. Модель линейна по параметрам и правильно специфицирована: $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$;
2. Матрица $\boldsymbol{X}$ — детерминированная матрица, имеющая максимальный ранг k;
3. $E\boldsymbol{\varepsilon} = \overline{0}$;
4. $\mathrm{Var}(\boldsymbol{\varepsilon}) = \Omega$ – положительно определенная симметричная матрица.

Таким образом, мы избавились от предпосылок гомоскедастичности и некоррелированности ошибок.

Мы формулируем такие условия потому, что МНК и ВМНК могут быть давать неэффективные и даже неверные результаты в таких условиях.

2.13 Вопрос 13. Теорема Айткена

Сформулируем теорему, доказанную Александром Айткеном в 1934 году.

Теорема I.5 (Теорема Айткена)**.** Если

1. Модель линейна по параметрам и правильно специфицирована: $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$;
2. Матрица $\boldsymbol{X}$ — детерминированная матрица, имеющая максимальный ранг k;
3. $E\boldsymbol{\varepsilon} = \overline{0}$;

4. $\mathrm{Var}(\boldsymbol{\varepsilon}) = \Omega$ – положительно определенная симметричная матрица,

то ОМНК-оценка $\hat{\boldsymbol{\beta}}^{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$ – несмещенная и эффективная в классе линейных по y оценок.

Доказательство. Доказательство теорема Айткена заключается в элегантной замене переменных, сводящих нашу модель к модели, удовлетворяющей всем условиям теоремы Гаусса-Маркова, в которой утверждаются и доказываются несмещенность и эффективность.

$\boxed{\text{а}}$ Согласно условию, Ω – положительно определенная симметричная матрица. Тогда Ω^{-1} существует, так как нет нулевых собственных значений, и тоже положительно определена, так как все собственные значения обратны.

Таким образом, мы можем разложить матрицу Ω^{-1} на произведение $\Omega^{-1} = P'P$. Например, разложение Холецкого дает уникальное такое разложение на пару обратимых матриц.

$\boxed{\text{б}}$ Делаем замену переменных, умножая на P:

$$y = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$Py = PX\boldsymbol{\beta} + P\boldsymbol{\varepsilon}$$

$\boxed{\text{в}}$ В этой модели выполнены предпосылки (4) и (5) теоремы Гаусса-Маркова:

4. $E(P\boldsymbol{\varepsilon}) = PE(\boldsymbol{\varepsilon}) = P\overline{0} = \overline{0}$

5. $\mathrm{Var}(P\boldsymbol{\varepsilon}) = P\,\mathrm{Var}(\boldsymbol{\varepsilon})P' = P\Omega P' = P\Omega\Omega^{-1}P^{-1} = I_n$.

$\boxed{\text{г}}$ Значит, для этой модели выполняются предпосылки теоремы Гаусса-Маркова. Получив оценку, можем перейти назад к исходной модели.

Оценка имеет вид:

$$\hat{\boldsymbol{\beta}}^{GLS} = ((PX)'PX)^{-1}(PX)'Py = (X'P'PX)^{-1}X'P'Py = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$

□

2.14 Вопрос 14. Ковариационная матрица коэффициентов в ОЛММР

Утверждение 2.14.1. $\mathrm{Var}(\hat{\boldsymbol{\beta}}^{GLS}) = (X'\Omega^{-1}X)^{-1}$[4]

Доказательство. Оценка имеет вид:

$$\hat{\boldsymbol{\beta}}^{GLS} = ((PX)'PX)^{-1}(PX)'Py = (X'P'PX)^{-1}X'P'y = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$

Чтобы показать справедливость утверждения, сделаем обычные преобразования[5]:

$$\hat{\boldsymbol{\beta}}^{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}(X\boldsymbol{\beta}+\boldsymbol{\varepsilon}) = \boldsymbol{\beta}+(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\boldsymbol{\varepsilon}$$

Теперь рассмотрим чистый хвост. Мы понимаем, что матожидание хвоста равно нулю:

$$\hat{\boldsymbol{\beta}}^{GLS} - \boldsymbol{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\boldsymbol{\varepsilon}$$

Теперь найдем ковариационную матрицу. Используем симметричность $X'\Omega^{-1}X$[6]:

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\beta}}^{GLS}) &= E\left((X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}\right) = \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Omega\Omega^{-1}X(X'\Omega^{-1}X)^{-1} = (X'\Omega^{-1}X)^{-1}\end{aligned}$$

□

2.15 Вопрос 15. Взвешенный МНК

Пусть есть некоторая модель:

$$y_i = \beta_1 + \beta_2 \cdot x_i^{(2)} + ... + \beta_k \cdot x_i^{(k)} + \varepsilon_i$$

---

[4] Для МНК-оценок матрица найдена на странице 63

[5] Из них еще видна несмещенность.

[6] Достаточно транспонировать:

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

Мы находимся в предпосылках КЛММР, отказавшись лишь от гомоскедастичности. В определенных условиях мы можем создать систему весов, ликвидирующую проблему гетероскедастичности[7].

$\boxed{\text{а}}$ Если мы знаем дисперсию каждой случайной ошибки, то мы можем разделить каждое наблюдение на соответствующее стандартное отклонение. Тогда мы получаем модель с гомоскедастичностью:

$$\frac{y_i}{\sigma_i} = \frac{\beta_1}{\sigma_i} + \beta_2 \frac{x_i^{(1)}}{\sigma_i} + \ldots$$

После такого преобразования, мы получаем все хорошие результаты, гарантируемые нам предпосылками.

$\boxed{\text{б}}$ Пусть мы знаем, что дисперсии случайных ошибок пропорциональны какой-то наблюдаемой переменной: $\sigma_i^2 = aw^n$.

Тогда мы делаем похожую замену переменных, переходя к модели с гомоскедастичностью:

$$\frac{y_i}{w^{\frac{n}{2}}} = \frac{\beta_1}{w^{\frac{n}{2}}} + \beta_2 \frac{x_i^{(1)}}{w^{\frac{n}{2}}} + \ldots$$

$\boxed{\text{в}}$ Пусть мы знаем, что наши дисперсии зависят от какого-то набора переменных. Тогда мы можем пытаться их оценить:

1. МНК-оцениваем нашу исходную модель. На ее основе мы получаем остатки регрессии e_i. Возводим их в квадрат.
2. Оцениваем вспомогательную регрессию[8]:

$$e_i^2 = a_0 + a_1 z_i^{(1)} + \ldots + a_p z_i^{(p)}$$

3. Получаем оценки в этой регрессии. По ней строим оценки остатков $\hat{\sigma}_i^2$.
4. Переходим к новой модели:

$$\frac{y_i}{\hat{\sigma}_i^2} = \frac{\beta_1}{\hat{\sigma}_i^2} + \beta_2 \frac{x_i^{(1)}}{\hat{\sigma}_i^2} + \ldots$$

[7]Но только при верном определении ее формы.

[8]Или оцениваем для логарифмов квадратов остатков

2.16 Вопрос 16. Доступный ОМНК

Пусть мы находимся в рамках предпосылок ОМНК. Однако очень часто мы не можем оценить матрицу Ω легко – это задача оценки треугольного блока, число элементов в котором заведомо превосходит число наблюдений.

Отсюда использование **доступного МНК**.

Его алгоритм прост:

- Мы делаем предположения о форме матрицы Ω.

  Пример I.13. Пусть $\sigma_i^2 = a^2 x_i^2$. Тогда матрица имеет вид:

$$\Omega = \begin{pmatrix} a^2x_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a^2x_n^2 \end{pmatrix}$$

  Вычисляем оценку $\hat{\Omega}$ для такой предположительной формы матрицы Ω.

  Вычисляем ОМНК-оценку:

$$\hat{\boldsymbol{\beta}}^{GLS} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y$$

2.17 Вопрос 17. Гетероскедастичность

Определение 2.17.1. Гетероскедастичность – случайные ошибки имеют непостоянную дисперсию: $\mathrm{Var}(\boldsymbol{\varepsilon}) \neq \sigma^2 I_n$.

Гетероскедастичность имеет ряд последствий, если мы проигнорируем ее наличие:

- МНК-оценки остаются несмещенными и асимптотически нормальными;
- МНК-оценки становятся неэффективными;
- Стандартные ошибки коэффициентов становятся смещенными и несостоятельными, поэтому тесты не работают корректно. Доверительные интервалы не строятся правильно.

Гетероскедастичность или хотя бы ее последствия можно устранить или использованием ВМНК, ОМНК, или переходом к робастным стандартным ошибкам.

О робастных ошибках. Вспомним, что ковариационная матрица случайных ошибок все еще имеет диагональный вид для нас, но элементы на диагонали мы не знаем. Однако мы можем получить состоятельную оценку:

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = (X'X)^{-1}(\sum \sigma_i^2 x_i x_i')(X'X)^{-1}$$

Оценка тогда имеет вид:

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}) = (X'X)^{-1}(\sum e_i^2 x_i x_i')(X'X)^{-1}$$

Это и есть оценки в форме Уайта.

2.18 Вопрос 18. Проблемы МНК

Пусть есть некоторая модель:

$$y_i = \beta_1 + \beta_2 \cdot x_i^{(2)} + ... + \beta_k \cdot x_i^{(k)} + \varepsilon_i$$

Если мы отказываемся от (4) и (5) предпосылок КЛММР, мы получаем довольно досадную ситуацию для МНК-оценок $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'y$:

- Они остаются несмещенными;
- Наша оценка ковариационной матрицы оценок оказывается смещенной, а сама ковариационная матрица имеет следующий вид:

$$\begin{gathered}\mathrm{Var}(\hat{\boldsymbol{\beta}}^{OLS}) = \mathrm{Var}((X'X)^{-1}X'y) = \mathrm{Var}((X'X)^{-1}X'(X\boldsymbol{\beta} + \boldsymbol{\varepsilon})) = \\ = \mathrm{Var}((X'X)^{-1}X'\boldsymbol{\varepsilon}) = ((X'X)^{-1}X')\,\mathrm{Var}(\boldsymbol{\varepsilon})((X'X)^{-1}X')' = \\ = ((X'X)^{-1}X')\Omega((X'X)^{-1}X')'\end{gathered}$$

  Это весьма далеко от $(X'X)^{-1}\sigma^2$.
- МНК-оценки решительно неэффективны.

Таким образом, из теоремы Айткена мы получаем вид для ОМНК-оценки:

$$\hat{\boldsymbol{\beta}}^{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$

Утверждение 2.18.1. $E((X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y) = \boldsymbol{\beta}$

Доказательство. Сделаем простейшие преобразования и сократим все лишнее:

$$\hat{\boldsymbol{\beta}}^{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}(X\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \\ = \boldsymbol{\beta} + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\boldsymbol{\varepsilon}$$

Берем матожидание:

$$E(\boldsymbol{\beta} + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\boldsymbol{\varepsilon}) = \boldsymbol{\beta} + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}E(\boldsymbol{\varepsilon}) = \boldsymbol{\beta}$$

□

2.19 Вопрос 19. Предпосылки со стохастическими регрессорами

Теперь откажемся от предпосылки детерминированности регрессоров.
Мы получаем предпосылки КЛММР со стохастическими регрессорами:

1. Модель линейна по параметрам и правильно специфицирована: $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$;
2. Значения регрессоров случайным образом извлечены из их генеральной совокупности;
3. $E\boldsymbol{\varepsilon} = 0$;
4. Регрессоры не связаны линейной связью;
5. Ковариационная матрица случайных ошибок имеет следующий вид:
$$\mathrm{Var}(\boldsymbol{\varepsilon}) = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix}$$
6. Случайные ошибки независимы: $\forall i,j, i \neq j \ \mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$;
7. $E(\varepsilon_i|\boldsymbol{x}) = 0$.

2.20 Вопрос 20. Теорема Гаусса-Маркова

Теорема I.6 (Теорема Гаусса-Маркова)**.** Если

1. Модель линейна по параметрам и правильно специфицирована: $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$;

2. Значения регрессоров случайным образом извлечены из их генеральной совокупности;

3. $E\boldsymbol{\varepsilon} = 0$;

4. Регрессоры не связаны линейной связью;

5. Ковариационная матрица случайных ошибок имеет следующий вид:

$$\mathrm{Var}(\boldsymbol{\varepsilon}) = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix}$$

6. Случайные ошибки независимы: $\forall i,j, i \neq j \; E(\varepsilon_i, \varepsilon_j) = 0$;

7. $E(\varepsilon_i|\boldsymbol{x}) = 0$,

то МНК-оценка $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'y$ несмещенная и состоятельная.

2.21 Вопрос 21. Предел по вероятности

Определение 2.21.1. Число X есть предел по вероятности последовательности случайных величин $\{x_n\}$ или же $\operatorname{plim} x_n = X$, если $\forall\varepsilon\,\forall\delta\,\exists n_0 : \forall N > n_0 \; P_n\{|x_n - X| \geq \varepsilon\} < \delta$

Интуитивный смысл: для произвольного ε мы задаем шар вокруг числа a. Мы хотим, чтобы для любой вероятности δ события, что член последовательности вылетит за его пределы, должен найтись номер члена, начиная с которой каждый последующий вылетает из шара с меньшей вероятностью.

2.22 Вопрос 22. Закон больших чисел

Теорема I.7 (Закон больших чисел I)**.** Пусть $x_1...x_n$ – i.i.d. случайные величины с матожиданием a и конечной дисперсией.

Тогда $\operatorname{plim} \overline{x} = a$

Дадим чуть более общую формулировку:

Теорема I.8 (Закон больших чисел II)**.** Пусть независимые случайные величина $x_1...x_n$ имеют математические ожидания $E(x_i)$ и дисперсии $\operatorname{Var}(x_i)$. Пусть при этом $\lim\limits_{n\to\infty} \dfrac{\sum \operatorname{Var}(x_i)}{n^2} = 0$

Тогда справдливо следующее равенство:

$$\operatorname{plim} \left(\frac{1}{n} \sum x_i - \frac{1}{n} \sum E(x_i) \right) = 0$$

2.23 Вопрос 23. Сходимость по распределению. ЦПТ

Пусть есть последовательность случайных величин x_n, каждая с функцией распределения $F_n(x_n)$. Можно сформулировать несколько иное понятие:

Определение 2.23.1. Говорят, что последовательность x_n **сходится по распределению**[9] к случайной величине x, если $\lim\limits_{n\to\infty} F_n(x) = F(x)$ во всех точках непрерывности $F(x)$.

При большом количестве случайных величин наблюдается одно чудесное свойство, выражаемое Центральной Предельной Теоремой.

Теорема I.9 (ЦПТ (Линдеберг-Леви))**.** Пусть $x_1...x_n$ – i.i.d. случайные величины с матожиданием a и конечной ненулевой дисперсией σ^2.

Тогда при $\dfrac{\sqrt{n}(\frac{1}{n}\sum x_i - a)}{\sigma} \xrightarrow{d} N(0{,}1)$.

ЦПТ имеет большое количество обобщений. Приведем одно:

Теорема I.10 (ЦПТ (Линдеберг))**.** Пусть $x_1...x_n$ – случайные величины с плотностями рапределения $p(x)$, с матожиданиями $E(x_i)$ и конечными дисперсиями $\operatorname{Var}(x_i)$

[9] Еще говорят о *слабой сходимости.*

Пусть выполняется **условие Линдеберга**: при любом $\varepsilon > 0$

$$\lim_{n\to\infty} \frac{\sum \int\limits_{|x-E(x_i)|>\varepsilon\sqrt{\sum \mathrm{Var}(x_i)}} (x - E(x_i)^2 p(x) dx}{\sum \mathrm{Var}(x_i)} = 0$$

Тогда $\dfrac{\sum x_i - \sum E(x_i)}{\sqrt{\sum \mathrm{Var}(x_i)}} \xrightarrow{d} N(0{,}1)$.

Отметим еще несколько свойств[10] здесь:

Утверждение 2.23.1. Сходимость по вероятности означает сходимость по распределению.

Утверждение 2.23.2. Сходимость по распределению означает сходимость по вероятности, если предельная случайная величина есть константа.

2.24 Вопрос 24. Состоятельность

Определение 2.24.1. Оценка параметра θ $\hat{\theta}$ называется **состоятельной**, если $\lim\limits_{n\to\infty} P\left\{|\hat{\theta} - \theta| \geq \varepsilon\right\} = 0$

Языком Вейерштрасса: $\forall \varepsilon > 0 \, \forall \delta > 0 \, \exists n_0 : \, \forall N > n_0 \, P_n \left\{|\hat{\theta} - \theta| \geq \varepsilon\right\} < \delta$

2.25 Вопрос 25. Асимптотический подход

Суть асимптотического подхода заключается в изучении свойств объектов в условиях неограниченно большой выборки.

В то время как точный подход занят изучением этих свойств именно на данных конечных выборках, асимптотический подход дает возможность посмотреть, что будет, если выборка бесконечно велика.

Он имеет ряд преимуществ:

- ЦПТ дает асимптотическую нормальность для сумм случайных величин при очень широких условиях. Хорошая изученность нормального распределения сильно упрощает некоторые выкладки.

[10] Еще крайне важно отметить, что суммы и произведения сходящихся по распределению величин, вообще говоря, не сходятся к суммам или произведениям предельных величин.

- В эконометрике использование асимптотического подхода позволяет отказаться от некоторых рестриктивных предпосылок, что добавляет реализма в анализ.
- Большой объем выборки положительно влияет на качество данных в ней. Это частично смягчает некоторые возможные проблемы в них.

2.26 Вопрос 26. Пример смещенности и состоятельности

Требования несмещенности и состоятельности часто могут нарушаться. Приведем пример смещенной, но состоятельной оценки.

Пусть есть $x_1 ... x_n$ из нормальной генеральной совокупности $N(a,\sigma^2)$. Оценим теоретическую дисперсию выборочной:

$$\hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \overline{x})^2$$

Очевидно из курса матстатистики, что:

$$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$$

При этом[11] $\operatorname{plim} \hat{\sigma}^2 = \sigma^2$.

Таким образом, мы имеем пример только асимптотически несмещенной, но состоятельной оценки.

2.27 Вопрос 27. Пример несмещенности и несостоятельности

Альтернативно можно предъявить и пример несмещенной и несостоятельной оценки.

Пусть есть $x_1 ... x_n$ из нормальной генеральной совокупности $N(a,\sigma^2)$
Тогда оценка матожидания по x_1 будет:

- Несмещенной: $E(x_1) = a$;

[11] Этот факт следует из того, что $\operatorname{Var}(\hat{\sigma}^2) = \dfrac{2\sigma^4(n-1)}{n^2}$. Неравенство Чебышева очевидно показывает, что выборочная оценка сходится по вероятности к теоретическому значению.

- Несостоятельной: при любом значении n распределение x_1 будет оставаться на месте, оно всегда будет $N(a,\sigma^2)$.

2.28 Вопрос 28. Теорема Слуцкого

Пусть $\{x_n\}, \{y_n\}$ – последовательности скалярных, векторных или матричных случайных величин.

Теорема I.11 (Слуцкий, 1925)**.** Если $x_n \xrightarrow{d} x$, а $y_n \xrightarrow{p} c$, то

- $x_n + y_n \xrightarrow{d} x + c$;
- $x_n y_n \xrightarrow{d} cx$;
- $\frac{x_n}{y_n} \xrightarrow{d} \frac{x}{c}$, если существует c^{-1}.

2.29 Вопрос 29. Теорема Манна-Вальда

Пусть $\{x_n\}$ – последовательность случайных величин на метрическом пространстве (X,ρ). Пусть $g(x) : (X, \rho) \to (X',\rho')$

Теорема I.12 (Манн, Вальд, 1943)**.** Если функция $g(x)$ имеет множество точек разрыва меры нуль[12], то

- $x_n \xrightarrow{d} x \Rightarrow g(x_n) \xrightarrow{d} g(x)$;
- $x_n \xrightarrow{p} x \Rightarrow g(x_n) \xrightarrow{p} g(x)$;

2.30 Вопрос 30. Состояльность при стохастических регрессорах

Рассматривается классическая линейная модель парной регрессии со стохастическим регрессором:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$$

[12] В самом простом случае она просто непрерывна.

Как мы помним,

$$\hat{\beta}_2 = \beta_2 + \frac{\sum(x_i - \overline{x})(\varepsilon_i - \overline{\varepsilon})}{\sum(x_i - \overline{x})^2}$$

Рассмотрим вопрос о состоятельности нашей оценки и перечислим, какие условия для этого требуются.

Как известно, выборочные оценки ковариации и дисперсии сходятся к своим теоретическим значениям – они состоятельны.

Мы имеем, что:

$$\hat{\beta}_2 = \beta_2 + \frac{\widehat{\mathrm{Cov}}(x,\varepsilon)}{\widehat{\mathrm{Var}}(x)}$$

На бесконечности:

$$\hat{\beta}_2 = \beta_2 + \frac{\widehat{\mathrm{Cov}}(x,\varepsilon)}{\widehat{\mathrm{Var}}(x)} \xrightarrow{p} \beta_2 + \frac{\mathrm{Cov}(x,\varepsilon)}{\mathrm{Var}(x)}$$

Отсюда очевидно, что мы должны иметь хоть немного отличающиеся значение регрессора, чтобы иметь ненулевой знаменатель, и его некоррелированость со случайной ошибкой, чтобы иметь нулевой числитель.

Значит, нам непосредственно потребовались предпосылки (1), (2) и **(7)**.

1. Модель линейна по параметрам и правильно специфицирована: $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$;

2. Значения регрессоров случайным образом извлечены из их генеральной совокупности.

7. $E(\varepsilon_i|\boldsymbol{x}) = 0$.

2.31 Вопрос 31. 2МНК

Рассматривается модель парной регрессии со стохастическим регрессором:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$$

x – эндогенная переменная, поэтому мы находим инструмент z, чтобы получить состоятельные оценки коэффициентов. Для этого мы используем *двухшаговый МНК*.

На первом шаге мы оцениваем регрессию каждого эндогенного регрессора на все инструменты[13] и все экзогенные переменные.

Полученные прогнозные значения уже не коррелируют со случайными ошибками, поэтому мы подставляем их в исходное уравнение регрессии и оцениваем его.

$$\begin{cases} x_i = a_1 + a_2 z_i + v_i \\ y_i = \beta_1 + \beta_2 \hat{x}_i + \varepsilon_i \end{cases}$$

$\boxed{1}$ Найдем оценки:

$$\hat{a}_2 = \frac{\widehat{\mathrm{Cov}}(x,z)}{\widehat{\mathrm{Var}}(z)}$$

$$\hat{\beta}_2^{TSLS} = \frac{\widehat{\mathrm{Cov}}(y,\hat{x})}{\widehat{\mathrm{Var}}(\hat{x})} = \frac{\widehat{\mathrm{Cov}}(y,\hat{a}_0 + \hat{a}_1 z)}{\widehat{\mathrm{Var}}(\hat{a}_0 + \hat{a}_1 z_i)} = \frac{\hat{a}_1 \widehat{\mathrm{Cov}}(y,z)}{\hat{a}_1^2 \widehat{\mathrm{Var}}(z)} = \frac{\widehat{\mathrm{Cov}}(y,z)}{\hat{a}_1 \widehat{\mathrm{Var}}(z)} =$$

$$= \frac{\widehat{\mathrm{Cov}}(y,z)}{\frac{\widehat{\mathrm{Cov}}(x,z)}{\widehat{\mathrm{Var}}(z)} \widehat{\mathrm{Var}}(z)} = \frac{\widehat{\mathrm{Cov}}(y,z)}{\widehat{\mathrm{Cov}}(x,z)} = \frac{\frac{1}{n}\sum(y_i - \overline{y})(z_i - \overline{z})}{\frac{1}{n}\sum(x_i - \overline{x})(z_i - \overline{z})}$$

$\boxed{2}$ Еще раз зададимся вопросом о состоятельности:

$$\hat{\beta}_2^{TSLS} = \frac{\widehat{\mathrm{Cov}}(y,z)}{\widehat{\mathrm{Cov}}(x,z)} = \frac{\widehat{\mathrm{Cov}}(z,\beta_1 + \beta_2 x_i + \varepsilon_i)}{\widehat{\mathrm{Cov}}(x,z)} = \frac{\beta_2 \widehat{\mathrm{Cov}}(z,x) + \widehat{\mathrm{Cov}}(z,\varepsilon)}{\widehat{\mathrm{Cov}}(x,z)}$$

Значит, $\hat{\beta}_2^{TSLS} = \beta_2 + \frac{\widehat{\mathrm{Cov}}(z,\varepsilon)}{\widehat{\mathrm{Cov}}(x,z)}$. Отсюда мы видим условия состоятельности, которые и накладываются на инструменты в предпосылках.

1. Экзогенность: $\mathrm{Cov}(z,\varepsilon) = 0$.

2. Релевантность: $\mathrm{Cov}(x,z) \neq 0$.

[13] Их не может быть меньше, чем эндогенных переменных.

Пусть выполняются эти условия. Тогда:

$$\hat\beta_2 = \beta_2 + \frac{\widehat{\mathrm{Cov}}(z,\varepsilon)}{\widehat{\mathrm{Cov}}(x,z)} \xrightarrow{p} \hat\beta_2 = \beta_2 + \frac{\mathrm{Cov}(z,\varepsilon)}{\mathrm{Cov}(x,z)} = \beta_2$$

Мы получаем состоятельную[14] оценку.

Утверждение 2.31.1. $\mathrm{Var}(\hat\beta_2^{TSLS}) = \dfrac{\sigma^2}{\sum(x_i - \overline{x})^2} \cdot \dfrac{1}{\widehat{\mathrm{Corr}}(x,z)^2}$[15].

Доказательство. Согласно утверждению 1.16.1:

$$\mathrm{Var}\,\hat\beta_2 = \frac{\sigma^2}{\sum(\hat x_i - \overline{\hat x})^2}$$

Нам надо преобразовать это выражение.

$$\frac{\sigma^2}{\sum(\hat\alpha_1 + \hat\alpha_2 z_i - \overline{\hat\alpha_1 + \hat\alpha_2 z})^2} = \sigma^2 \frac{\hat\alpha_2^{-2}}{\sum(z_i - \overline{z})^2} =$$

Как мы понимаем, $\hat\alpha_2 = \dfrac{\widehat{\mathrm{Cov}}(x,z)}{\widehat{\mathrm{Var}}(z)}$. Подставляем это:

$$= \sigma^2 \frac{\widehat{\mathrm{Var}}^2(z)}{\widehat{\mathrm{Cov}}^2(x,z)\sum(z_i - \overline{z})^2} = \sigma^2 \frac{\left(\frac{1}{n}\sum(z_i - \overline{z})^2\right)^2}{\left(\frac{1}{n}\sum(x_i - \overline{x})(z_i - \overline{z})\right)^2 \sum(z_i - \overline{z})^2} =$$

$$= \sigma^2 \frac{\frac{1}{n^2}\sum(z_i - \overline{z})^2}{\left(\frac{1}{n}\sum(x_i - \overline{x})(z_i - \overline{z})\right)^2} = \sigma^2 \frac{1}{n} \frac{\widehat{\mathrm{Var}}(z)}{\widehat{\mathrm{Cov}}^2(x,z)} =$$

Наконец, материализуем дисперсию x:

$$= \sigma^2 \frac{1}{n} \frac{\widehat{\mathrm{Var}}(z)\,\widehat{\mathrm{Var}}(x)}{\widehat{\mathrm{Cov}}^2(x,z)\,\widehat{\mathrm{Var}}(x)} = \sigma^2 \frac{1}{n} \frac{1}{\widehat{\mathrm{Corr}}^2(x,z)\,\widehat{\mathrm{Var}}(x)} =$$

$$= \frac{1}{n} \frac{\sigma^2}{\frac{1}{n}\sum(x_i - \overline{x})^2} \frac{1}{\widehat{\mathrm{Corr}}^2(x,z)} = \frac{\sigma^2}{\sum(x_i - \overline{x})^2} \cdot \frac{1}{\widehat{\mathrm{Corr}}(x,z)^2}$$

[14] Все это время мы предполагаем, что у нас верная спецификация, иначе, конечно, у нас никакой состоятельности не случится.

[15] Только в гомоскедастичном случае.

□

Следствие 2. Оценка[16] стандартной ошибки будет иметь вид:

$$se(\hat{\beta}_2) = \sqrt{\frac{S^2}{\sum(x_i - \overline{x})^2} \cdot \frac{1}{\widehat{\mathrm{Corr}}(x,z)^2}}$$

$$S^2 = \frac{ESS}{n-k} = \frac{\sum(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)}{n-2}$$

2.32 Вопрос 32. Эндогенность и экзогенность

Определение 2.32.1. Регрессор **эндогенный**, если он коррелирован со случайными ошибками модели

Определение 2.32.2. Регрессор **экзогенный**, если он не коррелирован со случайными ошибками модели

Работать с эндогенными регрессорами плохо, потому что они сбивают OLS с толку, и он дает несостоятельные оценки.

2.33 Вопрос 33. 2МНК при множественной регрессии

Пусть у нас есть модель множественной регрессии:

$$y_i = \beta_0 + \beta_1 x_i^{(1)} + ... + \beta_k x_i^{(k)} + \beta_{k+1} w_i^{(1)} + ... + \beta_{k+r} w_i^{(r)} + \varepsilon_i$$

Здесь x^j – эндогенные регрессоры, w^k – экзогенные регрессоры или контрольные переменные. Пусть у нас есть $m \geq k$ валидных инструментов z. Тогда нам доступен 2МНК.

Опишем его процедуру:

$\boxed{\mathrm{I}}$ На первом шаге делаем регрессии каждого из эндогенных регрессоров на константу, все инструменты и экзогенные переменные.

$$\begin{cases} x_i^{(1)} = \gamma_{11} z_i^{(1)} + ... + \gamma_{1m} z_i^{(m)} + \gamma_{1,m+1} w_i^{(1)} + ... + \gamma_{1,k+r} w_i^{(r)} + u_i \\ \vdots \\ x_i^{(k)} = \gamma_{k1} z_i^{(1)} + ... + \gamma_{km} z_i^{(m)} + \gamma_{k,m+1} w_i^{(1)} + ... + \gamma_{k,k+r} w_i^{(r)} + u_i \end{cases}$$

[16]ESS мы вычисляем по настоящим иксам.

Из этой системы получаем предсказанную матрицу значений $\hat{\boldsymbol{X}}$.

$\boxed{\text{II}}$ На втором шаге обычным МНК оцениваем регрессию, заменив истинные значения иксов на предсказанные на первом шаге:

$$y_i = \beta_0 + \beta_1 \hat{x}_i^{(1)} + ... + \beta_k \hat{x}_i^{(k)} + \beta_{k+1} w_i^{(1)} + ... + \beta_{k+r} w_i^{(r)} + \varepsilon_i$$

На этом получаются хорошие оценки.

2.34 Вопрос 34. Требования к инструментам

Чтобы мы могли использовать инструмент z, он должен быть валидным:

- Экзогенным: $\text{Cov}(z,\varepsilon) = 0$;
- Релевантным: для эндогенного регрессора x $\text{Cov}(x,z) \neq 0$.

2.35 Вопрос 35. Тест на эндогенность

Вообще говоря, мы предпочитаем МНК 2МНК, если мы не сталкиваемся с проблемой эндогенности. Если у нас нет эндогенности регрессоров, то и МНК, и 2МНК состоятельны[17], но 2МНК неэффективен. Поэтому нам бы хотелось иметь тест на эндогенность.

Мы имеем тест Хаусмана. Его идея – проверка, сильно ли отличаются результаты МНК и 2МНК. Самый простой способ – дополнительные регрессии. Через них и комбинацию стандартных тестов можно проверить эндогенность.

Выпишем чуть более общий вариант:

$$H^0: \text{ Регрессоры в модели экзогенны}$$
$$H^1: \text{ Есть эндогенные регрессоры}$$

Расчетная статистика имеет вид:

$$\chi^2_{\text{расч}} = (\hat{\boldsymbol{\beta}}^{TSLS} - \hat{\boldsymbol{\beta}}_{OLS})'(\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}^{TSLS}) - \widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_{OLS})^{-1}(\hat{\boldsymbol{\beta}}^{TSLS} - \hat{\boldsymbol{\beta}}_{OLS})$$

Мы сравниваем это с критическим значением $\chi^2_{\text{кр}}(\alpha, k)$, где k – число регрессоров.

[17]При условии хороших инструментов, которое критично для теста Хаусмана: он требует состоятельности хоть одной из оценок.

2.36 Вопрос 36. Тест на экзогенность

Если у нас число эндогенных регрессоров совпадает с числом инструментальных переменных, мы не можем проверить второе условие валидности – экзогенность. Однако когда у нас больше инструментов, мы можем провести тест Саргана.

В простейшем случае его проведение также связано со вспомогательными регрессиям, но мы рассмотрим общий метод.

Пусть есть модель:

$$y_i = \beta_0 + \beta_1 x_i^{(1)} + ... + \beta_k x_i^{(k)} + \beta_{k+1} w_i^{(1)} + ... + \beta_{k+r} w_i^{(r)} + \varepsilon_i$$

$$H^0: \text{ Инструменты экзогенны}$$

$$H^1: \text{ Есть эндогенные инструменты}$$

1. Оцениваем нашу исходную модель с помощью 2МНК.
2. Получаем остатки e_i.
3. Оцениваем вспомогательную регрессию остатков на инструменты и экзогенные переменные:

$$e_i = a_0 + a_1 z_i^{(1)} + ... + a_m z_i^{(m)} + a_{m+1} w_i^{(1)} + ... + a_{m+r} w_i^{(r)} + v_i$$

4. Вычиляем значение $F_{\text{расч}}$ теста с нулевой гипотезой $a_1 = ... = a_m = 0$.
5. Конструируем J-статистику: $J_{\text{расч}} = mF_{\text{расч}}$.
6. Сравниваем с $\chi^2_{\text{кр}}(m - k, \alpha)$.

2.37 Вопрос 37. Тест на релевантность. Слабые инструменты

Следующая заметная проблема с использованием инструментальных переменных возникает в вопросе их релевантности.

Если наши инструменты объясняют очень малую долю колебаний эндогенных переменных, то они называются *слабыми*. Работы со слабыми инструментами ведет к горе проблем:

- Теряются желанные асимптотические свойства:
- 2МНК оценки смещены, причем смещение остается даже с очень большими наборами данных. Происходит смещение в сторону МНК-оценок.
- Бессмысленно добавлять все больше инструментов, чтобы улучшить асимптотические свойства.
- С большой долей вероятности состоятельность не происходит.
- Тесты выполняются неправильно, а доверительные интервалы имееют не те размеры или просто смещены.

К счастью, на вопрос о релевантности наших инструментов можно ответить. Мы используем тест **Крэгга-Дональда**. Его расчетная статистика – минимальное собственное значение весьма ужасного выражения[18]

Это расчетное значение сравнивается с критическим. Статья (Stock, Yogo, 2005) в таблице 1 приводит критические значения в зависимости от числа эндогенных регрессоров (n), числа инструментов (K_2), и желаемого максимального смещения 2МНК оценки относительно МНК-оценок (b) на уровне значимости 5%.

Как оказывается, для одного эндогенного регрессора это значение совпадает с $F_{\text{расч}}$[19] первого шага 2МНК. Для принятия гипотезы о сильных инструментах она должна превышать 10.

2.38 Вопрос 38. Ситуации, ведущие к эндогенности

К сожалению, в мире довольно много ям, ведущих к коррелированности регрессора со случайной ошибкой. Перечислим пять основных из них:

[18] Конечно, его нельзя не выписать.

Пусть для произвольной матрицы A $P_A = A(A'A)^{-1}A'$, а $M_A = I - P_A$.

Пусть $\bar{Z}$ – матрица всех экзогенных переменных. k_1 – число включенных в основную модель экзогенных переменных, с матрицей W; k_2 - число инструментов, не включенных в основную модель, с матрицей Z. Пусть у нас T эндогенных переменных. Пусть X – их матрица.

Тогда статистика Крэгга-Дональда имеет следующий вид:

$$G_T = (X'M_{\bar{Z}}X)^{-1/2}(M_WX)'M_WZ((M_WZ)'M_WZ)^{-1}(M_WZ)'(M_WX)(X'M_{\bar{Z}}X)^{-1/2}\frac{(T-K_1-K_1)^2}{K_2}$$

[19] Значимость всего уравнения или короткая/длинная регрессия с отрубанием инструментов.

1. Пропуск существенных переменных;

2. Некорректная спецификация функциональной формы;

3. Ошибки измерения в регрессорах;

4. Sample selection bias: изначальное смещение в выборке. В отличие от проблем, наблюдаемых при просто плохом измерении регрессора, здесь у нас не хватает данных потому, что процесс отбора данных как-то сязан с зависисимой переменной, а не только регрессорами.

5. Одновременная причинно-следственная связь.

2.39 Вопрос 39. Ситуации

Рассмотрим разные случаи и уместность применения 2МНК. Как мы помним, если нет необходимости в нем, оценки будут состоятельными, но неэффективными.

(a) Используя квартальные данные за 20 лет, уважаемый исследователь анализирует зависимость индекса промышленного производства от линейного тренда (от переменной времени t) и от трех фиктивных переменных сезонности (соответствующих первому, второму и третьему кварталу каждого года).

Не требуется 2МНК.

(b) Уважаемый исследователь анализирует влияние численности полицейских в регионе на количество преступлений, совершаемых в этом регионе.

Очевидно требуется 2МНК, имеется двусторонняя причинно-следственная связь.

(c) Уважаемый исследователь анализирует зависимость реального ВВП от уровня неравенства распределения доходов в экономике (уровень неравенства измеряется с помощью коэффициента Джини).

Вопрос, в определенном смысле экономически коварный. Тем не менее, подозревается одновременная причинно-следственная связь. Нужен 2МНК.

(d) Моделируется цена квартиры на вторичном рынке жилья в зависимости от площади этой квартиры.

Здесь вполне очевидна некоррелированность, обычный МНК сойдет.

2.40 Вопрос 40. Ошибка измерения x

Пусть переменные y и x^* связаны точным линейным соотношением:

$$y_i = \beta_1 + \beta_2 x_i^*$$

Однако мы имеем лишь измеренные с ошибкой значения нашего регрессора: $x_i = x_i^* + u_i$, где u_i – i.i.d., $E(u_i) = 0$, $\mathrm{Var}(u_i) = \sigma^2$.

Пусть мы МНК-оцениваем регрессию:

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

Рассмотрим вопрос о свойствах ошибок и состоятельности оценок. Что есть эти ε_i?

Подставим в исходную модель:

$$y_i = \beta_1 + \beta_2 x_i^*$$

$$y_i = \beta_1 + \beta_2 (x_i - u_i)$$

$$y_i = \beta_1 + \beta_2 x_i - \beta_2 u_i$$

Значит, $\varepsilon_i = -\beta_2 u_i$.

Теперь очевидна коррелированность регрессоров в наблюдаемой модели со случайными ошибками:

$$\mathrm{Cov}(x,\varepsilon) = \mathrm{Cov}(x^* + u_i, -\beta_2 u_i) = -\beta_2 \mathrm{Cov}(x^* + u_i, u_i) =$$

$$= -\beta_2 \mathrm{Cov}(x^*, u_i) - \beta_2 \mathrm{Var}(u_i,u_i) = -\beta_2 \sigma^2$$

На этом состоятельность закончилась:

$$\hat{\beta}_2 \xrightarrow{p} \beta_2 - \frac{\beta_2 \sigma^2}{\mathrm{Var}(x)}$$

Мы видим, что наша оценка смещена к нулю.

2.41 Вопрос 41. Ошибка измерения y

Рассмотрим теперь альтернативную формулировку, не ведущую к таким проблемам.

Пусть переменные y^* и x связаны точным соотношением:

$$y^* = \beta_1 + \beta_2 x_i$$

Однако вместо точных значений зависимой переменной мы наблюдаем измеренные с ошибкой значения: $y_i = y_i^* + u_i$, где u_i – i.i.d. случайные величины с нулевым математическим ожиданием и дисперсией σ^2.

Мы оцениваем методом наименьших квадратов уравнение:

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

Рассмотрим, что происходит теперь с нашими оценками.

Рассмотри природу ε:

$$\begin{cases} y^* = \beta_1 + \beta_2 x_i \\ y_i^* = y_i - u_i \end{cases}$$

$$y_i = \beta_1 + \beta_2 x_i + u_i$$

Но случайные ошибки измерения зависимой переменной не связаны с регрессором. Таким образом, у нас выполняются все предпосылки ЛММР со стохастическими регрессорами.

Тогда легко видим несмещенность и состоятельность нашей МНК-оценки:

$$\operatorname{plim} \hat{\beta}_2 = \beta + \frac{\operatorname{Cov}(u, x)}{\operatorname{Var}(x)} = \beta_2$$

2.42 Вопрос 42. Оценка MPC I

Рассмотрим примитивную модельку:

$$C_t = \beta_1 + \beta_2 GDP_t + \varepsilon_t$$

$$GDP_t = C_t + I_t$$

Здесь мы говорим о ВВП, потреблении и инвестициях, причем инвестиции некоррелированы со случайными ошибками.

$\boxed{\text{а}}$ Проблема в том, что мы имеем одновременные уравнения, поэтому МНК не даст состоятельную оценку. Покажем это.

Как известно, для МНК-оценки верно:

$$\hat{\beta}_2 \xrightarrow{p} \beta_2 + \frac{\operatorname{Cov}(GDP,\varepsilon)}{\operatorname{Var}(GDP)}$$

Будем искать эту ковариацию. Подставим выражение для потребления в ВВП и приведем подобные члены:

$$GDP = \frac{\beta_1 + I + \varepsilon}{1 - \beta_2}$$

Возьмем ковариацию от этого выражения:

$$\operatorname{Cov}(GDP, \varepsilon) = \operatorname{Cov}\left(\frac{\beta_1 + I + \varepsilon}{1 - \beta_2}, \varepsilon\right) = \frac{\operatorname{Cov}(I, \varepsilon)}{1 - \beta_2} + \frac{\operatorname{Var}(\varepsilon)}{1 - \beta_2}$$

И вот, случается магия:

$$\hat{\beta}_2 \xrightarrow{p} \beta_2 + \frac{\operatorname{Var}(\varepsilon)}{(1 - \beta_2)\operatorname{Var}(GDP)}$$

Мы получили завышенную оценку, потому что имеем эндогенную переменную, нарушая наши предпосылки.

$\boxed{\text{б}}$ Для получения состоятельно оценки нам придется применить 2МНК. Единственный возможный инструмент – инвестиции. Из условия вопроса и начилия инвестиций в уравнении ВВП следует их валидность как инструмента.

$\boxed{\text{I}}$ Делаем регрессию ВВП на константу и инвестии.

$\boxed{\text{II}}$ Делаем регрессию потребления на предсказанные значения ВВП.
Это дает нам состоятельную оценку.

2.43 Вопрос 43. Оценка MPC II

Рассмотрим простую модель закрытой экономики со следующими предпосылками: инвестиции и госзакупки экзогенны:

$$C_t = \beta_1 + \beta_2 GDP_t + \varepsilon_t$$

$$GDP_t = C_t + I_t + G_t$$

$\boxed{\text{а}}$ Покажем, что МНК-оценка предельной склонности к потреблению несостоятельна.

Выразим как всегда из системы одновременных уравнений эндогенные переменные.

$$GDP = \beta_1 + \beta_2 GDP + \varepsilon + I + G$$

$$GDP = \frac{\beta_1 + I + G + \varepsilon}{1 - \beta_2}$$

Найдем теперь ковариацию ВВП и случайной ошибки:

$$\operatorname{Cov}(GDP, \varepsilon) = \frac{\sigma^2}{1 - \beta_2}$$

Это выражение больше нуля. Получается, что ВВП эндогенен.

Посчитаем здесь предел по вероятности: мы помним снова, что:

$$\hat{\beta}_2 \xrightarrow{p} \beta_2 + \frac{\operatorname{Var}(\varepsilon)}{(1 - \beta_2)\operatorname{Var}(GDP)}$$

Та же яма завышенности MPC.

$\boxed{\text{б}}$ Рассмотрим модель с инструментом $z = I + G$. Выпишем 2МНК оценку:

$$\hat{\beta}_2^{TSLS} = \frac{\widehat{\operatorname{Cov}}(C, I + G)}{\widehat{\operatorname{Cov}}(GDP, I + G)} \xrightarrow{p} \frac{\operatorname{Cov}(C, I + G)}{GDP, I + G} = \frac{\operatorname{Cov}(\beta_1 + \beta_2 GDP + \varepsilon, I + G)}{\operatorname{Cov}(GDP, I + G)} =$$

$$= \frac{\beta_2 \operatorname{Cov}(GDP, I + G) + \operatorname{Cov}(\varepsilon, I + G)}{\operatorname{Cov}(GDP, I + G)} = \beta_2$$

Валидность[20] нашего инструмента следует из условия. Мы получили состоятельную оценку.

[20]Чтобы доказать релевантность, просто посчитаем знаменатель:

$$(\operatorname{Cov}(GDP, I + G)) = \operatorname{Cov}\left(\frac{\beta_1 + \varepsilon + I + G}{1 - \beta_2}, I + G\right) = \frac{\operatorname{Var}(I + G)}{1 - \beta_2}$$

Из экономического смысла мы получаем, что это больше нуля.

$\boxed{\text{в}}$ Второй уважаемый исследователь сделает все то же, и получит ту же состоятельную оценку.

$\boxed{\text{г}}$ Теперь третий уважаемый исследователь собирается использовать два инструмента – I и G. С помощью 2МНК он проводит то же исследование.

Так как оба инструмента валидны, он получит состоятельную оценку. Таким образом, на бесконечность оценки третьего и двух других совпадут. В конечнх выборках допустимы различия, точность будет выше у третьего исследователя.

2.44 Вопрос 44. О телевидении

Уважаемый исследователь анализирует влияние телевидения на результаты выборов. Он располагает данными о популярности телеканала X-TB в каждом регионе (переменная x_i — количество людей, которые смотрят этот телеканал в i-м регионе), а также о количестве голосов полученных в этом регионе партией «Народное процветание», которую канал X-TB активно поддерживал во время предвыборной кампании 1999 года (переменная y_i).

Уважаемый исследователь предполагает наличие двусторонней причинно-следственной связи: с одной стороны, телеканал сильнее влияет на результаты выборов там, где он более популярен. С другой стороны, телеканал более популярен именно там, где много сторонников «Народного процветания», так как они знают, что телеканал часто хвалит их любимую партию, и охотно смотрят его. Помимо прочего, на переменную X влияет переменная z_i, которая равна единице, если канал X-TB транслировался в i-м регионе в 1980 году (задолго до образования партии «Народное процветание») и равна нулю в противном случае.

$\boxed{\text{а}}$ Описанная ситуация, может быть представлена как система одновременных уравнений:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$$

$$x_i = a_1 + a_2 \cdot y_i + a_3 \cdot z_i + u_i$$

МНК-оценка первого уравнения будет несостоятельной, так как регрессор коррелирован со случайной ошибкой. Чтобы в этом убедиться, выразим x через экзогенные z, ε, u.

$$x_i = \frac{a_1 + a_2(\beta_1 + \varepsilon_i) + a_3 \cdot z_i + u_i}{1 - a_2\beta_2}$$

Теперь мы точно можно доказать эндогенность, вычислив ковариацию между x и ε:

$$\mathrm{Cov}(\varepsilon{,}x) = \frac{a_2 \cdot \sigma^2}{1 - a_2\beta_2} \neq 0$$

Для этого, конечно, делаем шаткое предположение, что $a_2\beta_2 \neq 1$. Из этого немедленно очевидна несостоятельность:

$$\operatorname{plim} \hat{\beta}_2 = \beta_2 + \frac{Cov(\varepsilon, x)}{\mathrm{Var}(x)}$$

Как же нам быть? Придется оценивать первое из наших уравнений с помощью 2МНК с инструментов z. Раз у нас один эндогенный регрессор, нам хватает одного инструмента

Мы показали уже релевантность z. Здравый смысл и условие подсказывают еще и экзогенность для z.

$\boxed{\text{b}}$ Как мы помним,

$$\hat{\beta}_2^{TSLS} = \frac{\widehat{\mathrm{Cov}}(y{,}z)}{\widehat{\mathrm{Cov}}(x{,}z)}$$

Таким образом, мы получаем, что

$$\hat{\beta}_2^{TSLS} \xrightarrow{p} \frac{\mathrm{Cov}(y{,}z)}{\mathrm{Cov}(x{,}z)} = \frac{\mathrm{Cov}(y{,}z)}{\mathrm{Cov}(x{,}z)} = \frac{\mathrm{Cov}(\beta_1 + \beta_2 x + \varepsilon), z}{\mathrm{Cov}(x{,}z)} = \beta_2$$

Магия состоятельности.

2.45 Вопрос 45. Вывод оценки 2МНК

Пусть у нас есть модель со стохастическими регрессорами:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Здесь регрессоры коррелированы со случайными ошибками модели. Известно, что количество инструментов превышает количество эндогенных регрессоров. Пусть Z – матрица инструментов.

Утверждение 2.45.1. $\hat{\boldsymbol{\beta}}^{TSLS} = ((Z(Z'Z)^{-1}Z'X)'Z(Z'Z)^{-1}Z'X)^{-1}(Z(Z'Z)^{-1}Z'X)'y$

Доказательство. Проведем обычную процедуру 2МНК, чтобы показать наш результат.

$\boxed{\text{I}}$ Оценим сперва матрицу $\hat{X}$, чтобы использовать ее дальше.

Пусть:

$$X = Z\boldsymbol{a} + u_i$$

Тогда:

$$\hat{X} = Z\hat{\boldsymbol{a}}$$

Выражаем сначала оценку коэффициентов, а потом нашу матрицу:

$$\hat{\boldsymbol{a}} = (Z'Z)^{-1}Z'X$$

$$\hat{X} = Z(Z'Z)^{-1}Z'X$$

$\boxed{\text{II}}$ Теперь вычисляем МНК-оценки для исходной модели:

$$y_i = \hat{X}\beta + \boldsymbol{\varepsilon}$$

$$\hat{\boldsymbol{\beta}}^{TSLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'y$$

Наконец, подставляем наше выражение:

$$\hat{\boldsymbol{\beta}}^{TSLS} = ((Z(Z'Z)^{-1}Z'X)'Z(Z'Z)^{-1}Z'X)^{-1}(Z(Z'Z)^{-1}Z'X)'y$$

□

2.46 Вопрос 46. Разложение 2МНК-оценки

Рассмотрим здесь случай, когда число инструментов и эндогенных переменных совпадает.

Утверждение 2.46.1. $\hat{\boldsymbol{\beta}}^{TSLS} = (Z'X)^{-1}Z'y$

Доказательство. Воспользуемся парой свойств для доказательства этого результата. В наших предпосылках они верны:

- $(X'X)' = X(X')' = X'X$
- $(AB)^{-1} = B^{-1}A^{-1}$.

Как мы помним,

$$\hat{\boldsymbol{\beta}}^{TSLS} = ((Z(Z'Z)^{-1}Z'X)'Z(Z'Z)^{-1}Z'X)^{-1}(Z(Z'Z)^{-1}Z'X)'y =$$

$$= (X'Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y =$$

$$= (Z'X)^{-1}(Z'Z)(X'Z)^{-1}(X'Z)(Z'Z)^{-1}Z'y =$$

$$= (Z'X)^{-1}Z'y$$

□

2.47 Вопрос 47. Пример с ненаблюдаемой переменной

Пусть истинная модель имеет вид:

$$y_i = \beta_1 + \beta_2 x_i^{(2)} + \beta_3 x^{(3)} + \varepsilon_i$$

Мы ничего не знаем о третьем икс, но зато знаем второй, y, z. При этом $\mathrm{Cov}(z,x_i^{(2)}) \neq 0$, $\mathrm{Cov}(z,x_i^{(3)}) = 0$.

Утверждение 2.47.1. В таких условиях 2МНК может дать состояльную оценку β_2.

Доказательство. Оценим сокращенную версию модели:

$$y = \beta_1 + \beta_2 x_i^{(2)} + u_i$$
$$u_i = \beta_3 x_i^{(3)} + \varepsilon_i$$

Оценим эту модель с помощью 2МНК. Как мы понимаем, мы получим оценку коэффициента:

$$\hat{\beta}_2 = \frac{\widehat{\mathrm{Cov}}(y,z)}{\widehat{\mathrm{Cov}}(x_i^{(2)},z)} \xrightarrow{p} \frac{\mathrm{Cov}(y,z)}{\mathrm{Cov}(x_i^{(2)},z)} = \frac{\mathrm{Cov}(\beta_1 + \beta_2 x_i^{(2)} + \beta_3 x_i^{(3)} + \varepsilon_i, z)}{\mathrm{Cov}(x_i^{(2)},z)} =$$

$$= \frac{\beta_2 \mathrm{Cov}(x_i^{(2)},z) + \beta_3 \mathrm{Cov}(x_i^{(3)},z) + \mathrm{Cov}(\varepsilon, z)}{\mathrm{Cov}(x_i^{(2)},z)} = \beta_2$$

Все эти чудесные результаты мы получаем с опорой на условие.

□

Глава 3
Контрольная работа 2

3.1 Вопрос 1. Логит-модель

Уважаемый исследователь оценивает вероятность наступления некоторого события в зависимости от того, какое значение принимает регрессор x, используя логит-модель бинарного выбора.

Логит-модель Λ – обобщенная линейная модель вида:

$$p = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}} \in [0,1]$$

Это элементарно обобщается на много регрессоров:

$$p = \frac{1}{1+e^{-\boldsymbol{x}'\boldsymbol{\beta}}}$$

Оценка модели Довольно ясно, что нужно просто воспользоваться методом максимального правдоподобия, чтобы оценить эту модель. Запишум функцию правдоподобия:

$$L = L(y_1 \ldots y_n) = \prod_{y_i=0}(1-\Lambda(\boldsymbol{x_i'\beta}))\prod_{y_i=1}\Lambda(\boldsymbol{x_i'\beta})$$

Дальше пользуемся предпосылкой независимости, разделяем все:

$$L = \prod_i (\Lambda(\boldsymbol{x_i'\beta}))^{y_i}(1-\Lambda(\boldsymbol{x_i'\beta}))^{1-y_i}$$

Логарифмируем это:

$$\ln L = \sum_i (y_i \ln\Lambda(\boldsymbol{x_i'\beta}) + (1-y_i)(1-\Lambda(\boldsymbol{x_i'\beta})))$$

Дальше дифференцируем это:

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \sum_i \left(\frac{y_i\Lambda'}{\Lambda} - \frac{(1-y_i)\Lambda'}{1-\Lambda}\right)x_i = 0$$

Используем, что для логит-модели справедливо, что $\Lambda' = \Lambda(1-\Lambda)$. Тогда все упрощается до:

$$\sum_i (y_i - \Lambda(\boldsymbol{x_i'}\boldsymbol{\beta}))x_i = 0$$

Отсюда уже находится решение. Нетривиальный факт состоит в том, что $\ln L$ для логит-модели вогнута, поэтому необходимое условие экстремума становится и достаточным.

Предельный эффект Самая простая интерпретация происходящего – вычисление предельного эффекта. Стоит заметить, что он разный для разных значений регрессоров.

Сам предельный эффект:

$$\frac{\partial \Lambda}{\partial x_i} = \beta_i \Lambda'(\boldsymbol{x'}\boldsymbol{\beta})$$

Естественная суть происходящего – мы смотрим наклон функции, идя по направлению изучаемого регрессора. Чтобы дать содержательную интерпретацию, совершается маленький трюк.

Мы говорим, что предельный эффект по x_i, вычисленный в точке $\boldsymbol{x_0}$, означает прирост вероятности для объекта с прочими характеристиками $\boldsymbol{x_0}$ при росте x_i на единицу. Это как бы откидывает, что мы вычисляем производную в точке.

В мире дополнительно часто считаются предельный эффект для среднего наблюдения и средний предельный эффект.

Предельный эффект для среднего наблюдения, как можно догадаться, считается в точке $\overline{\boldsymbol{x}}$. Описывает приросты для среднего во всем наблюдения.

Средний предельный эффект считается как среднее от предельных эффектов для всех наблюдений.

3.2 Вопрос 2. Пробит-модель

Уважаемый исследователь оценивает вероятность наступления некоторого события в зависимости от того, какое значение принимает регрессор x, используя пробит-модель бинарного выбора.

Пробит-модель Φ – обобщенная линейная модель вида:

$$p = \int_{\infty}^{x} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} dt \in [0, 1]$$

Это элементарно векторизуется и обобщается на много регрессоров:

$$p = \int_{\boldsymbol{\infty}}^{\boldsymbol{x}} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{\boldsymbol{t}^2}{2}} d\boldsymbol{t}$$

Оценка модели Довольно ясно, что нужно просто воспользоваться методом максимального правдоподобия, чтобы оценить эту модель. Запишум функцию правдоподобия:

$$L = L(y_1 \dots y_n) = \prod_{y_i=0} (1 - \Phi(\boldsymbol{x_i'\beta})) \prod_{y_i=1} \Phi(\boldsymbol{x_i'\beta})$$

Дальше пользуемся предпосылкой независимости, разделяем все:

$$L = \prod_i \left(\Phi(\boldsymbol{x_i'\beta})\right)^{y_i} \left(1 - \Phi(\boldsymbol{x_i'\beta})\right)^{1-y_i}$$

Логарифмируем это:

$$\ln L = \sum_i (y_i \ln \Phi(\boldsymbol{x_i'\beta}) + (1 - y_i)(1 - \Phi(\boldsymbol{x_i'\beta})))$$

Дальше дифференцируем это:

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \sum_i \left(\frac{y_i \Phi'}{\Phi} - \frac{(1 - y_i)\Phi'}{1 - \Phi} \right) x_i = 0$$

Отсюда уже находится решение. Важно отметить факт, что $\ln L$ для пробит-модели тоже вогнута, поэтому необходимое условие экстремума становится и достаточным.

Предельный эффект Предельный эффект:

$$\frac{\partial \Phi}{\partial x_i} = \beta_i \Phi'(\boldsymbol{x'\beta})$$

Все абсолютно совпадает.

3.3 Вопрос 3. Качество модели бинарного выбора

Справедливо, что мы хотим сравнивать свои модели бинарного выбора. Для этого можно использовать некоторое количество разных критериев.

Псевдо-R^2 МакФаддена Если наши модели подогнаны на одних данных, можно воспользоваться R^2_{McF}:

$$R^2_{McF} = 1 - \frac{\ln L}{\ln L_0}$$

L_0 – модель только с константой. Довольно ясно, что логарифм функции правдопободия по жизни отрицателен. У модели с константой и еще чем-то будет меньшее по модулю значения логарифмической функции правдопободия. Чем правдоподобнее, тем ближе к нулю.

Таким образом, чем выше R^2_{McF}, тем большей объясняющей силой обладает модель.

Критерий Акаике Информационный критерий Акаике (AIC) распространен очень широко и может использоваться для большого класса моделей.

$$AIC = -2 \ln L + 2d$$

$\ln L$ – максимизированное значение функции правдопободия для изучаемой модели, d – число параметров. Предпочитается модель с минимальным значением AIC.

Здесь нужно отметить, что AIC ничего не говорит об абсолютном качестве модели, только о сравнении разных моделей на одних данных.

Прочее В мире есть еще много способов дать сравнение. Рядом с AIC стоит BIC, C_p. Тем не менее, их проблема в том, что нам требуется оценивать дисперсию ошибок для них.

Дополнительно мы можем провести тест Андерсона-Дарлинга на принадлежность данных из выборки какому-то конкретному распределению.

Тестовая ошибка Другой способ – сравнить совершаемые на какой-то тестовой выборке ошибки для каждой модели. Это можно имплементировать как кросс-валидицией, так и простым делением выбор на training set и test set.

Каждую из наших моделей моделей мы подгоняем на training set, после чего делаем предсказания для наблюдений из тестовой выборки, высчитывая ошибку. В целом мы должны отдавать предпочтения самой верной модели, делая поправку на сложность. Например, если нам доступна заметно более простая модель, делающая лишь на малый процент больше ошибок, нам стоит выбрать ее.

Стоит отметить, что это самый очевидный способ сравнивать модели, подогнанные на разных данных.

Более примитивный подход – смотреть процент верных предсказаний на всей выборке. Проблема такого подхода – модель видела эту свою выборку и может быть склонна к излишней подгонке, занижая оценку процента ошибок. Дополнительная тонкость здесь связана со сравнением моделей на разных данных.

3.4 Вопрос 4. Тест Вальда

Пусть мы рассматриваем линейную модель $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ с понятной нам матрицей Ω корреляции случайных ошибок.

В таким случаем методом максимального правдопободия мы получаем ОМНК-оценку:

$$\hat{\boldsymbol{\beta}}^{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$

Пусть мы желаем протестировать гипотезу:

$$H_0 : H\beta = r$$

$$H_1 : H\beta \neq r$$

F-тест – очень хрупкая конструкция, требующая выполнения большого числа условий. Тест Вальда же остается справедливым при широких отклонениях от этих условий.

Если выполняется нулевая гипотеза, то $H\hat{\boldsymbol{\beta}}$ должно быть близко к r, т.е.:

$$H\hat{\boldsymbol{\beta}} - r \sim N(0, H(X'\Omega^{-1}X)^{-1}H')$$

В итоге мы получаем симпатичную статистику:

$$W = (H\hat{\boldsymbol{\beta}} - r)'(H(X'\hat{\Omega}^{-1}X)^{-1}H')^{-1}(H\hat{\boldsymbol{\beta}} - r) \sim \chi^2(q)$$

3.5 Вопрос 5. Тест отношения правдоподобия

Тест отношения правдоподобия основан на той идее, что, если выполняется нулевая гипотеза, то отношение значений функции правдопободия должно быть близко к единице.

$$H_0 : H\beta = r$$

$$H_1 : H\beta \neq r$$

Отсюда выходит простая статистика:

$$LR = 2 \ln L_{UR} - 2 \ln L_R \sim \chi^2(q)$$

3.6 Вопрос 6. Тест множителей Лагранжа

В то время как тест Вальда использовал только оценки, полученные без учета ограничения, тест отношения правдоподобия использовал обе регрессии, тест множителей Лагранжа использует только оценки, полученные с использованием ограничений.

$$H_0 : H\beta = r$$

$$H_1 : H\beta \neq r$$

Сами множители Лагранжа берутся из решения задачи максимизации логарифма правдоподобия с учетом линейных ограничений. Главная идея в том, что при выполнении нулевой гипотезы все множители Лагранжа должны быть близки к нулю.

Из связи множителей и остатков u выходит милая статистика:

$$LM = \tilde{u}'\tilde{\Omega}^{-1}X(X'\tilde{\Omega}^{-1}X)^{-1}X'\tilde{\Omega}^{-1}\tilde{u} \sim \chi^2(q)$$

χ^2 из связи правдоподобия с ковариационной матрицей и ошибками.

Стоит отметить, что все эти замечательные тесты асимптотические.

3.7 Вопрос 7. Панельные данные

Панельные данные обладают очевидными преимуществами над пространственными выборками:

- Панельные данные включают много наблюдений, часто намного больше чем почти любая пространственная выборка.

- Они позволяют изучать динамику, да еще и для многих объектов;

- Панельные данные позволяют контролировать ненаблюдаемые и неизменные для объекта переменные.

3.8 Вопрос 8. Соотношение оценок

Рассмотрим истинную модель:

$$y_{it} = \beta x_{it} + \alpha_i + \varepsilon_{it}$$

При этом мы не наблюдаем α_i.

$\boxed{\text{а}}$ Нам хочется понять, когда самая простая МНК-оценка будет состоятельной и эффективной. Для этого должны выполняться условия теоремы Гаусса-Маркова.

Определенно нам нужен диагональный вид матрицы ковариаций $\alpha_i + \varepsilon_{it}$, да еще и гомоскедастичность. Значит, один способ – иметь только один период.

В другом варианте α_i обладает нулевой дисперсией между всеми объектами – просто единое постоянное значение для всех единиц и периодов.

Разумеется, при этом мы требуем выполнения разумных условий, не должно быть ужасных автокорреляций и корреляций между ошибками и регрессорами.

$\boxed{\text{б}}$ В определенных условиях мы готовы пожертвовать эффективностью МНК-оценок, если они состоятельны. При этом мы хотим, чтобы RE-оценки были состоятельны и эффективны.

В рамках предпосылок RE-моделей мы можем получить состоятельные оценки с помощью МНК, хотя для получение эффективных как раз и потребуется использовать ОМНК – оценивать RE-модель.

Условия:

- ε_{it}, u_i не коррелированы внутри себя и друг с другом для произвольных значений индексов;
- ε_{it}, u_i не коррелированы с $\boldsymbol{x}$ для любых индексов;
- $E(\varepsilon_{it}) = E(u_i) = 0$ во все моменты и места;
- $\operatorname{Var}(\varepsilon_{it}) = \sigma^2_\varepsilon$, $\operatorname{Var}(u_i) = \sigma^2_u$

Хотя можно просто гетероскедастичность.

$\boxed{\text{в}}$ Осталось понять, при каких условиях все оценки, кроме FE, будут несостоятельными.

Такая катастрофа происходит, когда α начинает коррелировать с x. После этого остается состоятельной только FE-оценка. Конечно, все еще не должно быть всяких эндогенностей и прочих катастроф.

3.9 Вопрос 9. Within-преобразование

Рассмотрим истинную модель:

$$y_{it} = \beta x_{it} + \alpha_i + \varepsilon_{it}$$

При этом мы не наблюдаем α_i. Выведем с помощью внутригруппового преобразования оценку для параметра β. Чтобы сделать это, для каждого объекта вычисляем среднее во времени:

$$\overline{y}_i = \beta_0 + \beta_1 \overline{x}_i + \alpha_i + \overline{\varepsilon}_i$$

Вычитаем и переходим к новым обозначениям:

$$y_{it} - \overline{y}_i = \beta_1 (x_{it} - \overline{x}_i) + (\varepsilon_{it} - \overline{\varepsilon}_i)$$

$$\tilde{y}_{it} = \beta_1 \tilde{x}_{it} + \tilde{\varepsilon}_{it}$$

Теперь мы можем применить обычный МНК без константы, средние берутся по обоим измерениям:

$$\hat{\beta}_1 = \frac{\overline{\overline{\tilde{x}_{it} \tilde{y}_{it}}}}{\overline{\overline{\tilde{x}^2_{it}}}}$$

3.10 Вопрос 10. Модель с первыми разностями

Рассмотрим истинную модель:

$$y_{it} = \beta x_{it} + \alpha_i + \varepsilon_{it}$$

При этом мы не наблюдаем α_i. Выведем с помощью модели в первых разностях оценку для параметра β.

$$y_{it} = \beta_0 + \beta x_{it} + \alpha_i + \varepsilon_{it}$$

$$y_{it-1} = \beta_0 + \beta x_{it-1} + \alpha_i + \varepsilon_{it-1}$$

Берем разницу:

$$\Delta y_{it} = \beta \Delta x_{it} + \Delta \varepsilon_{it}$$

Важный факт – в иксах только то, что меняется. И снова же:

$$\hat{\beta}_1 = \frac{\overline{\Delta x \Delta y}}{\overline{(\Delta x)^2}}$$

Магия.

3.11 Вопрос 11. R^2 для FE-моделей

Есть два эквивалентных подхода в FE-моделям: LSDV-подход и within-преобразование. Они дают одинаковые оценки. При этом для них высчитывается R^2.

LSDV-модель Запишем исходную регрессию:

$$y_{it} = \beta_0 + \sum \beta_j x_{it}^{(j)} + \gamma z_i + \varepsilon_{it}$$

Создаем n фиктивных переменных:

$$D1i = \begin{cases} 1, \text{ для 1-ого объекта} \\ 0, else \end{cases}$$

$$Dni = \begin{cases} 1, \text{ для n-ого объекта} \\ 0, else \end{cases}$$

В итоге модель размазывается:

$$y_{it} = \sum \beta_j x_{it}^{(j)} + \alpha_1 D1_i + \ldots \alpha_n Dn_i + \varepsilon_{it}$$

R^2_{LSDV} – обычный R^2 для этой модели. С ним связана та проблема, что из-за целых i фиктивных переменных случается сильный overfit, а R^2 раздувается до астрономических значений.

Within R^2 Таким образом, разумнее использовать что-то более приличное для оценки качества подгонки. Для модели, прошедшей внутригрупповое преобразование и оцененной, мы считаем R^2_{within}. Он демонстрирует, сколько дисперсии зависимой переменной во времени наши регрессоры объясняют.

3.12 Вопрос 12. Комментарий к FE-моделям

Очевидно, что в нашу модель с фиксированными эффектами можно всунуть не только какие-то неизменные во времения для каждого отдельного объекта черты, но и меняющиеся во времени общие для всех объектов параметры. Там получается двунаправленная модель.

Пусть у нас есть симпатичная однонаправленная моделька:

$$y_{it} = \beta_0 + \sum \beta_j x_{it}^{(j)} + \gamma z_i + \varepsilon_{it}$$

Если мы засунем в нее время, то мы получим двунаправленную модель. Время лучше засовывать своими фиктивными переменными, а не линейно.

$$y_{it} = \beta_0 + \sum \beta_j x_{it}^{(j)} + \gamma z_i + \sum \delta dT_j + \varepsilon_{it}$$

Такая модель позволяет учитывать влияние времени на переменные. Это может спасать нас, может вытаскивать влияние времени, сильно давящее на переменные в однонаправленной модели.

3.13 Вопрос 13. RE-модель

Процедура оценки Пусть у нас есть некая модель со случайными эффектами. Чтобы оценить ее, мы проводим следующую процедуру:

- В рамках предпосылок модели со случайными эффектами у нас будут состоятельны МНК-оценки. Тем не менее, они будут неэффективными.
- Если мы хотим эффективные оценки, то надо обратиться к GLS:

  1. Если мы знаем матрицу Ω, то сразу записываем ОМНК-оценки;
  2. Если мы не знаем ее, то строим оценку, оценивая дисперсию случайного эффекта и ошибки;
  3. Потом находим GLS-оценку:

$$\hat{\boldsymbol{\beta}}_{RE}^{GLS} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y$$

Пример I Пусть вот теперь в нашей чудесной жизни есть два объекта в три момента времени. Модель:

$$y_{it} = \beta_0 + \boldsymbol{x}'\boldsymbol{\beta} + u_i + \varepsilon_i$$

Сформулируем невероятные предпосылки для RE-модели:

- ε_{it}, u_i не коррелированы внутри себя и друг с другом для произвольных значений индексов;
- ε_{it}, u_i не коррелированы с $\boldsymbol{x}$ для любых индексов;
- $E(\varepsilon_{it}) = E(u_i) = 0$ во все моменты и места;
- $\mathrm{Var}(\varepsilon_{it}) = \sigma^2_\varepsilon$, $\mathrm{Var}(u_i) = \sigma^2_u$

Найдем теперь вид матрицы Ω для модели:

$$y_{it} = \beta_0 + \boldsymbol{x}'\boldsymbol{\beta} + \varkappa_{it}$$

$$\varkappa = u_i + \varepsilon_{it}$$

Довольно очевидно, что из этих предпосылок следует, что что-то ненулевое может быть только в рамках жизни одного объекта.

Если мы стоим на диагональном элементе, то мы смотрим $\mathrm{Cov}(\varkappa_\Upsilon, \varkappa_\Upsilon)$:

$$\mathrm{Cov}(\varkappa_\Upsilon, \varkappa_\Upsilon) = \mathrm{Var}(\varkappa_\Upsilon) = \sigma^2_\varepsilon + \sigma^2_u$$

Здесь мы воспользовались нашей невероятной предпосылкой о независимости.

На элементе вне диагональки слегка другое:

$$\mathrm{Cov}(\varkappa_\Upsilon, \varkappa_\triangle) = \sigma^2_u$$

Как бы увидели, что есть совпадающий индекс у u, вставили дисперсию, несовпадающий у ε – воспользовались некоррелированностью.

Вне диагонали стоят все нули, потому что все некоррелировано. Тогда матрица:

$$\Omega = \begin{pmatrix} \sigma_\varepsilon^2+\sigma_u^2 & \sigma_u^2 & \sigma_u^2 & 0 & 0 & 0 \\ \sigma_u^2 & \sigma_\varepsilon^2+\sigma_u^2 & \sigma_u^2 & 0 & 0 & 0 \\ \sigma_u^2 & \sigma_u^2 & \sigma_\varepsilon^2+\sigma_u^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_\varepsilon^2+\sigma_u^2 & \sigma_u^2 & \sigma_u^2 \\ 0 & 0 & 0 & \sigma_u^2 & \sigma_\varepsilon^2+\sigma_u^2 & \sigma_u^2 \\ 0 & 0 & 0 & \sigma_u^2 & \sigma_u^2 & \sigma_\varepsilon^2+\sigma_u^2 \end{pmatrix}$$

Пример II Пусть теперь случилось три объекта в два момента времени. Аналогичной логикой мы получаем матрицу:

$$\Omega = \begin{pmatrix} \sigma_\varepsilon^2+\sigma_u^2 & \sigma_u^2 & 0 & 0 & 0 & 0 \\ \sigma_u^2 & \sigma_\varepsilon^2+\sigma_u^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_\varepsilon^2+\sigma_u^2 & \sigma_u^2 & 0 & 0 \\ 0 & 0 & \sigma_u^2 & \sigma_u^2+\sigma_\varepsilon^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_\varepsilon^2+\sigma_u^2 & \sigma_u^2 \\ 0 & 0 & 0 & 0 & \sigma_u^2 & \sigma_\varepsilon^2+\sigma_u^2 \end{pmatrix}$$

3.14 Вопрос 14. Сравнение FE и RE

Уважаемый исследователь хочет воспользоваться формальным тестом, чтобы выбрать, использовать ли ему FE или RE-модель.

Чтобы сравнить модели с фиксированными и случайными эффектами, делаем тест Хаусмана. Если $\mathrm{Cov}(u_i, x_{i_0t_0}) = 0$, то оценка со случайными эффектами состоятельна.

$$H_0 : \mathrm{Cov}(u_i, x_{i_0t_0}) = 0$$
$$H_1 : \mathrm{Cov}(u_i, x_{i_0t_0}) \neq 0$$

В данном случае мы проводим обычный тест Хаусмана, проверяя состоятельность ОМНК-оценок. Расчетная статистика имеет вид:

$$\chi^2_{\text{расч}} = (\hat{\boldsymbol{\beta}}^{FE} - \hat{\boldsymbol{\beta}}^{RE})'(\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}^{FE}) - \widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}^{RE})^{-1}(\hat{\boldsymbol{\beta}}^{FE} - \hat{\boldsymbol{\beta}}^{RE})$$

Мы сравниваем это с критическим значением $\chi^2_{\text{кр}}(\alpha, k)$, где k – число регрессоров.

3.15 Вопрос 15. Сравнение FE и pooled-регрессии

Иногда мы задаемся, есть ли в нашей индивидуальные эффекты, что эквивалентно вопросу, имеют ли группы разные константы.

$$H_0: \text{ Индивидуальных эффектов нет}$$

Чтобы сравнить, оцениваем сначала LSDV-модель с $D1_i \dots Dn_i$. Делаем тест на короткую и длинную регрессию, проверяя, равны ли нулю коэффициенты. И все.

Чтобы проверить это, используем любой тест на линейное ограничение.

3.16 Вопрос 16. Сравнением RE и pooled-регрессии

$$H_0: \text{ Случайных эффектов нет: } \operatorname{Var}(u_i) = \sigma_u^2 = 0$$

Если мы приняли нулевую гипотезу, то используем pooled регрессию. Для тестирования это гипотезы используем тест множителей Лагранжа.

Статистика:

$$LM = \frac{nT}{2(T-1)} \left(\frac{\sum_t (\sum_i e_{it})^2}{\sum_t (\sum_i e_{it}^2)} - 1 \right)^2 \sim \chi^2(1)$$

Сравнивается очевидным образом.

3.17 Вопрос 17. Динамическая панель

Иногда в жизни все несколько печальнее, чем хотелось бы. Под динамической панелью мы понимаем ситуацию, когда в числе регрессоров оказывается лагированное значение зависимой переменной. Что-то в духе:

$$y_{it} = \beta_0 + \boldsymbol{x_t'\beta} + \alpha y_{it-1} + \dots + \varepsilon_{it}$$

Сразу же становится ясно, что совсем простыми способами такое не возьмешь: если мы возьмем разность, то получим эндогенность:

$$y_{it} - y_{it-1} = \boldsymbol{(x_t' - x_{t-1}')\beta} + \alpha(y_{it-1} - y_{it-2}) + \dots + \varepsilon_{it} - \varepsilon_{it-1}$$

Самое очевидное решение – воспользоваться TSLS-оценками. На первом шаге мы регрессируем $(y_{it-1} - y_{it-2})$ на y_{it-2}. Кажется вполне валидным инструментом.

Другое очевидное решение возникает из GMM. Мы можем использовать не по одному значению как инструменту, а целую пирамиду. Так, в разности второго и первого периода мы будем иметь лишь один инструмент такой природы – значение первого периода; в разности третьего и второго – уже два, значения первого и второго периодов. Так выстраивается целая пирамида инструментов.

Из всего этого возникают горы моментных условий, из которых мы в итоге находим оценку наших параметров.

3.18 Вопрос 18. Об экспериментах

Определение 3.18.1. Treatment effect – изменения значение интересующего нас параметра в результате воздействия.

$$TE = Y_i(1) - Y_i(0)$$

Определение 3.18.2. Average treatment effect – математическое ожидание treatment effect для всей генеральной совокупности:

$$ATE = E(Y_i(1) - Y_i(0))$$

Определение 3.18.3. Selection bias – ситуация, при которой от характеристик объекта зависит, подвергался ли он воздействию.

Определение 3.18.4. Random assignment – метод назначения воздействия, при котором отнесение объекта в ту или иную группу не зависит от его характеристик.

3.19 Вопрос 19. Метод разности разностей

Уважаемый исследователь хочет оценить влияние повышения минимального размера оплаты труда на занятость в ресторанах быстрого питания.

Исследователь располагает панельными данными об уровне занятости в ресторанах быстрого питания в двух штатах, в одном из которых минимальный размер оплаты труда был увеличен. Опишем, как исследователь может оценить интересующий его эффект при помощи метода разность разностей.

Мы делим наши регионы на treatment и control group.

$$y_{ist} = \mu_i + \lambda_t + \delta D_{ist} + \varepsilon_{it}$$

Ясно, что:

$$E(y_{it}|treatm,1) = \mu_{tr} + \lambda_1 + \delta$$

$$E(y_{it}|treatm,0) = \mu_{tr} + \lambda_0$$

Для контрольной группы:

$$E(y_{it}|contr,1) = \mu_{cont} + \lambda_1$$

$$E(y_{it}|cont,0) = \mu_{cont} + \lambda_0$$

Дальше вычитаем:

$$\Delta_{tr} = E(y_{it}|treatm,1) - E(y_{it}|treatm,0) = \lambda_1 - \lambda_0 + \delta$$

$$\Delta_{contr} = E(y_{it}|contr,1) - E(y_{it}|cont,0) = \lambda_1 - \lambda_0$$

В итоге:

$$\Delta_{tr} = \Delta_{contr} + \delta$$

$$\delta = \Delta_{tr} - \Delta_{contr}$$

Методом разности разностей мы нашли нашу дельту.

3.20 Вопрос 20. Разрывная регрессия

Суть идеи, стоящей за разрывной регрессией, элементарна. Рассмотрим примитивнейшую постановку:

$$y_i = beta_0 + \beta_1 w_i + \beta_2 x_i + \varepsilon_i$$

$$w_i = \begin{cases} 1, x_i \geqslant x_0 \\ 0, x_i < x_0 \end{cases}$$

Очевидно, что такая регрессия имеет разрыв в точке x_0, если $\beta_1 \neq 0$. В этом точке происходит великолепное смещение линии на β_1. Эффект воздействия здесь тогда изображается β_1.

Суть метода в том, что это квазиэкспериментальный дизайн, который рассматривает эффекты воздействий, назначая порог, при переходе которого осуществляется воздействие. Так как часто возле порога лежат в целом похожие наблюдения, становится возможно довольно хорошо оценить ATE.

Классический пример здесь – награждения scholarship в зависимости от успеваемости и последующая успеваемость. Здесь сложно достичь чего-то простыми методами, потому что возникнет эндогенность; из соображений академической справедливости нельзя провести полноценную рандомизацию.

Таким образом, вариант оценки – ввести жесткий порог по успеваемости в 80%. При этом очень вероятно, что люди, имевшие 81% весьма похожи на людей, имевших 79%. На этом можно и оценить эффект.

Глава 4
Экзамен

4.1 Вопрос 1. Определение временного ряда

К этому моменту мы уже рассмотрели пространственные выборки и панельные данные. Естественным образом для рассмотрения остаются данные, имеющие только временную характеристику.

Определение 4.1.1. Временным рядом мы будем называть последовательность измерений, сделанных:

- За непрерывный временной интервал;
- Из последовательных измерений на этом интервале;
- С равным расстоянием во времени между каждыми двумя последовательными измерениями;
- Не более чем с одним измерением на каждый отдельный момент времени.

Достаточно очевидны принципиальные различия с пространственными данными:

- Во временных рядах есть упорядочивание по времени, данные жестко привязаны ко своему времени;
- Многие свойства МНК полагались на случайное извлечение наблюдений из генеральной совокупности. Здесь нет такой простой случайности.

Альтернативно можно определять временной ряд как последовательность случайных переменных, индексируемых индексом t. Это определение связывает временные ряды со случайными процессами[1].

[1] С дискретным временем, конечно.

4.2 Вопрос 2. Стационарность

Достаточное большое количество результатов в теории временных рядов опираются на идею стационарности. Ее определение:

Определение 4.2.1. Случайный процесс $\{x_t : t = 1, 2, \dots\}$ называется стационарным, если для любого набора индексов $1 \leqslant t_1 < t_2 < \dots < t_m$ совместное распределение $(x_{t_1}, x_{t_2}, \dots, x_{t_m})$ совпадает с совместным распределением $(x_{t_1+h}, x_{t_2+h}, \dots, x_{t_m+h})$ для любого целого $h \geqslant 1$.

Это свойство еще называют стационарностью в узком смысле (сильной стационарностью). Иногда нам достаточно более слабой формы:

Определение 4.2.2. Случайный процесс $\{x_t : t = 1, 2, \dots\}$ с конечным вторым моментом[2] называется ковариационно (слабо) стационарным, если

1. $\mathrm{E}(x_t) = const$;
2. $\mathrm{Var}(x_t) = const$;
3. Для любых $t, h \geqslant 1$ $\mathrm{Cov}(x_t, x_{t+h})$ зависит только от h.

Из этих определений немедленно ясно, что из сильной стационарности тривиальным образом следует слабая. Конечно, в обратную сторону такой связи нет, потому что слабая стационарность предполагает совпадения только первых и вторых моментов.

Надо заметить, что понятие стационарности описывает совместные распределения процесса во времени, поэтому нам нужно ввести еще одно определение – слабой зависимости.

Определение 4.2.3. Стационарный случайным процесс называется слабо зависимым, если $\mathrm{Corr}(x_t, x_{t+h}) \to 0$ при $h \to 0$ «достаточно быстро»[3].

Суть этого определения – требование асимптотической некоррелированности: чем дальше переменные во времени друг от друга, тем слабее между ними должна быть связь.

При этом требование слабой зависимости очень важно – именно оно заменяет предпосылку случайного извлечения из генеральной совокупности в модификациях Центральной Предельной Теоремы и Закона больших чисел.

---

[2] $\mathrm{E}(x_t^2) < \infty$.

[3] Есть несколько формальных определений слабой зависимости. Одно из них требует равномерной ограниченности последовательности вторых моментов, определенного поведения частичных сумм дисперсий и частичных сумм обратных к дисперсиям величин и асимптотической нормальности преобразованных величин x_t.

4.3 Вопрос 3. Интегрированность

Достаточно естественно, что многие ряды не обладают свойством стационарности. Таким образом, важная черта – порядок интегрированности таких рядов.

Определение 4.3.1. Временной ряд называется интегрированным порядка k, если только его k-ая разность стационарна (∇_k), а все предыдущие разности нестационарны.

4.4 Вопрос 4. Информационные критерии

Часто для сравнения моделей мы используем информационные критерии, возникающие естественным образом из теории информации и статистики.

Критерий Шварца Байесов (или Шварца) информационный критерий определяется как:

$$BIC = -2\ln \hat{L} + k\ln n$$

Здесь $\hat{L}$ – максимизированное значение функции правдоподобия, k – число свободных параметров, оцениваемых в модели, n – число наблюдений.

Дополнительно в нормальном случае критерий может быть переписан иначе[4]:

$$BIC = n\ln\left(\frac{\sum(x_i - \hat{x}_i)^2}{n}\right) + k\ln(n)$$

Предпочитаются модели с меньшим значением BIC. Ограничение критерия – сложности в его работе в пространствах высокой размерности.

Критерий Акаике Другой возможный критерий – критерий Акаике:

$$AIC = 2k - 2\ln L$$

Через критерий Акаике можно сравнивать модели в терминах вероятности потери информации. Предпочитается модель с минимальным значением критерия. Стоит отметить, что критерий Акаике меньше пенализует число

[4]Иногда встречается версия, где происходит деление на n

параметров и способен приводить к выбору более сложных моделей. В случае линейной модели он эквивалентен C_p Мэллоу.

Есть скорректированная версия критерия – $AICc$. Для парной регрессии с нормальным шумом:

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}$$

Суть коррекции – больший штраф за количество параметров в малых выборках. Если использовать сам критерий, а не скорректированную версию в выборках, когда число наблюдений n не превосходит во много раз k^2, можно получить перепараметризованную модель.

Критерий Ханнана-Куинна Еще один критерий – критерий Ханнана-Куинна:

$$HQC = -2L_{max} + 2k \log \log n$$

Это достаточно слабый критерий с ограниченным применениям, во многом подобный BIC по свойствам.

Дополнительно используются (в разных контекстах) DIC, FIC и C_p Мэллоу.

4.5 Вопрос 5. $AR(p)$ процессы

Одна из самых простых моделей – одномерная модель авторегрессии порядка p:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p} + \varepsilon_t$$

Заменяя в процессе лаги переменной на лаговый оператор, мы получаем следующее:

$$y_t = \alpha_0 + \alpha_1 L y_t + \alpha_2 L^2 y_t + \cdots + \alpha_p L^p y_t + \varepsilon_t$$

Из этого немедленно следует обычным образом характеристическое уравнение:

$$1 - \alpha_1 z - \alpha_2 z^2 - \cdots - \alpha_p z^p = 0$$

Автокорреляционная функция $AR(p)$ процесса:

$$\rho(\tau) = \sum_{k=1}^{p} \alpha_k y_k^{-|\tau|}$$

Таким образом, становится ясно, что каждый действительный корень уравнения добавляет в функцию экспоненциально угасающую компоненту, а каждая пара сопряженных комплексных корней – экспоненциально подавленные колебания.

Если все $|z| > 1$, то процесс стационарен.

4.6 Вопрос 6. $MA(q)$ процессы

Альтернативный подход – модели скользящего среднего $MA(q)$:

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

Достаточно ясно, что модели скользящего среднего конечных порядков всегда стационарны. Это не всегда так для $MA(\infty)$, для обеспечения стационарности в этом случаются требуется налагать еще определенные ограничения.

4.7 Вопрос 7. Выбор $ARIMA$ модели

Естественное обобщение AR и MA моделей – модели $ARIMA$[5]. Модель $ARIMA(p,d,q)$ описывает ряд, интегрированный порядка d, с компонентами $AR(p)$ и $MA(q)$.

Традиционный способ выбрать $ARIMA$ модель – метод Бокса-Дженкинса, включающий три шага в индентификации модели:

1. Вопрос стационарности и сезонности в данных: первый шаг – определение, стационарен ли ряд и есть ли какая-то существенная сезонность. Стационарность можно определять любым стандартным способом, сезонность обычно определяется по графикам автокорреляции или спектральным анализом.

[5] $ARIMAX$, если есть еще экзогенные переменные.

Если ряд нестационарен, то Бокс и Дженкинс рекомендуют брать разности ряда. Тем не менее, возможно и подгонять кривую, а потом вычитать подогнанные значения.

При обнаружении сезонности ее обычно не убирают, а корректируют порядки лагов для учета этой сезонности

2. После определения порядка интегрированности выбирают параметры p и q. Есть два главных подхода сравнения $ARMA(p,q)$ моделей на этом шаге – использование $AICc$ и анализ графиков ACF и $PACF$ – автокорреляции и частной автокорреляции. Характерные варианты по форме ACF:

   - Экспоненциально убывающая – AR модель. Порядок p определяется по $PACF$;
   - Альтернирующая между положительными и отрицательными значениями, стремится к нулю – AR модель. Порядок p определяется по $PACF$;
   - Один или больше пиков, остальные не отличаются значимо от нуля – MA, порядок определяется периодом, где ACF становится равна нулю;
   - Убывание начинается через несколько лагов – смешанная модель $ARMA$;
   - Ноль или очень близко к тому – данные случайны;
   - Высокие значения на фиксированных интервалах – есть сезонный AR член;
   - Нет стремления к нулю – ряд не стационарен.

   В первом и третьем случае порядки p и q, соответственно, определяются порядом p лага, при котором происходит пик в $PACF$, порядком q, на котором происходит пик ACF.

3. Модель оценивается ММП. При этом для сравнения моделей используют информационные критерии, а адекватность проверяется анализом остатков модели, которые не должны иметь автокорреляцию. Для этого можно проверять на равенство нулю каждый отдельный коэффициент или тестировать их совместно.

4.8 Вопрос 8. Автокорреляционная функция

Остается ввести два существенных понятия:

Определение 4.8.1. Для произвольного случайного процесса функцией автокорреляции между моментами времени s и t называется функция вида:

$$R(s,t) = \frac{\mathrm{E}\left[(X_t - \mu_t)(X_s - \mu_s)\right]}{\sigma_t \sigma_s}$$

Если мы работаем с дискретным стационарным в широком смысле процессом, мы получаем следующую эмпирическую функцию:

$$\hat{R}(k) = \frac{1}{(n-k)\sigma^2} \sum_{t=1}^{n-k} (X_t - \mu)(X_{t+k} - \mu)$$

Достаточно ясно, что для AR процессов автокорреляционная функция очень медленно сходится к нулю, а для MA она обрывается после лага q.

Рядом с простой функцией обитает и ее частный аналог. Неформально говоря, частная автокорреляция x_t и x_{t+k} – автокорреляция между этими двумя переменными, очищенная от линейного влияния $x_{t+1} \dots x_{t+k-1}$:

$$\alpha(k) = \mathrm{Corr}(x_{t+k} - P_{t,k}(x_{t+k}), x_t - P_{t,k}(x_t))$$

Здесь P – проекция на линейную оболочку $x_{t+1} \dots x_{t+k-1}$. На основе выборочных коэффициентов корреляции вычисляется $PACF$.

Понятно, что для AR процессов близка к нулю после лага p, а для MA $PACF$ бесконечна и постепенно сходится к нулю.

4.9 Вопрос 9. Тест Дики-Фуллера

Один из самых стандартных тестов в рамках теории временных рядов – тест Дики-Фуллера, проверяющий, если ли в авторегрессионной модели единичный корень. Альтернативные гипотезы – стационарность или тренд-стационарность.

Самый простой пример – $AR(1)$:

$$y_t = \rho y_{t-1} + \varepsilon_t$$

Эта модель нестационарна, если $\rho = 1$[6]. Перепишем нашу модель:

$$\nabla y_t = (\rho - 1)y_{t-1} + \varepsilon_t = \delta y_{t-1} + \varepsilon_t$$

После этого оценивается одна из трех основных версий и тестируется гипотеза $H_0 : \delta = 0$. Возникает единственная неловкая ситуация в том, что в рамках нулевой гипотезы процесс интегрирован порядка 1, поэтому нельзя использовать ЦПТ. Таким образом, t-статистика теста Дики-Фуллера имеет распределение Дики-Фуллера, которое следует использовать в рамках теста.

Таким образом, мы оцениваем одну из версий:

1. Тест на единичный корень:
$$\nabla y_t = \delta y_{t-1} + \varepsilon_t$$
2. Тест на единичный корень со сдвигом:
$$\nabla y_t = \alpha_0 + \delta y_{t-1} + \varepsilon_t$$
3. Тест на единичный корень с детерминированным[7] трендом:
$$\nabla y_t = \alpha_0 + \alpha_1 t + \delta y_{t-1} + \varepsilon_t$$

Все это время мы тестируем гипотезы:

$$H_0 : \delta = 0$$

Статистика имеет вид:

$$DF = \frac{\hat{\delta}}{se(\hat{\delta})}$$

Если она меньше критического значения из нужной таблицы, отвергаем релевантную нулевую гипотезу.

Если же мы имеем более сложный процесс $AR(p)$, мы можем использовать расширенный тест Дики-Фуллера (ADF).

В общем случае рассматривается модель:

$$\nabla y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \nabla y_{t-1} + \cdots + \delta_{p-1} \nabla y_{t-p+1} + \varepsilon_t$$

[6]Конечно, при $\rho > 1$ мы получаем банальную взрывную модель.

[7]В отличие от стохастического тренда для нестационарного ряда (как в нулевой гипотезе)

На нее можно наложить ограничения отсутствия константы и тренда. Изучается поведение γ. Статистика:

$$DF = \frac{\hat{\gamma}}{se(\hat{\gamma})}$$

Нулевая гипотеза остается той же. Дополнительный вопрос в использовании этого тест – выбор порядка лагов, который может быть осуществлен с помощью информационных критериев.

4.10 Вопрос 10. Тест KPSS

Другой подход к определению тренд-стационарности процесса – тест KPSS (Kwiatkowski–Phillips–Schmidt–Shin). Он обратен тесту ДФ в том, что его нулевая гипотеза – тренд-стационарность процесса.

$$H_0: \text{ ряд тренд-стационарен}$$

$$H_1: \text{ ряд нестационарен}$$

Для проведения теста оцениваем регрессию y на константу и линейный тренд:

$$y_t = \delta + \gamma t + \varepsilon_t$$

По остаткам регрессии строится статистика:

$$KPSS = \sum_{i=1}^{T} \frac{\left(\sum_{j=1}^{i} e_j\right)^2}{\hat{\sigma}_\varepsilon^2}$$

Традиционно используется критическое значение 0.146, если статистика больше, мы отвергаем нулевую гипотезу. Для стационарности оценивается регрессия на константу и используется критическое значение 0.463.

Прочие тесты на единичный корень включают ADF-GLS тест и тест Филлипса-Перрона, использующие разные подходы к расширению теста Дики-Фуллера. Хотя тест Филлипса-Перрона и робастен к произвольной автокорреляции и гетероскедастичности, он работает хуже ADF-теста.

4.11 Вопрос 11. Тест Льюнг-Бокса

Тест Льюнг[8]-Бокса – тест, проверяющий отличается ли произвольная группа автокорреляций от нуля, проверяя общую случайность по некоторому числу лагов.

$$H_0 : \rho_1 = \cdots = \rho_k = 0$$

Тогда тестовая статистика имеет вид:

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}$$

В стандартном случае в рамках нулевой гипотезы эта статистика имеет распределение $\chi^2(h)$. При этом важно обратить внимание, что, если мы работает с $ARIMA(p,d,q)$ моделью, мы тестируем автокорреляцию остатков этим тестом. Тогда требуется поправка степеней свободы, распределение статистики – $\chi^2(h-p-q)$.

4.12 Вопрос 12. QLR-тест

Иногда в жизни и в данных происходят структурные сдвиги. Наша работа с ними различается в зависимости от того, знаем ли мы, когда гипотетически произошел этот сдвиг.

1. Если мы знаем мы можем специфицировать простую модификацию модели, в которую мы включаем фиктивную переменную сдвига. Проверка сдвига осуществляется проверкой равенства нулю коэффициентов при переменных, с которыми стоит переменная сдвига в момент времени t_0. Обычно можно использовать F-тест.

2. Если мы не знаем дату сдвига, нам становится несколько труднее. Тогда один из способов – задать начало и конец тестируемых моментов потенциального сдвига (например, искать между $t_0 = 10$ и $t_1 = 90$ из 100 периодов, но должно хватать данных для разумной оценки).

   После этого итеративно стоит проводить F-тест на структурный сдвиг в каждый момент между t_0 и t_1. Максимальная из F-статистик – QLR-статистика, критическое значение которой зависит от числа ограничений и обрезания краев по времени.

[8] Назван по Грете М. Льюнг

Тест может обнаруживать единственный разрыв, множественные разрывы и медленное изменение параметров. Если в данных есть один разрыв, то его дата оценивается самой большой из статистик.

4.13 Вопрос 13. Тест Грейнджера

Достаточно естественно обобщить AR-модель в векторный вид. Тогда мы получаем векторную модель авторегрессии (VAR):

$$\begin{cases} y_t = \rho + \alpha_1 y_{t-1} + \gamma_1 x_{t-1} + \alpha_2 y_{t-2} + \gamma_2 x_{t-2} + \dots + \varepsilon_t \\ x_t = \varkappa + \beta_1 y_{t-1} + \delta x_{t-1} + \beta_2 y_{t-2} + \delta_2 x_{t-2} + \dots + v_t \end{cases}$$

Такие модели достаточно удобны в прогнозировании. Суть выражения выше – позволить нам понять, может ли, например, x помогать нам прогнозировать y (при этом мы контролируем прошлые значения y). Если x полезен, то мы говорим, что он – причина по Грейнджеру для y.

Если формально, то:

Определение 4.13.1. x – причина по Грейнджеру для y, если

$$\mathrm{E}(y_t | I_{t-1}) \neq \mathrm{E}(y_t | J_{t-1})$$

Здесь I_{t-1} содержит информацию по y и x до момента времени $t-1$, а J_{t-1} содержит только по y.

Таким образом, мы тестируем (для влияния x на y) гипотезу:

$$H_0 : \boldsymbol{\gamma} = \overline{0}$$

Мы используем уместную тестовую статистику. Дополнительно мы можем проверять обратную связь.

4.14 Вопрос 14. PooS-прогноз

В мире есть несколько подходов для оценки качества прогнозов:

- In-sample: мы оцениваем модель на основе всех имеющихся данных и сравниваем получившиеся результаты с фактическими данными. Этот подход не дает адекватной оценки качество прогнозов: он занижает фактическую оценку и ведет к оверфиту;

- Out-of-sample: оценить модель, сделать на ее основе прогноз на следующий период. В следующем периоде оценить модель заново, посчитав ошибку, и так далее. В итоге будет достаточно точная картина качества предсказаний модели;

- Pseudo-out-of-sample: настоящие OOS оценки требует много времени, поэтому часто используется имитация этого подхода. Процедура оценки начинается с какого-то прошлого периода t_0 и ведется до последних доступных данных. Это позволяет получить простую оценку качества.

4.15 Вопрос 15. Тест Бреуша-Годфри

Тест Бреуша-Годфри дает нам мощный инструмент для тестирования наличия автокорреляции в остатках. Он действует и для стохастических регрессоров, может проверять любую степень автокорреляции и разрешает использовать лагированные значения объясняемой переменной как регрессоры.

$$H_0 : \rho_1 = \cdots = \rho_p = 0$$

Пусть есть некоторая линейная модель:

$$y_t = \alpha_0 + \alpha_1 x_t + \cdots + u_t$$

При этом остатки u_t подчиняются процессу $AR(p)$:

$$u_t = \rho_1 u_{t-1} + \rho u_{t-2} + \cdots + \rho_p u_{t-p} + \varepsilon_t$$

Тогда исходная модель МНК-оценивается, из нее получаются $\hat{u}_t$ – подогнанные остатки. Тогда подгоняется следующая модель:

$$\hat{u}_t = \alpha_0 + \alpha_1 x_t + \cdots + \rho_1 \hat{u}_{t-1} + \rho \hat{u}_{t-2} + \cdots + \rho_p \hat{u}_{t-p} + \varepsilon_t$$

При этом статистика nR^2 имеет асимптотическое распределение $\chi^2(p)$. Важно заметить, что $n = T - p$ здесь.

4.16 Вопрос 16. Тест Дарбина-Уотсона

Классический тест на автокорряляцию первого порядка – тест Дарбина-Уостона. Его тестовая статистика имеет следующий вид:

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

Проверяется следующая гипотеза:

$$H_0 : \rho = 0$$

$$u_t = \rho u_{t-1} + \varepsilon_t$$

При этом есть аппроксимация:

$$DW \approx 2(1 - \hat{\rho})$$

При этом выходит достаточно сложная конструкция теста из-за особого распределения этой статистики:

- Если $DW < d_L$, то мы можем сделать вывод, что $\rho > 0$.
- Если $d_u < DW < 4 - d_u$, то $\rho = 0$.
- Если $DW > 4 - d_L$, то $\rho < 0$

В остальных случаях тест не дает результатов. При этом тест очень хрупкий – включать лаг зависимой переменной нельзя, убирать константу нельзя, автокорреляция только первого порядка может быть.

Надо заметить, что статистика Дарбина-Уотсона смещена для $ARMA$-моделей, автокорреляция недооценивается. Дарбин предложил дополнительно вычислять h-статистику:

$$h = \left(1 - \frac{1}{2}DW\right)\sqrt{\frac{T}{1 - T\,\widehat{\mathrm{Var}}(\hat{\beta}_1)}}$$

β_1 – коэффициент при y_{t-1}. Такая формулировка позволяет включать лаг, h-статистика имеет нормальное распределение, но в некотороых случаях она не определена[9].

[9] Когда $T\,\widehat{\mathrm{Var}}(\hat{\beta}_1) \geqslant 1$

4.17 Вопрос 17. Фильтр Ходрика-Прескотта

Достаточно большое количество реальных данных имеют циклические или сезонные колебания. Чтобы убрать их, часто используют сглаживание данных.

Один из классических инструментов – фильтр Ходрика-Прескотта. Пусть y_t – логарифмы некоторого ряда. Пусть y_t состоит из трендовой компоненты τ и циклической компоненты c, т.е.:

$$y_t = \tau_t + c_t + \varepsilon_t$$

Тогда для адекватных положительных значений λ существует решение минимизационной задачи тренд τ:

$$\min_\tau = \left(\sum_{t=1}^{T} (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} \left[(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t_1}) \right]^2 \right)$$

Первый член этого выражения штрафует циклическую компоненту, а второй – колебания в приростах трендовой компоненты. Чем выше λ, тем сильнее сглаживание.

Утверждение 4.17.1. λ должна варьироваться в четвертой степени частоты наблюдений.

Конечно, фильтр оптимален только для $I(2)$ данных с нормальным шумом и чисто историческим анализом.

4.18 Вопрос 18. Декомпозиция $AR(1)$

Теорема I.13 (Теорема Вольда)**.** Любой ковариационно стационарный временной ряд может быть записан как сумма двух временных рядов – стохастического и детерминированного.

В частности, с помощью этой теоремы можем преобразовать $AR(1)$-процесс в $MA(\infty)$. Покажем это для процесса (с $|\phi| < 1$:

$$y_t = \phi y_{y-1} + \varepsilon_t$$

Запишем через лаговый оператор:

$$(1 - \phi L) y_t = \varepsilon_t$$

Преобразуем:

$$y_t = (1 - \phi L)^{-1} \varepsilon_t$$

Воспользуемся разложением в ряд:

$$(1 - z)^{-1} = 1 + z + z^2 + z^3 + \dots, |z| < 1$$

Тогда:

$$y_t = (1 - \phi L)^{-1} \varepsilon_t = (1 + \phi L + \phi^2 L^2 + \dots) \varepsilon_t = \sum_{h=0}^{\infty} \phi^h \varepsilon_{t-h}$$

Аналогичным образом можно преобразовать $AR(p)$ в $MA(\infty)$, $MA(q)$ в $AR(\infty)$.

4.19 Вопрос 19. Прогноз для $AR(1)$

Рассмотрим теперь простейший пример с прогнозированием. Пусть есть процесс:

$$y_t = \alpha + \theta y_{t-1} + \varepsilon_t, |\theta| < 1$$

Построим прогноз. Достаточно очевидно, что лучший прогноз – условное матожидание вида:

$$\mathrm{E}(y_{t+1}|y_t) = \mathrm{E}(\alpha + \theta y_t + \varepsilon_t) = \alpha + \theta y_t$$

Если мы хотим получить еще на шаг вперед, получается так:

$$\mathrm{E}(y_{t+2}|y_t) = \mathrm{E}(\alpha + \theta y_{t+1} + \varepsilon_{t+1}|y_t) = \alpha + \theta(\alpha + \theta y_t)$$

Этот тренд достаточно ясен. Если мы назовем $f(z) = \alpha + \theta z$, то:

$$\mathrm{E}(y_{t+s}|y_t) = f^s(y_t)$$

Здесь степень функции берется в плане композиции.

Достаточно легко вычислить и дисперсию такой конструкции[10]:

$$\mathrm{Var}(\hat{y}_{t+1} - y_{t+1}) = \mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2$$

[10]Или возможна какая-то форма гетероскедастичности.

Дальше уже чуть менее хорошо:

$$\mathrm{Var}(\hat{y}_{t+2} - y_{t+2}) = \mathrm{Var}(\alpha + \theta(\alpha + \theta y_t + \varepsilon_{t+1}) - \alpha + \theta(\alpha + \theta y_t) - \varepsilon_{t+2}) =$$

$$= \mathrm{Var}(\theta\varepsilon_{t+1} - \varepsilon_{t+2}) = \theta^2\sigma^2 + \sigma^2 = (1 + \theta^2)\sigma^2$$

Продолжая в том же катастрофическом духе, мы получим:

$$\mathrm{Var}(\hat{y}_{t+s} - y_{t+s}) = (1 + \theta^2 + \cdots + \theta^{2(s-1)})\sigma^2$$

Это тоже можно свернуть, конечно, как простую сумму геометрической прогрессии.

4.20 Вопрос 20. $ADL(p,q)$ модель

Одна из самых очевидных возможных модель – модель распределенных лагов (потенциально с авторегрессией). Ее общий вид:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_p y_{t-p} + \delta_0 x_t + \delta_1 x_{t-1} + \cdots + \delta_q x_{t-q} + \varepsilon_t$$

Состоятельность МНК-оценок для такой модели может существовать в достаточно широких условиях, описанных Песараном. Для модели $ADL(0,q)$ можно дать жесткие, но простые условия:

1. Существует слабая форма экзогенности: $\mathrm{E}(\varepsilon_t | x_{t-1} \ldots x_{t-q})$;

2. Нет линейной связи между регрессорами;

3. Ряды стационарны;

4. Ряды обладают конечными восьмыми моментами распределения.

Естественно, что существуют и бесконечные модели распределенных лагов, имеющие вид:

$$y_t = \beta_0 + \delta_0 x_t + \delta_1 x_{t-1} + \cdots + \varepsilon_t$$

Такую модель очень сложно состоятельно оценить, но мы можем наложить некоторые ограничения на вид лагов. Простейшее – геометрический лаг Койка:

$$\delta_j = \gamma \rho^j, |\rho| < 1$$

Тогда простым преобразованием мы получаем:

$$y_t = \beta_0 + \gamma x_t + \rho y_{t-1} + v - i$$

Дополнительно можно рассматривать модель с обратным полиномиальным лагом, гамма-лагом, рациональным лагом и прочими видами.

4.21 Вопрос 21. Мультипликатор в $ADL(1,1)$

Рассмотрим теперь $ADL(1,1)$:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \delta_0 x_t + \delta_1 x_{t-1} + \varepsilon_t$$

Выведем эффект от перманентного увеличения x на единицу:

- В первом периоде эффект будет равен δ_0
- Во втором периоде эффект становится $\beta_1\delta_0 + \delta_1$
- В третьем периоде $\beta_1(\beta_1\delta_0 + \delta_1)$, после чего просто копятся степени β_1 на этой скобке.

Тогда получаются два ряда, которые можно собрать вместе:

$$LRP = \frac{\delta_0}{1-\beta_1} + \frac{\delta_1}{1-\beta_1}$$

4.22 Вопрос 22. Проблема с автокорреляцией

Достаточно естественно, что во временных рядах случается связь между случайными ошибками во времени:

$$u_t = \rho_1 u_{t-1} + \rho u_{t-2} + \cdots + \rho_p u_{t-p} + \varepsilon_t$$

Автокорреляция не вызывает великой катастрофы в оценке моделей, оставляя оценки состоятельными, но делает их неэффективными, а ошибки несостоятельными.

Теорема I.14. Если

1. Случайный процесс следует линейной модели;
2. В выборке (и генеральной совокупности) нет констант или линейно зависящих от других переменных;
3. Объясняющие переменные экзогенны в смысле современности: $\mathrm{E}(u_t|\boldsymbol{x}_t) = 0$;
4. Ошибки гомоскедастичны в смысле современности: $\mathrm{E}(u_t|\boldsymbol{x}_t) = \sigma^2$;
5. Для любых $t \neq s$ $\mathrm{E}(u_t u_s|\boldsymbol{x}_s \boldsymbol{x}_t) = 0$,

то МНК-оценки состоятельны, асимптотически нормально распределены, все ошибки и статистики асимптотически корректны.

С автокорреляцией можно бороться двумя способами:

- Робастными ошибками;
- ОМНК: вообще, не зная точной формы, сложно получить внятное понимание, как надо действовать. Тем не менее, часто мы можем получить смещенные, но состоятельные и более эффективные оценки с помощью доступного ОМНК. Два способа для $AR(1)$ для серийной корреляции:

  1. Оценка Кохрейна-Оркатта: мы оцениваем нашу регрессию, после чего по остаткам оцениваем ρ:

     $$e_t = \rho e_{t-1} + v_t$$

     После этого мы делаем преобразование переменных:

     $$y_t - \hat{\rho} y_{t-1} = \alpha - \hat{\rho}\alpha + \beta x_t - \hat{\rho}\beta x_{t-1} + v_t$$

     Мы можем повторять эту процедуру, пока не будем получать одни и те же оценки;
  2. Оценка Праиса-Винстена не выбрасывает первое наблюдение, что может быть важно в малых выборках. Оно преобразует его так:

     $$y_1* = \sqrt{1 - \hat{\rho}^2 y_1}$$
     $$x_1* = \sqrt{1 - \hat{\rho}^2 x_1}$$

4.23 Вопрос 23. HAC-ошибки

Рассмотрим произвольную линейную модель:

$$y_t = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_k x_{tk} + u_t$$

Из этого достаточно просто получить HAC-ошибки. Пусть $\varkappa(\hat{\beta}_1)$ – стандартная ошибка коэффициента из МНК. Пусть $\hat{\sigma}$ – стандартная ошибка модели. Пусть $\hat{r}_t$ – остатки из регрессии x_{t1} на $x_{t2} \ldots x_{tk}$.

Тогда для данного целого числа g мы определим:

$$\hat{v} = \sum_{t=1}^{n} \hat{\alpha}_t^2 + 2 \sum_{h=1}^{g} \left[1 - \frac{h}{g+1}\right] \left(\sum_{t=h+1}^{n} \hat{a}_t \hat{a}_{t-h}\right)$$

Здесь $\hat{a}_t = \hat{r}_t \hat{u}_t$. Тогда мы получаем HAC-ошибку:

$$se(\hat{\beta}_1) = \left(\frac{\varkappa(\hat{\beta}_1)}{\hat{\sigma}}\right)^2 \sqrt{\hat{v}}$$

4.24 Вопрос 24. Ложная регрессия

Если в рамках обсуждения пространственных выборок мы называли ложной регрессией ситуацию, когда две переменных тесно связаны, но только через связь через третью переменную (или просто очень похожи случайным образом), то во временных рядах такая ситуация тоже возможна. Дополнительно такая связь может обнаруживаться, если у временных рядом есть возрастающий или убывающие тренды. В таком случае эта проблема решается включение тренда в уравнение.

Дополнительно такая проблема может возникать с $I(1)$-процессами, даже если в них нет тренда. Грейнджер и Ньюболд показали симуляциями, что очень часто при регрессии одного независимого процесса на другой наблюдается значимая связь, причем проблема слишком частого отклонения нулевой гипотезы становится только серьезнее с ростом выборки.

Проблемы ситуации с ложной регрессией заключаются в том, что статистики в таком случае не имеют асимптотически нормального распределения, а R^2 не сходится по вероятности в R^2 генеральной совокупности, а сходится к некоторой случайной переменной.

4.25 Вопрос 25. Коинтеграция

Конечно, есть случай, когда регрессия с нестационарными рядами оказывается не ложной, а отражает действительную долгосрочную связь. Это случай коинтегрированных рядов:

Определение 4.25.1. Два ряда x_t, $y_t \in I(1)$ коинтегрированы, если существует число θ такое, что ряд $y_t - \theta x_t$ стационарен.

Тогда это число θ – коинтеграционный коэффициент, а вектор $\begin{pmatrix} 1 \\ -\theta \end{pmatrix}$ – коинтегрирующий вектор.

Этим определением достаточно легко пользоваться. Для проверки коинтегрированности надо рассмотреть сами ряды, проверить их нестационарность, сделать регрессию одного на другой и проверить стационарность остатков регрессии.

4.26 Вопрос 26. GARCH

Иногда в модели случается достаточно ехидная ситуация, когда в ней проявляется гетероскедастичность особого вида. Такой особый вид – $GARCH(p,q)$:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_p u_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_q \sigma_{t-q}^2$$

Достаточно ясно, что это всего лишь гетероскедастичность, от которой легко излечить HAC-ошибками, робастными к произвольным формам гетероскедастсчности.

Саму модель достаточно легко специфицировать – надо сперва подогнать к исходному процессу лучшую $AR(q)$ модель, вычислить автокорреляции ошибок в ней. После этого можно использовать тест Льюнг-Бокса.

На тему ARCH есть очень много модификаций – NGARCH, EGARCH, QGARCH, GARCH-M, *etc.*

4.27 Вопрос 27. ARIMA с GARCH

Рассмотрим, например, $ARIMA(1{,}0{,}3)$ с $GARCH(0{,}2)$. Тогда:

$$y_t = \alpha + \beta y_{t-1} + \delta_1 \varepsilon_{t-1} + \delta_2 \varepsilon_{t-2} + \delta_3 \varepsilon_{t-3} + \varepsilon_t$$

$$\sigma_t^2 = \gamma_0 + \gamma_1 \sigma_{t-1}^2 + \gamma^2 \sigma_{t-2}^2$$

4.28 Вопрос 28. Парная регрессия с $ARCH(1)$

Рассмотрим модель:

$$y_t = \theta_0 + \theta_1 x_t + \varepsilon_t$$

При этом у нас $GARCH(1{,}0)$:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$$

Достаточно просто найти теперь безусловную дисперсию в модели[11]:

$$\mathrm{E}(\sigma_t^2) = \alpha_0 + \alpha_1 \mathrm{E}(\varepsilon_{t-1}^2)$$

Легко используем предпосылку, что ошибки имеют нулевое матожидание:

$$\sigma^2 = \alpha_0 + \alpha_1 \sigma^2$$

Значит:

$$\sigma^2 = \frac{\alpha_0}{1 - \alpha_1}$$

Если мы вдруг узнаем, что в прошлый период $\varepsilon = -0.1$, мы можем запросто посчитать и условную дисперсию:

$$\mathrm{E}(\sigma_t^2 | \varepsilon_{t-1} = -0.1) = \alpha_0 + 0.01\alpha_1$$

4.29 Вопрос 29. Парная регрессия с $GARCH(1{,}1)$

Рассмотрим модель:

$$y_t = \theta_0 + \theta_1 x_t + \varepsilon_t$$

При этом у нас $GARCH(1{,}1)$:

[11] При условии, что сам процесс не сходит с ума.

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

Найти здесь в разумных условиях безусловную дисперсию тоже элементарно:

$$\mathrm{E}(\sigma_t^2) = \mathrm{E}(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2) = \alpha_0 + \alpha_1 \mathrm{E}(\varepsilon_{t-1}^2) + \beta_1 \mathrm{E}(\sigma_{t-1}^2)$$

$$\sigma^2 = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$$

4.30 Вопрос 30. Векторная авторегрессия

Вернемся в векторной модели:

$$\begin{cases} y_t = \rho + \alpha_1 y_{t-1} + \gamma_1 x_{t-1} + \alpha_2 y_{t-2} + \gamma_2 x_{t-2} + \dots + \varepsilon_t \\ x_t = \varkappa + \beta_1 y_{t-1} + \delta x_{t-1} + \beta_2 y_{t-2} + \delta_2 x_{t-2} + \dots + v_t \end{cases}$$

Пусть мы в итоге некоторых глубоких измышлений сократили ее и поменяли названия коэффициентов на максимально неинтуитивные:

$$\begin{cases} y_t = \beta_1 y_{t-1} + \gamma_1 x_{t-1} + u_{1t} \\ x_t = \beta_2 y_{t-1} + \gamma_2 x_{t-1} + u_{2t} \end{cases}$$

Достаточно естественно использовать VAR-модели для прогнозирования. Запишем векторно модель:

$$\boldsymbol{y}_t = \boldsymbol{\beta} \boldsymbol{y}_{t-1} + \boldsymbol{u}_t$$

Как и всегда с прогнозами, лучше условного матожидания не уйти:

$$\mathrm{E}(\boldsymbol{y}_{t+1} | \boldsymbol{y}_t) = \boldsymbol{\beta} \boldsymbol{y}_t$$

Тогда можно построить и второй прогноз:

$$\mathrm{E}(\boldsymbol{y}_{t+2} | \boldsymbol{y}_t) = \boldsymbol{\beta} \boldsymbol{\beta} \boldsymbol{y}_t$$

Тогда, например, мы получаем, что:

$$\hat{y}_{t+2} = (\beta_1^2 + \gamma_1 \beta_2) y_t + (\beta_1 \gamma_1 + \gamma_1 \gamma_2) x_t$$

Часть II

Практика

Глава 5
Семинары 1-8

5.1 Семинар 1. Введение

5.1.1 Задача 1

В этом задании нам предлагается вспомнить свойства математического ожидания, дисперсии и ковариации.

Предположим, что мы располагаем определенной суммой денег для инвестирования. Для простоты будем считать, что эта сумма равна одному доллару.

Мы планируем инвестировать долю w в ценную бумагу А, а оставшуюся долю $(1-w)$ в ценную бумагу .

Предположим, что годовая доходность от вложения в ценную бумагу А составляет R_A (это означает, например, что вложение одного доллара в эту ценную бумагу приносит через год $1+R_A$ долларов). Известно, что это случайная величина, а с математическим ожиданием 0.3 и стандартным отклонением 0.1.

Годовая доходность от вложения в ценную бумагу В составляет R_B. Известно, что – это случайная величина с математическим ожиданием 0.5 и стандартным отклонением 0.2.

Корреляция между случайными величинами составляет 0.25.

Если мы вложим долю w в ценную бумагу А, а долю $(1-w)$ – в ценную бумагу В, то доходность вашего портфеля ценных бумаг составит:

$$R = wR_A + (1-w)R_B$$

.

$\boxed{\text{а}}$ Пусть w=0.5.

Найдем сперва математическое ожидание R:

$$\mathrm{E}(R) = \mathrm{E}(wR_A + (1-w)R_B) = w\,\mathrm{E}(R_A) + (1-w)\,\mathrm{E}(R_B) = 0.15 + 0.25 = 0.4.$$

Найдем теперь дисперсию:

$$\begin{gathered}\operatorname{Var}(R)=\operatorname{Var}((wR_A+(1-w)R_B)= \\ =\operatorname{Var}(wR_A)+\operatorname{Var}((1-w)R_B)+2\operatorname{Cov}(wR_A+(1-w)R_B)= \\ =w^2\operatorname{Var}(R_A)+(1-w)^2\operatorname{Var}(R_B)+2w(1-w)\operatorname{Cov}(R_A.R_B)\end{gathered}$$

Найдем ковариацию:

$$\operatorname{Cov}(x,y)=\operatorname{Corr}(x,y)\sqrt{\operatorname{Var}(x)\operatorname{Var}(y)}$$

$$\operatorname{Cov}(R_A,R_B)=\operatorname{Corr}(R_A,R_B)\sqrt{\operatorname{Var}(R_A)\operatorname{Var}(R_B)}=0{,}25\cdot 0{,}1\cdot 0{,}2=0{,}005$$

Подставляем в формулу:

$$\operatorname{Var}(R)=0.5^2+0.1^2+0.5^2\cdot 0.2^2+2\cdot 0.5\cdot 0.5\cdot 0.005=0.015$$

Значит, $\sqrt{\operatorname{Var}(R)}=0.123$

$\boxed{\text{б}}$ При $w=0.5$ найдем ковариацию доходности портфеля и доходности ценной бумаги A.

$$\begin{gathered}\operatorname{Cov}(R,R_A)=\operatorname{Cov}(wR_A+(1-w)R_B,R_A)= \\ =\operatorname{Cov}(wR_A,R_A)+\operatorname{Cov}((1-w)R_B,R_A) \\ =w\operatorname{Cov}(R_A,R_A)+(1-w)\operatorname{Cov}(R_B,R_A)= \\ =0.5\cdot 0.1^2+0.5\cdot 0.005=0.0075\end{gathered}$$

$\boxed{\text{в}}$ Определим значение w, при котором стандартное отклонение доходности портфеля будет минимальным.

$$w:\operatorname{Var}(R)\to\min$$

Ошибочно было бы считать, что задача минимизации равна задаче максимизации доли бумаги A, обладающей меньшей дисперсией.

Покажем это:

$$\operatorname{Var}(R)=w^2\operatorname{Var}(R_A)+(1-w)^2\operatorname{Var}(R_B)+2w(1-w)\operatorname{Cov}(R_A,R_B)\to\min, w\in[0,1]$$

$$\frac{\partial\operatorname{Var}(R)}{\partial w}=2w\operatorname{Var}(R_A)+2(w-1)\operatorname{Var}(R_B)+(2-4w)\operatorname{Cov}(R_A,R_B)=0$$

Подставим:

$$\frac{\partial \mathrm{Var}(R)}{\partial w} = 2 \cdot 0.1^w + 2 \cdot 0.2^(w-1) + (2-4w) \cdot 0.005 = 0$$

$$0.08w = 0{,}07$$

$$w = 0.875$$

Вторая производная положительна, значит, мы нашли точку минимума.

5.1.2 Задача 2

Пусть $y_1, y_2 ... y_{10}$ - выборка из генеральной совокупности с математическим ожидание μ и дисперсией σ^2.

$\boxed{\text{а}}$ Разница между μ и $\hat{\mu}$, некоторой оценкой μ:

- μ - теоретическая характеристика, присущая всей генеральной совокупности. Мы не наблюдаем ее прямо;
- $\hat{\mu}$ исчисляется по некоторой выборке, поэтому сама $\hat{\mu}$ есть случайная величина, обладающая своим матожиданием и дисперсий.

Исследователя интересует математическое ожидание μ, и он рассматривает три возможных оценки этого параметра, которые можно вычислить на основе имеющихся данных.

1. $\hat{\mu} = \dfrac{y_1 + ... + y_{10}}{10}$
2. $\hat{\mu} = \dfrac{y_1 + ... + y_{10}}{9}$
3. $\hat{\mu} = \dfrac{1}{15}(y_1 + y_5) + \dfrac{2}{15}(y_6 + ... + y_{10})$

$\boxed{\text{б}}$ Проверим их на несмещенность:

1. Первая оценка несмещенная:

$$\mathrm{E}(\hat{\mu}) = \mathrm{E}\left(\frac{y_1 + ... + y_{10}}{10}\right) = \frac{1}{10}\sum \mathrm{E}(y_i) = \frac{10}{10}\mu = \mu$$

2. Вторая оценка смещена:

$$\mathrm{E}(\hat{\mu}) = \mathrm{E}\left(\frac{y_1 + ... + y_{10}}{9}\right) = \frac{1}{9}\sum \mathrm{E}(y_i) = \frac{10}{9}\mu$$

3. Третья оценка несмещенная:

$$\mathrm{E}(\hat{\mu}) = E\left(\frac{1}{15}(y_1 + y_5) + \frac{2}{15}(y_6 + ... + y_{10})\right) = \frac{1}{15}\sum_{i=1}^{5}\mathrm{E}(y_i) + \frac{2}{15}\sum_{i=6}^{10}\mathrm{E}(y_i) = \frac{15}{15}\mu = \mu$$

$\boxed{\text{в}}$ Найдем самую эффективную из оценок, вычислив и сравнив их дисперсии

1. Первая оценка:

$$\mathrm{Var}(\hat{\mu}) = \mathrm{Var}\left(\frac{y_1 + ... + y_{10}}{10}\right) = \frac{1}{100}\sum \mathrm{Var}(y_i) = \frac{\sigma^2}{10}$$

2. Третья оценка:

$$\mathrm{Var}(\hat{\mu}) = \mathrm{Var}\left(\frac{1}{15}(y_1 + y_5) + \frac{2}{15}(y_6 + ... + y_{10})\right) = \frac{1}{225}5\sigma^2 + \frac{4}{225}5\sigma^2 = \frac{25\sigma^2}{225} = \frac{\sigma^2}{9}$$

Первая оценка лучше, таким образом.

5.1.3 Задача 3

Пусть дана модель:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i \tag{5.1}$$

Имеются следующие данные о переменных x и y:

y	1	3	1	3	7
x	2	2	6	4	6

$\boxed{\text{а}}$ Найдем МНК-оценки модели $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x_i$.

Воспользуемся формулами:

$$\hat{\beta}_2 = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2}$$

$$\hat{\beta}_1 = \overline{y} - \hat{\beta}_2 \overline{x}$$

Вычислим все необходимые промежуточные значения: $\overline{xy} = 13.6, \overline{x^2} = 19.6, \overline{x} = 4, \overline{y} = 3$.

Подставляя, получи, что $\hat{\beta}_1 = 1, \hat{\beta}_2 = 0.5$

$$\hat{y}_i = 1 + 0.5x_i$$

$\boxed{\text{б}}$ Найдем RSS и R^2.
Как мы помним, $RSS = \sum(\hat{y}_i - \overline{y})^2$.
Вычисляя, получим, что RSS=4.

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS} = \frac{1}{6}$$

5.1.4 Задача 4

Утверждение 5.1.1. Пусть дана модель $y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$.

Тогда коэффициент детерминации R^2 равен квадрату коэффициента корреляции между переменными x и y.

Доказательство. Распишем подробно R^2.

$$R^2 = \frac{\widehat{\mathrm{Var}}(\hat{y})}{\widehat{\mathrm{Var}}(y)} = \frac{\widehat{\mathrm{Var}}(\hat{\beta}_1 + \hat{\beta}_2 x)}{\widehat{\mathrm{Var}}(y)} = \hat{\beta}_2^{\,2} \frac{\widehat{\mathrm{Var}}(x)}{\widehat{\mathrm{Var}}(y)}$$

Помним, что $\hat{\beta}_2 = \dfrac{\widehat{\mathrm{Cov}}(x,y)}{\widehat{\mathrm{Var}}(x)}$. Подставляем:

$$\hat{\beta}_2^{\,2} \frac{\widehat{\mathrm{Var}}(x)}{\widehat{\mathrm{Var}}(y)} = \frac{\widehat{\mathrm{Cov}}^2(x,y)}{\widehat{\mathrm{Var}}^2(x)} \cdot \frac{\widehat{\mathrm{Var}}(x)}{\widehat{\mathrm{Var}}(y)} = \frac{\widehat{\mathrm{Cov}}^2(x,y)}{\widehat{\mathrm{Var}}(x)\,\widehat{\mathrm{Var}}(y)} = \widehat{\mathrm{Corr}}^2(x,y)$$

□

5.1.5 Задача 5

Пусть дана модель:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i. \tag{5.2}$$

$\boxed{\text{а}}$ Пусть мы добавим к каждому наблюдение x константу c.
Пусть мы имели модель:

$$\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x_i.$$

Теперь мы получаем другую модель:

$$\begin{cases} \hat{y}_i = \hat{\alpha}_1 + \hat{\alpha}_2 z_i \\ z_i = x_i + c \end{cases}$$

Очевидно, что $\overline{z} = \overline{x} + c$.
Подробно распишем коэффициенты:

$$\hat{\alpha}_2 = \frac{\widehat{\mathrm{Cov}}(y,z)}{\widehat{\mathrm{Var}}(z)} = \frac{\widehat{\mathrm{Cov}}(y,x+c)}{\widehat{\mathrm{Var}}(x+c)} = \frac{\widehat{\mathrm{Cov}}(y,x)}{\widehat{\mathrm{Var}}(x)} = \hat{\beta}_2$$

Коэффициент при x не изменился.

$$\hat{\alpha}_1 = \overline{y} - \hat{\alpha}_2 \overline{z} = \overline{y} - \hat{\beta}_2(\overline{x} + c) = \overline{y} - \hat{\beta}_2 \overline{x} - \hat{\beta}_2 c = \hat{\beta}_1 - \hat{\beta}_2 c$$

Рассмотрим теперь R^2 в этой модели.

$$R^2 = \frac{\widehat{\mathrm{Cov}}^2(y,z)}{\widehat{\mathrm{Var}}(y)\,\widehat{\mathrm{Var}}(z)} = \frac{\widehat{\mathrm{Cov}}^2(y,x+c)}{\widehat{\mathrm{Var}}(y)\,\widehat{\mathrm{Var}}(x+c)} = \frac{\widehat{\mathrm{Cov}}^2(y,x)}{\widehat{\mathrm{Var}}(y)\,\widehat{\mathrm{Var}}(x)} = R^2_{\text{исходный}}$$

Коэффициент детерминации не изменился.

$\boxed{\text{б}}$ Пусть каждое наблюдение x домножится на константу $c \neq 0$.

$$\begin{cases} \hat{y}_i = \hat{\alpha}_1 + \hat{\alpha}_2 z_i \\ z_i = x_i \cdot c \end{cases}$$

Очевидно, что $\overline{z} = \overline{x} \cdot c$.
Подробно распишем коэффициенты:

$$\hat{\alpha}_2 = \frac{\widehat{\mathrm{Cov}}(y,z)}{\widehat{\mathrm{Var}}(z)} = \frac{\widehat{\mathrm{Cov}}(y,cx)}{\widehat{\mathrm{Var}}(cx)} = \frac{c\,\widehat{\mathrm{Cov}}(y,x)}{c^2\,\widehat{\mathrm{Var}}(x)} = \frac{\hat{\beta}_2}{c}$$

Коэффициент при x изменился.

$$\hat{\alpha}_1 = \overline{y} - \hat{\alpha}_1 \overline{z} = \overline{y} - \frac{\hat{\beta}_2}{c}(\overline{x} \cdot c) = \overline{y} - \hat{\beta}_2 \overline{x} = \hat{\beta}_1$$

Рассмотрим теперь R^2 в этой модели.

$$R^2 = \frac{\widehat{\mathrm{Cov}}^2(y,z)}{\widehat{\mathrm{Var}}(y)\,\widehat{\mathrm{Var}}(z)} = \frac{\widehat{\mathrm{Cov}}^2(y,x \cdot c)}{\widehat{\mathrm{Var}}(y)\,\widehat{\mathrm{Var}}(x \cdot c)} = \frac{c^2\,\widehat{\mathrm{Cov}}^2(y,x)}{c^2\,\widehat{\mathrm{Var}}(y)\,\widehat{\mathrm{Var}}(x)} = R^2_{\text{исходный}}$$

Коэффициент детерминации не изменился.

5.1.6 Задача 6

В результате МНК-оценивания модели по 150 наблюдениях мы получили следующий реультат:

$$\hat{y}_i = 10.4 + 20 \cdot x_i, R^2 = 0.8$$

Пусть теперь мы МНК-оцениванием параметры модели:

$$\hat{x}_i = \hat{\alpha}_1 + \hat{\alpha}_2 \cdot y_i$$

Найдем сперва выражение параметров:

$$\hat{\beta}_2 = \frac{\widehat{\mathrm{Cov}}(x,y)}{\widehat{\mathrm{Var}}(x)}$$

$$\hat{\alpha}_2 = \frac{\widehat{\mathrm{Cov}}(y,x)}{\widehat{\mathrm{Var}}(y)}$$

$$\hat{\alpha}_1 = \overline{x} - \hat{\alpha}_2 \overline{y}$$

Обратим внимание на один замечательный факт:

$$\hat{\beta}_2 \cdot \hat{\alpha}_2 = \frac{\widehat{\mathrm{Cov}}^2(y,x)}{\widehat{\mathrm{Var}}(y)\,\widehat{\mathrm{Var}}(x)} = R^2$$

Таким образом, мы видим совпадение R^2.
Итак, найдем отсюда коэффициенты:

$$20 \cdot \hat{\alpha}_2 = 0.8.$$

$$\hat{\alpha}_2 = \frac{0.8}{20} = 0.04.$$

$$\hat{\beta}_1 = \overline{y} - \hat{\beta}_2 \overline{x}$$

$$10.4 = \overline{y} - 20\overline{x}$$

$$\overline{y} = 10.4 + 20\overline{x}$$

Подставляем и получаем оценку для α_1:

$$\hat{\alpha}_1 = \overline{x} - 0.04(10.4 + 20\overline{x})$$

5.1.7 Задача 7

Имеем модель, для которой мы делаем МНК-оценки:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i \tag{5.3}$$

Для таких оценок верны пять утверждений:

Утверждение 5.1.2.

$$\sum e_i = 0$$

Доказательство. Для доказательства вернемся к условиям первого порядка:

$$\begin{cases} \sum_{i=1}^n 2(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0 \\ \sum_{i=1}^n 2x_i(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0 \end{cases}$$

В первом уравнении записано:

$$\sum_{i=1}^{n} 2(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = \sum_{i=1}^{n} 2(y_i - \hat{y}_i) = \sum_{i=1}^{n} 2e_i = 0$$

□

Утверждение 5.1.3.

$$\sum x_i e_i = 0$$

Доказательство. Для доказательства вернемся к условиям первого порядка:

$$\begin{cases} \sum_{i=1}^n 2(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0 \\ \sum_{i=1}^n 2x_i(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0 \end{cases}$$

Второе уравнение дает нам:

$$\sum_{i=1}^n 2x_i(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 2\sum x_i e_i = 0$$

Следовательно, $\sum x_i e_i = 0$

□

Утверждение 5.1.4. $\sum y_i = \sum \hat{y}_i$

Доказательство. Так как мы знаем, что $\sum e_i = 0$, мы можем записать, что $\sum (y_i - \hat{y}_i) = 0$, откуда следует, что $\sum y_i = \sum \hat{y}_i$. □

Утверждение 5.1.5. $\sum (\hat{y}_i - \overline{y}) e_i = \mathrm{Cov}(\hat{y}, e) = 0$

Доказательство. Будем раскрывать скобки:

$\sum (\hat{y}_i - \overline{y}) e_i = \sum \hat{y}_i \cdot e_i - \overline{y} \sum e_i = \sum (\hat{\beta}_1 + \hat{\beta}_2 x_i) e_i =$

$= \hat{\beta}_1 \sum e_i + \hat{\beta}_2 \sum x_i e_i = 0$

Таким образом, видим, что прогнозные значения игрека некоррелированы с ошибками. □

Утверждение 5.1.6. $\sum (y_i - \overline{y})^2 = \sum (\hat{y}_i - \overline{y})^2 + \sum e_i^2$ или же, другими словами, $TSS = RSS + ESS$

Доказательство. Отметим факт о дисперсиях:

$\widehat{\mathrm{Var}}(y) = \widehat{\mathrm{Var}}(e + \hat{y}) = \widehat{\mathrm{Var}}(e) + \widehat{\mathrm{Var}}(\hat{y}) + 2\,\widehat{\mathrm{Cov}}(e, \hat{y}) =$

$\widehat{\mathrm{Var}}(e) + \widehat{\mathrm{Var}}(\hat{y})$

Запишем, расшифровав:

$$\frac{1}{n} \sum (y_i - \overline{y})^2 = \frac{1}{n} \sum e_i^2 + \frac{1}{n} \sum (\hat{y} - \overline{y})^2$$

$$\sum (y_i - \overline{y})^2 = \sum e_i^2 + \sum (\hat{y} - \overline{y})^2$$

$$TSS = ESS + RSS$$

□

5.1.8 Задача 8

Пусть переменная y связана с переменной x точной линейной связью $y = b_1 + b_2 x$.

Пусть у нас есть выборки наблюдений для x, y и z.

Утверждение 5.1.7. Если $b_2 > 0$, то $\widehat{\text{Corr}}(y{,}z) = \widehat{\text{Corr}}(x{,}z.)$

Доказательство. Как мы помним,

$$\text{Corr}(x{,}y) = \frac{\text{Cov}(x{,}y)}{\sqrt{\text{Var}(x)\,\text{Var}(y)}}$$

Значит, нам нужно доказать, что

$$\frac{\widehat{\text{Cov}}(z{,}y)}{\sqrt{\widehat{\text{Var}}(z)\,\widehat{\text{Var}}(y)}} = \frac{\widehat{\text{Cov}}(x{,}z)}{\sqrt{\widehat{\text{Var}}(x)\,\widehat{\text{Var}}(z)}}.$$

Избавимся сразу от дисперсии z в знаменателях.
Рассмотрим дисперсию y:

$$\widehat{\text{Var}}(y) = \widehat{\text{Var}}(b_1 + b_2 X) = b_2^2\,\widehat{\text{Var}}(X)$$

Добавим в наше выражение:

$$\widehat{\text{Cov}}(z{,}y) = |b_2|\,\widehat{\text{Cov}}(x{,}z)$$

Преобразуем:

$$\widehat{\text{Cov}}(z{,}y) = \widehat{\text{Cov}}(z{,}b_1 + b_2 X) = \widehat{\text{Cov}}(z{,}b_2 X) = b_2\,\widehat{\text{Cov}}(z, x)$$

$$b_2\,\widehat{\text{Cov}}(z, x) = |b_2|\,\widehat{\text{Cov}}(x{,}z)$$

Отсюда мы получаем наше утверждение, если $b_2 > 0$.

□

5.1.9 Задача 9

Утверждение 5.1.8. $\widehat{\mathrm{Corr}}(x,y)$ не меняется при линейной перемене единиц измерения одной из переменных.

Доказательство. Как мы помним,

$$\mathrm{Corr}(x,y) = \frac{\mathrm{Cov}(x,y)}{\sqrt{\mathrm{Var}(x)\,\mathrm{Var}(y)}}.$$

Изменим единицы измерения величины у. Пусть $z = ay + b$ – монотонная функция перевода.

Тогда $\widehat{\mathrm{Var}}(z) = \widehat{\mathrm{Var}}(ay + b) = a^2\,\widehat{\mathrm{Var}}(y)$.

$$\widehat{\mathrm{Cov}}(x,z) = \widehat{\mathrm{Cov}}(x,ay + b) = a\,\widehat{\mathrm{Cov}}(x,y)$$

$$\widehat{\mathrm{Corr}}(x,z) = \frac{a\,\mathrm{Cov}(x,y)}{|a|\sqrt{\mathrm{Var}(x)\,\mathrm{Var}(y)}}.$$

Отсюда видим, что коэффициенты корреляции равны с точностью до знака. □

5.2 Семинар 2-3. Парная регрессия

5.2.1 Задача 1

Исследователь анализирует зависимость потребления риса от уровня дохода (кривую Энгеля) для однородной группы из 20 потребителей. Все потребители из этой группы сталкиваются с одинаковыми ценами на рис и другие товары, и только уровни дохода у них различны, поэтому исследователь использует модель парной регрессии.

Пусть x_i – ежемесячный располагаемый доход i-го потребителя (в тысячах денежных единиц), а y_i – ежемесячное потребление риса i-м потребителем (в килограммах).

Имеются наблюдения:

$$\sum_{i=1}^{20} x_i = 20, \sum_{i=1}^{20} x_i^2 = 40, \sum_{i=1}^{20} y_i = 42, \sum_{i=1}^{20} y_i^2 = 108, \sum_{i=1}^{20} x_i y_i = 60$$

$\boxed{\text{a}}$ Найдем МНК-оценки модели $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x_i$.
Воспользуемся формулами:

$$\hat{\beta}_2 = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2}$$

$$\hat{\beta}_1 = \overline{y} - \hat{\beta}_2 \overline{x}$$

Подставляем значения:

$$\hat{\beta}_2 = \frac{3 - 1 \cdot 2.1}{2 - 1} = 0.9$$

$$\hat{\beta}_1 = 2.1 - 0.1 \cdot 1 = 1.2$$

Получаем модель:

$$\hat{y}_i = 1.2 + 0.9 x_i$$

Вычислим теперь R^2.

$$R^2 = \frac{RSS}{TSS} = \frac{\sum(\hat{y}_i - \overline{y})^2}{\sum(y_i - \overline{y})^2}$$

$$\sum(\hat{y}_i - \overline{y})^2 = \sum(1.2 + 0.9x_i - 2.1)^2 = \sum(0.9x_i - 0.9)^2 = 0.9^2 \sum(x_i - 1)^2 = 16.2$$

$$\sum(y_i - \overline{y})^2 = \sum(y_i - 2.1)^2 = 19.8$$

$$R^2 = \frac{16.2}{19.8} = 0.82$$

$\boxed{\text{б}}$ Проверим значимость переменной x на уровне значимости 5%.
Формулируем гипотезы:

$H_0 : \beta_2 = 0$
$H_1 : \beta_2 \neq 0$

Сравниваем $t_{\text{расч}}$ и $t_{\text{кр}} \sim t(n-2, \alpha)$.

$$t_{\text{расч}} = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)}$$

$$se(\hat{\beta}_2) = \sqrt{\frac{S^2}{\sum(x_i - \overline{x})^2}}$$

$$S^2 = \frac{\sum e_i^2}{n-2}$$

Сначала посчитаем ESS:

$$ESS = TSS - RSS = 19.8 - 16.2 = 3.6$$

$$S^2 = \frac{3.6}{18} = 0.2$$

Теперь вычислим стандартную ошибку:

$$\sum(x_i - \overline{x})^2 = n\,\widehat{\mathrm{Var}}(x) = 20(\overline{x^2} - \overline{x}^2) = 20$$

$$se(\hat{\beta}_2) = 0.1$$

$$t_{\text{расч}} = 9,\ \text{ а } t_{\text{кр}}(18, 0.005) = 2{,}1$$

Значит, с вероятностью 95% $\hat{\beta}_2 \neq 0$, и x оказывает значимое влияние на y. Смысл в том, что росте располагаемого дохода на 1000 единиц потребитель начнет потреблять на 0.9 кг риса больше.

$\boxed{\text{в}}$ Вычислим эластичность спроса на рис по доходу:

$$E_{\frac{y}{x}} = \frac{\partial \hat{y}}{\partial x} \cdot \frac{x}{\hat{y}} = 0.9\frac{x_i}{1.2 + 0.9x_i} = \frac{0.9}{1.2 + 0.9x_i}$$

Так как в любой момент времени $x_i > 0$, это выражение всегда лежит между нулем и единицей. Таким образом, рис - товар первой необходимости.

$\boxed{\text{г}}$ Проверим гипотезу о равенства $\beta_2 = 1$ на уровне значимости 5%.
Формулируем гипотезы:

$H_0 : \beta_2 = 1$
$H_1 : \beta_2 \neq 1$

Сравниваем $t_{\text{расч}}$ и $t_{\text{кр}} \sim t(n-2, \alpha)$.

$$t_{\text{расч}} = \frac{\hat{\beta}_2 - 1}{se(\hat{\beta}_2)} = -1$$

$$t_{\text{кр}}(18, 0.05) = 2{,}1$$

Таким образом, $|t_{\text{расч}}| < t_{\text{кр}}$. Не отклоняем нулевую гипотезу и отклоняем H_1

$\boxed{\text{д}}$ Построим 95-процентный доверительный интервал для коэффициента β_2.

Общий вид интервала:

$$\beta_2 \in (\hat{\beta}_2 - se(\hat{\beta}_2)t_{\text{кр}}(n-2{,}0.05); \hat{\beta}_2 + se(\hat{\beta}_2)t_{\text{кр}}(n-2{,}0.05))$$

Подставляем:

$$\beta_2 \in (0.69, 1.11)$$

5.2.2 Задача 4

Рассматривается альтернативная версии регрессии:

$$y_i = \theta x_i + \varepsilon_i \tag{5.4}$$

Необходимо вычислить МНК-оценку коэффициента и показать ее несмещенность.

Утверждение 5.2.1. МНК-оценка θ $\quad \hat{\theta} = \dfrac{\overline{xy}}{\overline{x^2}}$

Доказательство. Выпишем задачу и условие первого порядка.

Подбираем параметры так, чтобы сумма квадратов отклонений была минимальна:

$$e_1^2 + ... + e_n^2 = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{\theta} x_i)^2 \to \min$$

$$\frac{\partial S}{\partial \hat{\theta}} = 0$$

Все весьма хорошо.

$$\frac{\partial S}{\partial \hat{\theta}} = \sum_{i=1}^{n} -2x_i(y_i - \hat{\theta} x_i) = 0$$

$$\sum x_i y_i - \hat{\theta} \sum x_i^2 = 0$$

$$\hat{\theta} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\overline{xy}}{\overline{x^2}}$$

□

Утверждение 5.2.2. $\hat{\theta}$ – несмещенная оценка параметра θ.

Доказательство. Преобразуем:

$$\hat{\theta} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i(\theta x_i + \varepsilon_i)}{\sum x_i^2} = \frac{\theta \sum x_i^2 + \sum x_i \varepsilon_i}{\sum x_i^2} =$$

$$= \theta + \frac{\sum x_i \varepsilon_i}{\sum x_i^2}.$$

Проверяем несмещенность:

$$\mathrm{E}(\hat{\theta}) = \mathrm{E}\left(\theta + \frac{\sum x_i \varepsilon_i}{\sum x_i^2}\right) = \theta + \mathrm{E}\left(\frac{\sum x_i \varepsilon_i}{\sum x_i^2}\right) = \theta + 0 = \theta$$

□

Утверждение 5.2.3. $\operatorname{Var}(\hat{\theta}) = \dfrac{\sigma^2}{\sum x_i^2}$.

Доказательство.

$$\operatorname{Var}\left(\theta + \frac{\sum x_i \varepsilon_i}{\sum x_i^2}\right) = \operatorname{Var}\left(\frac{\sum x_i \varepsilon_i}{\sum x_i^2}\right) =$$

$$= \frac{\operatorname{Var}(\sum x_i \varepsilon_i)}{(\sum x_i^2)^2} =$$

Используем предпосылки модели о независимости и дисперсии ошибок:

$$= \frac{\sum x_i^2 \operatorname{Var}(\varepsilon_i)}{(\sum x_i^2)^2} = \frac{\sigma^2}{\sum x_i^2}.$$

□

Вычислим R^2:

$$R^2 = \frac{\widehat{\operatorname{Var}}(\hat{y})}{\widehat{\operatorname{Var}}(y)} = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS} = \frac{\frac{1}{n}\sum(\hat{y} - \overline{y})^2}{\frac{1}{n}\sum(y_i - \overline{y})^2} = 1 - \frac{\sum e_i^2}{\sum(y_i - \overline{y})^2}$$

Легко предъявить пример данных, для которых R^2 принимает необычные значения.

Пусть $x_1 = 0, x_2 = 2; y_1 = 1, y_2 = 0$. Оцениваем модель:

$$\hat{\theta} = \frac{0}{0.5} = 0$$

$$\hat{y}_i = 0$$

Тогда, вычисляя R^2:

$$R^2 = 1 - \frac{\sum e_i^2}{\sum(y_i - \overline{y})^2} = 1 - \frac{1}{0.5} = -1$$

Не слишком осмысленно выглядит.

5.2.3 Задача 5

Рассматривается альтернативная версии регрессии:

$$y_i = \theta + \varepsilon_i \tag{5.5}$$

Необходимо вычислить МНК-оценку коэффициента и показать ее несмещенность.

Утверждение 5.2.4. МНК-оценка $\theta \quad \hat{\theta} = \overline{y}$

Доказательство. Выпишем задачу и условие первого порядка.

Подбираем параметры так, чтобы сумма квадратов отклонений была минимальна:

$$e_1^2 + ... + e_n^2 = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{\theta})^2 \to \min$$

$$\frac{\partial S}{\partial \hat{\theta}} = 0$$

Очевидно, что

$$\frac{\partial S}{\partial \hat{\theta}} = \sum_{i=1}^n -2(y_i - \hat{\theta}) = 0$$

$$-2\sum y_i = n\hat{\theta} = 0$$

$$\hat{\theta} = \overline{y}$$

□

Утверждение 5.2.5. $\hat{\theta}$ – несмещенная оценка параметра θ.

Доказательство. Проверяем несмещенность:

$$\mathrm{E}(\hat{\theta}) = \mathrm{E}(\overline{y}) = \frac{1}{n}\sum \mathrm{E}(y) = \frac{1}{n}\,\mathrm{E}(n\theta + n\varepsilon_i) = \theta + \mathrm{E}(\varepsilon_i) = \theta$$

□

Утверждение 5.2.6. $\mathrm{Var}(\hat{\theta}) = \dfrac{\sigma^2}{n}$.

Доказательство.

$$\mathrm{Var}(\hat{\theta}) = \mathrm{E}(\overline{y}) = \frac{1}{n^2}\,\mathrm{Var}\left(\sum y_i\right) = \frac{1}{n^2}\,\mathrm{Var}\left(n\theta + \sum \varepsilon_i\right) =$$

$$= \frac{1}{n^2}\,\mathrm{Var}\left(\sum \varepsilon_i\right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

□

Вычислим дополнительно R^2:

$$R^2 = \frac{\widehat{\mathrm{Var}}(\hat{y})}{\widehat{\mathrm{Var}}(y)} = \frac{0}{\widehat{\mathrm{Var}}(y)} = 0$$

Всегда ноль.

5.2.4 Задача 7

Рассматриваем модель $y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$.

Имеются наблюдения:

$$\sum_{i=1}^{10} x_i = 20, \sum_{i=1}^{10} x_i^2 = 50, \sum_{i=1}^{10} y_i = 8, \sum_{i=1}^{10} y_i^2 = 26, \sum_{i=1}^{10} x_i y_i = 10$$

$\boxed{\text{а}}$ Вычислим МНК-оценци коэффициентов в регрессии:
Воспользуемся формулами:

$$\hat{\beta}_2 = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2}$$

$$\hat{\beta}_1 = \overline{y} - \hat{\beta}_2 \overline{x}$$

Подставляем значения:

$$\hat{\beta}_2 = \frac{1 - 2 \cdot 0.8}{5 - 4} = -0.6$$

$$\hat{\beta}_1 = 0.8 + 0.6 \cdot 2 = 2$$

Получаем модель:

$$\hat{y}_i = 2 - 0.6 x_i$$

$\boxed{\text{б}}$ Имеем наблюдение $x_{11} = 5$. Построим лучший линейный несмещенный прогноз и оценим его точность 95% доверительным интервалом.

Согласно утверждению 1.29.2, лучше всего подставить в уравнение регрессии.

$$y_{11} = 2 - 0.6 \cdot 5 = -1$$

Построим теперь 95-процентный доверительный интервал для этого значения.

Запишем общий интервала:

$$\begin{cases} y_{11} \in (\hat{y}_{11} - t_{\text{кр}} S; \hat{y}_{11} + t_{\text{кр}} S) \\ S = \sqrt{S^2 \left(1 + \frac{1}{n} + \frac{(x_{n+1} - \overline{x})^2}{\sum (x_i - \overline{x})^2}\right)} \\ S^2 = \frac{ESS}{n-2} \end{cases}$$

Мы помним, что

$$\sum e_i^2 = \sum e_i e_i = \sum e_i (y_i - \hat{y}_i) =$$

$$= \sum e_i y_i - \sum e_i (2 - 0{,}6 x_i) = \sum e_i y_i - 2 \sum e_i + 0{,}6 \sum e_i x_i =$$

$$= \sum e_i y_i = \sum (y_i - 2 + 0{,}6 x_i) y_i = \sum y_i^2 - 2 \sum y_i + 0{,}6 \sum x_i y_i =$$

$$= 26 - 2 \cdot 8 + 0.6 \cdot 10 = 16$$

Запишем теперь:

$$\sum (x_i - \overline{x})^2 = n \widehat{\operatorname{Var}}(x) = n(\overline{x^2} - \overline{x}^2) = 10(5 - 4) = 10$$

$$S = \sqrt{2 \left(1 + \frac{1}{10} + \frac{15 \cdot 25}{10}\right) = 2}$$

$$t_{\text{кр}}(8{,}0.05) = 2.306$$

Значит,

$$y_{11} \in (-5.612; 3.612)$$

5.2.5 Задача 8

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i \tag{5.6}$$

Имеются следующие данные о переменных x и y:

y	16	9	7	5	3
x	12	9	6	3	0

$\boxed{\text{а}}$ Найдем МНК-оценки модели $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x_i$.
Воспользуемся формулами:

$$\hat{\beta}_2 = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2}$$

$$\hat{\beta}_1 = \overline{y} - \hat{\beta}_2 \overline{x}$$

Подставляем значения:

$$\hat{\beta}_2 = \frac{66 - 8 \cdot 6}{84 - 64} = 0.9$$

$$\hat{\beta}_1 = 6 - 0.9 \cdot 8 = -1.2$$

Получаем модель:

$$\hat{y}_i = -1.2 + 0.9 x_i$$

$\boxed{\text{б}}$ Найдем RSS и R^2.

$$R^2 = \frac{RSS}{TSS} = \frac{\sum(\hat{y}_i - \overline{y})^2}{\sum(y_i - \overline{y})^2}$$

$$\sum(\hat{y}_i - \overline{y})^2 = \sum(-1.2 + 0.9x_i - 6)^2 = \sum(0.9x_i - 7.2)^2 = 0.9^2 \sum(x_i - 8)^2 = 81$$

$$\sum(y_i - \overline{y})^2 = \sum(y_i - 6)^2 = 90$$

$$R^2 = \frac{81}{90} = 0.9$$

$\boxed{\text{в}}$ Построим 95% доверительный интервла для коэффициента β_2.
Общий вид интервала:

$$\beta_2 \in (\hat{\beta}_2 - se(\hat{\beta}_2)t_{\text{кр}}(n-2{,}0.05); \hat{\beta}_2 + se(\hat{\beta}_2)t_{\text{кр}}(n-2{,}0.05))$$

$$se(\hat{\beta}_2) = \sqrt{\frac{S^2}{\sum(x_i - \overline{x})^2}}$$

$$S^2 = \frac{\sum e_i^2}{n-2}$$

Сначала посчитаем ESS:

$$ESS = TSS - RSS = 90 - 81 = 9$$

$$S^2 = \frac{9}{3} = 3$$

Теперь вычислим стандартную ошибку:

$$\sum(x_i - \overline{x})^2 = n\,\widehat{\mathrm{Var}}(x) = 5(\overline{x^2} - \overline{x}^2) = 100$$

$$se(\hat{\beta}_2) = \sqrt{\frac{3}{100}} = 0.173$$

$$\text{а } t_{\text{кр}}(3, 0.005) = 3.182$$

Подставляем:

$$\beta_2 \in (0.35, 1.45)$$

5.3 Семинар 4. Множественная регрессия

5.3.1 Задача 1

На основе 20 наблюдений была оценена следующая модель регрессии (в скобках указаны стандартные ошибки оценок коэффициентов):

$$\widehat{\mathrm{y_i}} = \underset{(0.6)}{2.4} + \underset{(0.3)}{6.9}\,\mathrm{x}_i + \underset{(9.8)}{5.1}\,\mathrm{w}_i \tag{5.7}$$

Кроме того, известно, что TSS=2000, а сумма квадратов остатков равна 200.

$\boxed{\text{а}}$ Вычислим R^2, R^2_{adj} и SEE.

$$R^2 = 1 - \frac{200}{2000} = 0.9$$

$$R^2_{adj} = R^2 - \frac{k-1}{n-k}(1-R^2) = 0.9 - \frac{3-1}{20-3} \cdot 0.1 = 0.89$$

$$S^2 = \frac{\sum e_i^2}{n-k} = \frac{200}{17} = 11.765$$

$$SEE = \sqrt{11.765} = 3.43$$

$\boxed{\text{б}}$ Проверим все уравнение на значимость: формулируем гипотезы и строим F-статистику.

$H_0 : \boldsymbol{\beta} = \overline{0}$
$H_1 : \boldsymbol{\beta} \neq \overline{0}$

$$F_{\text{расч}} = \frac{R^2}{1-R^2}\frac{n-k}{k-1}$$

$$F_{\text{кр}}(0.01; 2; 17) = 6{,}1$$

$$F_{\text{расч}} = \frac{0.9}{0.1}\frac{20-3}{3-1} = 76.5$$

Очевидно, что

$$F_{\text{расч}} > F_{\text{кр}}$$

. Поэтому мы отклоняем нулевую гипотезу: с уверенностью в 99% мы утверждаем, что уравнение значимо в целом.

$\boxed{\text{в}}$ Проверим коэффициент при переменной на значимость.
Строим гипотезы и t-статистику:

$H_0 : \beta_2 = 0$
$H_1 : \beta_2 \neq 0$

$$t_{\text{расч}} = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} = \frac{6.9}{0.3} = 23$$

$$t_{\text{кр}}(0.01; 17) = 2.898$$

Мы отклоняем нулевую гипотезу: с вероятностью 99% x оказывает значимое влияние на y.

$\boxed{\text{г}}$ Проверим равенство коэффициента при x семи.
Строим гипотезы и t-статистику:

$H_0 : \beta_2 = 7$
$H_1 : \beta_2 \neq 7$

$$t_{\text{расч}} = \frac{\hat{\beta}_2 - 7}{se(\hat{\beta}_2)} = \frac{6.9 - 7}{0{,}3} = -0.333$$

$$t_{\text{кр}}(0.01; 17) = 2.898$$

Мы принимаем нулевую гипотезу, так как расчетное значение статистики меньше критического по модулю: с вероятностью 99% истинное значение коэффициента перед переменной x равно семи.

$\boxed{\text{д}}$ Построим 99-процентный доверительный интервал для коэффициента при переменной x.

Общий вид интервала:

$$\beta_2 \in (\hat{\beta}_2 - se(\hat{\beta}_2)t_{\text{кр}}(n-2{,}0.01); \hat{\beta}_2 + se(\hat{\beta}_2)t_{\text{кр}}(n-2{,}0.01))$$

Подставляя известные нам значения, получим, что с вероятностью в 99% истинное значение коэффициента перед переменной x лежат в таких пределах:

$$\beta_2 \in (6.03; 7.77)$$

$\boxed{\text{e}}$ После того, как уважаемый исследователь добавил в модель еще две переменных (p и s), R^2 в этой модели увеличился до 0.95. Проверим, стоит ли добавлять эти переменные.

Выпишем эти модели:

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 w_i + \beta_4 p_i + \beta_5 s_i + \varepsilon_i$$

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 w_i + \varepsilon_i$$

Сделаем тест на короткую и длинную регрессию.

Формулируем гипотезы:

$H_0 : \beta_4 = \beta_5 = 0$
$H_1 : \exists j \in (4{,}5) : \beta_j \neq 0$

Таким образом, мы видим, что q=2, так как мы проверяем два коэффициента, а k=5 (число регрессоров длинной регрессии). При этом нам известно, что $R^2_R = 0.9$, а $R^2_{UR} = 0.95$.

Вводим F-статистику:

$$F_{\text{кр}} = \frac{(R^2_{UR} - R^2_R)/q}{(1 - R^2_{UR})/(n-k)}$$

Подставляем наши значения:

$$F_{\text{расч}} = \frac{(0.95 - 0.9)/2}{(1 - 0.95)/(20 - 5)} = 7.5$$

$$F_{\text{кр}}(\alpha{,}q{,}n-k) = F_{\text{кр}}(0.01; 2; 15) = 6.36$$

Таким образом, расчетное значение статистики больше критического. Значит, мы отклоняем нулевую гипотезу: с вероятностью 99% хотя бы одна из новых переменных значимо влияет на зависимую переменную.

5.4 Семинар 5. Множественная регрессия

5.4.1 Задача 1

МНК-оценивалась модель $y_i = \beta_1 + \beta_2 \cdot x_i^{(2)} + \beta_3 \cdot x_i^{(3)} + \varepsilon_i$ по 30 наблюдениям.

$$\widehat{\mathrm{y}_\mathrm{i}} = \underset{(3.1)}{9.1} - \underset{(2.1)}{10.2}\,\mathrm{x}_i^{(2)} + \underset{(3.0)}{4.5}\,\mathrm{x}_i^{(3)}, R^2 = 0.21 \tag{5.8}$$

После этого мы МНК-оценили модель $y_i = \beta_1 + \beta_2 \cdot x_i^{(2)} + \beta_3 \cdot x_i^{(3)} + \beta_4 \cdot x_i^{(4)} + \beta_5 \cdot x_i^{(5)} + \varepsilon_i$ по этим же 30 наблюдениям.

При этом мы получили:

$$\hat{y}_i = 8.2 - 8.5x_i^{(2)} + 9.0x_i^{(3)} + 5.0x_i^{(4)} + 6.1x_i^{(5)}, R^2 = 0.23 \tag{5.9}$$

$\boxed{\text{а}}$ Проверим значимость первого уравнения на уровне значимости 5%.
Формулируем гипотезы и строим F-статистику.

$H_0 : \boldsymbol{\beta} = \overline{0}$
$H_1 : \boldsymbol{\beta} \neq \overline{0}$

$$F_{\text{расч}} = \frac{R^2}{1-R^2}\frac{n-k}{k-1}$$

$$F_{\text{кр}}(0.05; 2; 27) = 3.35$$

$$F_{\text{расч}} = \frac{0.21}{0.79}\frac{30-3}{3-1} = 3.59$$

$$F_{\text{расч}} > F_{\text{кр}}.$$

Поэтому мы отклоняем нулевую гипотезу: с уверенностью в 95% мы утверждаем, что уравнение значимо в целом.

$\boxed{\text{б}}$ Проверим коэффициент при переменной $x^{(3)}$ на значимость при уровне значимости 5%.

Строим гипотезы и t-статистику:

$H_0 : \beta_3 = 0$

$H_1 : \beta_3 \neq 0$

$$t_{\text{расч}} = \frac{\hat{\beta}_3}{se(\hat{\beta}_3)} = \frac{4.5}{3} = 1.5$$

$$t_{\text{кр}}(0.05; 27) = 2.052$$

Мы принимаем нулевую гипотезу: с вероятностью 95% $x^{(3)}$ не оказывает значимое влияние на y.

$\boxed{\text{в}}$ Сравним модели используя скорректированный коэиффициент детерминации R^2_{adj}:

$$R^2_{adj} = R^2 - \frac{k-1}{n-k}(1 - R^2)$$

Подставляем значения:

$$R^{2;(1)}_{adj} = 0.21 - \frac{2}{27}(0.79) = 0.151$$

$$R^{2;(2)}_{adj} = 0.23 - \frac{4}{27}(0.77) = 0.11$$

Таким образом, мы видим, что первая модель лучше второй, хотя у второй и выше нескорректированный коэффициент.

$\boxed{\text{г}}$ Проверим, стоило ли добавлять новые переменные в нашу модель на уровне значимости 5%. Сделаем тест на короткую и длинную регрессию.

Формулируем гипотезы:

$H_0 : \beta_4 = \beta_5 = 0$
$H_1 : \exists j \in (4{,}5) : \beta_j \neq 0$

Таким образом, мы видим, что q=2, так как мы проверяем два коэффициента, а k=5 (число регрессоров длинной регрессии). При этом нам известно, что $R^2_R = 0.21$, а $R^2_{UR} = 0.23$.

Вводим F-статистику:

$$F_{\text{расч}} = \frac{(R^2_{UR} - R^2_R)/q}{(1 - R^2_{UR})/(n-k)}$$

Подставляем наши значения:

$$F_{\text{расч}} = \frac{(0.23 - 0.21)/2}{(1 - 0.23)/(30 - 5)} = 0.323$$

$$F_{\text{кр}}(\alpha, q, n - k) = F_{\text{кр}}(0.05; 2; 25) = 3.39$$

Таким образом, расчетное значение статистики меньше критического. Значит, мы принимаем нулевую гипотезу: с вероятностью 95% новые переменные не влияют значимо на зависимую переменную.

5.4.2 Задача 2

Исследуется зависимость среднедушевого потребления алкоголя по странам мира от различных факторов:

Имеется модель:

$$ALCO_i = \beta_1 + \beta_2 GDP_i + \beta_3 MUSL_i + \beta_4 BUDD_i + \beta_5 HINDU_i + \varepsilon_i \tag{5.10}$$

$ALCO_i$ – среднедушевое потребление чистого спирта на человека (л), GDP_i – ВВП на душу населения (долларов США), $MUSL_i, BUDD_i, HINDU_i$ – доли населения исповедующего, соответственно, мусульманство, буддизм и индуизм (в % от общей численности населения

В ходе МНК-оценивания модели на основе данных о 180 странах получены сле дующие результаты: сумма квадратов остатков ESS=200, объясненная сумма квадратов RSS=300.

Также для проверки гипотезы о том, что религия не оказывает существенного влияния на потребление алкоголя, были оценены параметры второй модели:

$$ALCO_i = \beta_1 + \beta_2 GDP_i + \varepsilon_i \tag{5.11}$$

Во второй модели, по сравнению с первой, значение объясненной суммы квадратов RSS изменилось на 100.

$\boxed{\text{а}}$ Вычислим R^2 в модели (1).

$$R^2 = \frac{RSS}{TSS} = \frac{300}{500} = 0.6$$

$\boxed{\text{б}}$ Покажем, что во второй модели значение RSS уменьшилось на 100, а не увеличилось.

Действительно, рассмотрим R^2:

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS} = 1 - \frac{\sum(\widehat{ALCO} - \overline{ALCO})^2}{\sum(ALCO_i - \overline{ALCO})^2}$$

Как мы видим, TSS остается без изменений. При этом количество регрессоров упало, следовательно, упала объясняющая сила модели: ESS растет. Таким образом, $RSS \downarrow$ на 100.

Значит, в короткой модели $R^2 = \dfrac{200}{500} = 0.4$

$\boxed{\text{в}}$ Проверим, влияет ли религия на потребление алкоголя на уровне значимости 5%, с помощью теста на короткую и длинную регрессию.

Формулируем гипотезы

$H_0 : \beta_3 = \beta_4 = \beta_5 = 0$
$H_1 : \exists j \in (3{,}4{,}5) : \beta_j \neq 0$

Таким образом, мы видим, что q=3, так как мы проверяем два коэффициента, а k=5 (число регрессоров длинной регрессии). При этом нам известно, что $R^2_R = 0.4$, а $R^2_{UR} = 0.6$.

Вводим F-статистику:

$$F_{\text{кр}} = \frac{(R^2_{UR} - R^2_R)/q}{(1 - R^2_{UR})/(n-k)}$$

Подставляем наши значения:

$$F_{\text{расч}} = \frac{(0.6 - 0.4)/3}{(1 - 0.6)/(180 - 5)} = 29.17$$

$$F_{\text{кр}}(\alpha, q, n-k) = F_{\text{кр}}(0.05; 3; 175) = 2.6$$

Таким образом, расчетное значение статистики больше критического. Значит, мы отклоняем нулевую гипотезу: с вероятностью 95% новые переменные влияют значимо на зависимую переменную, а длинная регрессия лучше.

5.5 Семинар 6-7. Фиктивные переменные, нелинейные модели

5.5.1 Задача 1

По 1000 наблюдений было оценено следующее уравнение регрессии (в скобках указаны стандартные отклонения оценок коэффициентов):

$$\widehat{\mathrm{y}}_t = -\underset{(19.1)}{117.0} + \underset{(4.2)}{20.0} \ln \mathrm{x}_i + \underset{(5.1)}{20.0}\, \mathrm{z}_i \tag{5.12}$$

$\boxed{\text{а}}$ Дадим интерпретацию коэффициента при переменной z:

При увеличении переменной z на одну единицу переменная y увеличится на 20 единиц.

$\boxed{\text{б}}$ Дадим интерпретацию коэффициента при переменной x:

При увеличении переменной x на один процент переменная y увеличится на $\dfrac{20}{100}$ *единиц.*

5.5.2 Задача 2

По 1000 наблюдений было оценено следующее уравнение регрессии (в скобках указаны стандартные отклонения оценок коэффициентов):

$$\ln \widehat{\mathrm{y}_t} = \underset{(2.0)}{10.0} + \underset{(0.01)}{0.07} \ln \mathrm{x}_i + \underset{(0.01)}{0.07}\, \mathrm{z}_i + \underset{(0.1)}{0.9}\, \mathrm{d}_i \tag{5.13}$$

$\boxed{\text{а}}$ Дадим интерпретацию коэффициента при переменной $\ln x$:

При увеличении переменной x на один процент переменная y увеличится на 0.07 процентов.

$\boxed{\text{б}}$ Дадим интерпретацию коэффициента при переменной z:

При увеличении переменной z на единицу переменная y увеличится на $0.07 \cdot 100$ *процентов.*

Здесь мы используем приблизительную формулу ($\hat{\beta}_i \cdot 100\%$), так как коэффициент меньше 0.1. В противном случае мы были бы вынуждены точно считать.

$\boxed{\text{в}}$ Дадим интерпретацию коэффициента при переменной d:

При увеличении переменной d на единицу переменная y увеличится на 146%.

Покажем это:

$$\Delta \ln y = 0.9\Delta d$$

$$\ln \frac{y_1}{y_0} = 0.9$$

$$\frac{y_1}{y_0} = e^{0.9} = 2.46.$$

5.5.3 Задача 3

Исследуется зависимость потребления индивида от его располагаемого дохода:

$$c_i = \beta_1 + \beta_2 \cdot income_i + \varepsilon_i \tag{5.14}$$

$\boxed{\text{а}}$ Предположим, что мы хотим учесть возможный сдвиг в автономном потреблении при переходе от городского населения к сельскому.

Для этого мы введем новую фиктивную переменную $city_i$:

$$city_i = \begin{cases} 1, & \text{если i-ый индивид из города;} \\ 0, & \text{если i-ый индивид из села;} \end{cases}$$

Тогда мы получаем новую модификацию модели:

$$c_i = \beta_1 + \beta_2 \cdot income_i + \beta_3 \cdot city_i + \varepsilon_i \tag{5.15}$$

Значит, при разных значениях переменной мы получаем:

$city_i = 1 : c_i = (\beta_1 + \beta_3) + \beta_2 income_i + \varepsilon_i$
$city_i = 0 : c_i = \beta_1 + \beta_2 income_i + \varepsilon_i$

$\boxed{\text{б}}$ Теперь мы бы хотели отразить возможность того, что предельные склонности к потреблению индивидуумов с доходом выше и ниже уровня y_o различаются.

Вводим новую фиктивную переменную:

$$level_i = \begin{cases} 1, \text{ если } y_i > y_o \\ 0, \text{ если } y_i \leq y_o \end{cases}$$

Тогда получаем новую модель:

$$c_i = \beta_1 + \beta_2 \cdot income_i + \beta_3 \cdot level_i \cdot income_i \varepsilon_i \tag{5.16}$$

Значит, при разных значениях переменной мы получаем:

$y_i > y_o : c_i = \beta_1 + (\beta_2 + \beta_3) income_i + \varepsilon_i$
$y_i \leq y_o : c_i = \beta_1 + \beta_2 income_i + \varepsilon_i$

5.5.4 Задача 4

На основе квартальных данных с 1971 по 1976 г. с помощью МНК получено следующее уравнение (в скобках указаны стандартные отклонения):

$$\widehat{\mathrm{y}_\mathrm{t}} = \underset{(2.14)}{1.12} - \underset{(0.0034)}{0.0098}\, \mathrm{x}_t^{(1)} - \underset{(3.42)}{5.62}\, \mathrm{x}_t^{(2)} + \underset{(0.009)}{0.044}\, \mathrm{x}_t^{(3)} \tag{5.17}$$

$RSS = 110.21$; $ESS = 21.43$ для нашей модели.

$\boxed{\text{a}}$

$$R^2 = \frac{RSS}{TSS} = \frac{110.21}{131.75} = 0.837$$

Проверим коэффициент при всех переменной на значимость на уровне значимости 1%.

Строим гипотезы и t-статистику:

$H_0^k : \beta_k = 0$
$H_1^k : \beta_k \neq 0, k \in (2, 3, 4)$

Для всех случаев $t_{\text{кр}}(0.01; 20) = 2.845$.

- k=2: $t_{\text{расч}} = \dfrac{-0.0098}{0.0034} = -2.88$. $|t_{\text{расч}}| > t_{\text{кр}}$, значит, нулевая гипотеза отклоняется, и коэффициент значим.
- k=3: $t_{\text{расч}} = \dfrac{-5.62}{3.42} = -1.67$. $|t_{\text{расч}}| < t_{\text{кр}}$, значит, нулевая гипотеза принимается, и нет значимого влияния.

- k=4: $t_{\text{расч}} = \dfrac{0.044}{0.009} = 4.89$. $|t_{\text{расч}}| > t_{\text{кр}}$, значит, нулевая гипотеза отклоняется, и коэффициент значим.

$\boxed{\text{б}}$ Были раздельно проведены две регрессии на основе данных: 1-й квартал 1971 г. – 1-й квартал 1975 г. и 2-й квартал 1975 г. – 4-й квартал 1976 г. Соответственно получены следующие значения сумм квадратов остатков: $ESS_1 = 12.25$, $ESS_2 = 2.32$.

Мы подозреваем, что произошел структурный сдвиг. Чтобы проверить это, сделаем тест Чоу.

Разбиваем нашу выборку на две подвыборки: до сдвига/после сдвига (с признаком/без признака).

При этом наши модели имеют один набор регрессоров.

$$1: y = \beta_1 + \sum \beta_k x^{(k)} + \varepsilon \tag{5.18}$$

$$2: y = \alpha_1 + \sum \alpha_k x^{(k)} + \varepsilon \tag{5.19}$$

Формулируем конкурирующие гипотезы:

$H_0 : \forall i \ \alpha_i = \beta_i$;
$H_1 : \exists j \in (1{,}2{,}3) : \alpha_j \neq \beta_j$.

Для такого теста мы вводим F-статистику:

$$F_{\text{расч}} = \frac{(ESS_{\sum} - ESS_A - ESS_B)/k}{(ESS_A + ESS_B)/(n-2k)} \sim F(k, n-2k)$$

Подставляем наши значения:

$$F_{\text{расч}} = \frac{(21.43 - 12.23 - 2.32)/4}{(12.25 + 2.32)/(24 - 4)} = 1.88$$

Сравниваем с $F_{\text{кр}}(\alpha, k, n-2k) = F_{\text{кр}}(0.05; 4; 16) = 3.01$. Расчетное значение меньше критического, значит, мы принимаем нулевую гипотезу: с уверенностью 95% мы можем утверждать, что не произошло структурного сдвига.

5.5.5 Задача 5

Оценивается производственная функция типичной фирмы некоторой отрасли:

$$\ln Y_i = \beta_1 + \beta_2 \ln K_i + \beta_3 \ln L_i + \varepsilon_i \tag{5.20}$$

Выборка состоит из 100 отечественных фирм и еще 100 иностранных фирм. Исследователь предполагает, что у двух этих групп фирм различаются эластичности выпуска по труду, в то время как все остальные параметры производственной функции идентичны.

Чтобы отразить это различие, введем новую фиктивную переменную $foreign_i$:

$$foreign_i = \begin{cases} 1, & \text{если i-ая фирма иностранная;} \\ 0, & \text{если i-ый фирма не иностранная;} \end{cases}$$

Тогда мы получаем модель вида:

$$\ln Y_i = \beta_1 + \beta_2 \ln K_i + \beta_3 \ln L_i + \beta_4 foreign_i \ln L_i + \varepsilon_i$$

$$\ln Y_i = \beta_1 + \beta_2 \ln K_i + (\beta_3 + \beta_4 foreign_i) \ln L_i + \varepsilon_i$$

Значит, при разных значениях переменной мы получаем:

$foreign_i = 1 : \ln Y_i = \beta_1 + \beta_2 \ln K_i + (\beta_3 + \beta_4) \ln L_i + \varepsilon_i$
$foreign_i = 0 : \ln Y_i = \beta_1 + \beta_2 \ln K_i + \beta_3 \ln L_i + \varepsilon_i$

5.5.6 Задание 7

На основе данных о 200 квартирах города Готэм было оценено следующее уравнение регрессии (все переменные оказались значимыми, $R^2 = 0.94$):

$$\ln \hat{P}_i = 1.00 + 0.90 \cdot \ln S_i + 0.20 \cdot Center_i \cdot \ln S_i + 0.03 \cdot Center_i +$$
$$+ 0.04 \cdot Metro_i + 0.05 \cdot Metro_i \cdot Center_i$$

P_i – цена i-ой квартиры, тысяч долларов, S_i – площадь i-ой квартиры, квадратных метров, $Center_i$ – фиктивная переменная равная единице, если i-ая квартира расположена в центре города, и равная нулю в про-тивном случае, $Metro_i$ – фиктивная переменная равная единице, если i-ая квартира расположена в пешей доступности от метро, и равная нулю если от квартиры до метро надо добираться на общественном транспорте.

$\boxed{\text{а}}$ Проверим, на сколько процентов при прочих равных условиях увеличивается цена квартиры в центре города, при увеличении ее площади на 1%.

Просто подставляем в нашу модель наши данные:

$$\ln \hat{P}_i = (1.00 + 0.03) + (0.90 + 0.20) \cdot \ln S_i + (0.04 + 0.05) \cdot Metro_i$$

Видим, что при увеличении площади на 1% цена квартиры растет на 1.1%.

$\boxed{\text{б}}$ Проверим, на сколько процентов дороже квартира в центре рядом с метро по сравнению с такой же квартирой, расположенной не рядом с метро.

Рядом с метро:

$$\ln \hat{P}_i = (1.00 + 0.03 + 0.09) + (0.90 + 0.20) \cdot \ln S_i$$

Не рядом с метро:

$$\ln \hat{P}_i = (1.00 + 0.03) + (0.90 + 0.20) \cdot \ln S_i$$

Значит, получаем, что

$$\frac{P_1}{P_0} = e^{0.09} = 1.094$$

Значит, на 9.4%.

$\boxed{\text{в}}$ Проверим, на сколько процентов дороже квартира не в центре рядом с метро по сравнению с такой же квартирой, расположенной не рядом с метро.

Рядом с метро:

$$\ln \hat{P}_i = (1.00 + 0.04) + 0.90 \cdot \ln S_i$$

Не рядом с метро:

$$\ln \hat{P}_i = 1.00 + 0.90 \cdot \ln S_i$$

Значит, получаем, что

$$\frac{P_1}{P_0} \approx 1.04$$

Значит, на 4%.

5.5.7 Задача 8

На рынке телевизоров некоторого города продаются телевизоры только трех фирм: «Альфа», «Бета» и «Гамма». Исследователь анализирует зависимость цены телевизора от диагонали экрана и марки производителя. В его распоряжении имеется информация о 100 моделях телевизоров. Для каждого наблюдения ему известна цена телевизора в долларах (обозначим ее P), длина диагонали экрана в дюймах (обозначим ее $Diag$) и марка производителя. Исследователь ввел следующие фиктивные переменные:

1. $Alfa_i$ — фиктивная переменная равная единице, если i-ая модель телевизора произведена фирмой «Альфа» и равная нулю во всех остальных случаях;
2. $Beta_i$— фиктивная переменная равная единице, если i-ая модель телевизора произведена фирмой «Бета» и равная нулю во всех остальных случаях.

На основе доступных данных было оценено следующее уравнение регрессии (все переменные оказались значимыми):

$$\ln \hat{P}_i = 2.00 + 0.05 \cdot Diag_i + 0.07 \cdot Alfa_i + 0.06 \cdot Beta_i + 0.03 \cdot Diag_i \cdot Beta_i$$

$\boxed{\text{а}}$ Проверим, на сколько процентов увеличивается цена телевизора фирмы «Бета» при увеличении диагонали его экрана на один дюйм.

Для телевизора "Бета"имеем модель:

$$\ln \hat{P}_i = 2.06 + (0.05 + 0.03) \cdot Diag_i$$

Таким образом, при увеличении диагонали на один дюйм, цена вырастет на 8%.

$\boxed{\text{б}}$ Проверим, на сколько процентов увеличивается цена телевизора фирмы «Гамма» при увеличении диагонали его экрана на один дюйм.

Наша модель для "Гаммы"имеет вид:

$$\ln \hat{P}_i = 2.00 + 0.05 \cdot Diag_i$$

Значит, на 5%.

$\boxed{\text{в}}$ Дадим интерпретацию коэффициенту при переменной $Alfa_i$: *при прочих равных условиях телевизоры фирмы «Альфа» на 7 процентов дороже, чем телевизоры "Гамма".*

5.5.8 Задание 9

Исследователь анализирует взаимосвязь между площадью квартиры x и ее ценой y. В его выборке есть 30 квартир в кирпичных домах и 40 — в панельных. Он оценил регрессию для квартир в кирпичных домах и получил следующий результат:

$$\hat{y}_i = 30 + 50x_i$$

Затем он оценил регрессию для квартир в панельных домах и получил уравнение:

$$\hat{y}_i = 10 + 21x_i$$

.

Теперь он, используя все 70 наблюдений, оценивает регрессию:

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 \cdot x_i + \hat{\beta}_3 \cdot D_i + \hat{\beta}_4 \cdot D_i \cdot X_i$$

Здесь D — переменная, которая равна единице для квартир в кирпичных домах и нулю для квартир в панельных. Мы хотим понять, какие численные значения оценок коэффициентов он получит.

Подставляя разные значения фиктивной переменной, мы можем заключить, опираясь на свойства многочленов, что результаты должны получаться такими же, как и для отдельных регрессии.

Покажем это формально. Формулируем оптимизационную задачу:

$$ESS = \sum e_i^2 = \sum (y_i - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_i - \hat{\beta}_3 \cdot D_i - \hat{\beta}_4 \cdot D_i \cdot X_i)^2 \to \min$$

Выписываем условия первого порядка:

$$\begin{cases} \dfrac{\partial ESS}{\partial \hat{\beta}_1} = -2\sum(y_i - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_i - \hat{\beta}_3 \cdot D_i - \hat{\beta}_4 \cdot D_i \cdot X_i) = 0 \\ \dfrac{\partial ESS}{\partial \hat{\beta}_2} = -2x_i\sum(y_i - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_i - \hat{\beta}_3 \cdot D_i - \hat{\beta}_4 \cdot D_i \cdot X_i) = 0 \\ \dfrac{\partial ESS}{\partial \hat{\beta}_3} = -2D_i\sum(y_i - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_i - \hat{\beta}_3 \cdot D_i - \hat{\beta}_4 \cdot D_i \cdot X_i) = 0 \\ \dfrac{\partial ESS}{\partial \hat{\beta}_4} = -2D_i x_i\sum(y_i - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_i - \hat{\beta}_3 \cdot D_i - \hat{\beta}_4 \cdot D_i \cdot X_i) = 0 \end{cases}$$

Слегка преобразуем систему:

$$\begin{cases} \sum(y_i - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_i - \hat{\beta}_3 \cdot D_i - \hat{\beta}_4 \cdot D_i \cdot X_i) = 0 \\ x_i\sum(y_i - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_i - \hat{\beta}_3 \cdot D_i - \hat{\beta}_4 \cdot D_i \cdot X_i) = 0 \\ D_i\sum(y_i - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_i - \hat{\beta}_3 \cdot D_i - \hat{\beta}_4 \cdot D_i \cdot X_i) = 0 \\ D_i x_i\sum(y_i - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_i - \hat{\beta}_3 \cdot D_i - \hat{\beta}_4 \cdot D_i \cdot X_i) = 0 \end{cases}$$

Разобьем каждую из сумм относительно D. При этом первые слагаемые третьего и четвертого уравнений обнулятся.

$$\begin{cases} \sum\limits_{D=0}(y_i - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_i) + \sum\limits_{D=1}(y_i - (\hat{\beta}_1 + \hat{\beta}_3) - (\hat{\beta}_3 + \hat{\beta}_4) \cdot X_i) = 0 \\ \sum\limits_{D=0}x_i(y_i - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_i) + \sum\limits_{D=1}x_i(y_i - (\hat{\beta}_1 + \hat{\beta}_3) - (\hat{\beta}_3 + \hat{\beta}_4) \cdot X_i) = 0 \\ \sum\limits_{D=1}(y_i - (\hat{\beta}_1 + \hat{\beta}_3) - (\hat{\beta}_3 + \hat{\beta}_4) \cdot X_i) = 0 \\ \sum\limits_{D=1}x_i(y_i - (\hat{\beta}_1 + \hat{\beta}_3) - (\hat{\beta}_3 + \hat{\beta}_4) \cdot X_i) = 0 \end{cases}$$

Вычтем из первого уравнения третье, а из второго – четвертое.
Получаем следующую систему:

$$\begin{cases} \sum\limits_{D=0}(y_i - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_i) = 0 \\ \sum\limits_{D=0}x_i(y_i - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_i) = 0 \\ \sum\limits_{D=1}(y_i - (\hat{\beta}_1 + \hat{\beta}_3) - (\hat{\beta}_3 + \hat{\beta}_4) \cdot X_i) = 0 \\ \sum\limits_{D=1}x_i(y_i - (\hat{\beta}_1 + \hat{\beta}_3) - (\hat{\beta}_3 + \hat{\beta}_4) \cdot X_i) = 0 \end{cases}$$

Первые два уравнения в точности повторяют систему, из которой находятся коэффициенты для регрессии на основе выборки с панельными домами.

$$\begin{cases} \sum\limits_{D=0} (y_i - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_i) = 0 \\ \sum\limits_{D=0} x_i (y_i - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_i) = 0 \end{cases}$$

Эта система совместна и имеет единственное решение, как мы видим из условия.

$$\begin{cases} \hat{\beta}_1 = 10 \\ \hat{\beta}_2 = 21 \end{cases}$$

Отсюда следует, что единственное решение относительно $\hat{\beta}_3, \hat{\beta}_4$ имеет и вторая система:

$$\begin{cases} \sum\limits_{D=1} (y_i - (\hat{\beta}_1 + \hat{\beta}_3) - (\hat{\beta}_3 + \hat{\beta}_4) \cdot X_i) = 0 \\ \sum\limits_{D=1} x_i (y_i - (\hat{\beta}_1 + \hat{\beta}_3) - (\hat{\beta}_3 + \hat{\beta}_4) \cdot X_i) = 0 \end{cases}$$

Мы обнаруживаем, что:

$$\begin{cases} \hat{\beta}_3 + 10 = 30 \\ \hat{\beta}_4 + 21 = 50 \end{cases}$$

Получаем, таким образом:

$$\begin{cases} \hat{\beta}_3 = 20 \\ \hat{\beta}_4 = 29 \end{cases}$$

Регрессия имеет вид:

$$\hat{y}_i = 10 + 21 \cdot x_i + 20 \cdot D_i + 29 \cdot D_i \cdot X_i.$$

Глава 6
Семинары 9-15

6.1 Семинар 9. Векторно-матричная форма записи

6.1.1 Задача 1

Имеются следующие данные о 20 наблюдениях двух регрессоров и зависимой переменной:

$$\sum y_i = 50, \sum x_i^{(2)} = 0, \sum x_i^{(3)} = 10, \sum x_i^{(3)} y_i = 40, \sum (x_i^{(2)})^2 = 10$$

$$\sum (x_i^{(3)})^2 = 10, \sum x_i^{(2)} x_i^{(3)} = 0, \sum x_i^{(2)} y_i = 20$$

Вычислить МНК-оценки коэффициентов в регрессии:

$$y_i = \beta_1 + \beta_2 x_i^{(2)} + \beta_3 x_i^{(3)} + \varepsilon_i$$

Чтобы осознать суть, выпишем векторно-матрично нашу регрессию:

$$y = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_{20} \end{pmatrix} = \beta_1 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_2 \begin{pmatrix} x_1^2 \\ \vdots \\ x_{20}^2 \end{pmatrix} + \beta_3 \begin{pmatrix} x_1^3 \\ \vdots \\ x_{20}^3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{20} \end{pmatrix}$$

Таким образом, матрица X – матрица 20x3.

Считаем вектор оценок:

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1} X' y$$

Матрица X'X, как легко проверить, состоит из сумм произведений вида $\sum x_i x_j$. Таким образом, формируем ее, опираясь на условие:

$$X'X = \begin{pmatrix} 20 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 10 \end{pmatrix}, (X'X)^{-1} = \begin{pmatrix} 0.1 & 0 & -0.1 \\ 0 & 0.1 & 0 \\ -0.1 & 0 & 0.2 \end{pmatrix}$$

Аналогичным образом получаем и правую часть выражения:

$$X'y = \begin{pmatrix} 50 \\ 20 \\ 40 \end{pmatrix}$$

Таким образом, получаем, что:

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

Заметим, что матрица X'X близка к вырождению. Это намек на наличие мультиколлинеарности.

6.1.2 Задача 2

МНК-оценивание модели производственной функции типичной фирмы некоторой отрасли по 27 наблюдения дало следующие результаты:

$$\ln y_i = \beta_1 + \beta_2 \cdot \ln x_i^{(2)} + \beta_3 \ln x_i^{(3)} + \varepsilon_i$$

$$\ln \hat{y}_i = 5 + 0.5 \cdot \ln x_i^{(2)} + 0.4 \ln x_i^{(3)}$$

При этом $ESS = 40$.

Еще известно, что

$$(X'X)^{-1} = 10^{-3} \begin{pmatrix} 4890 & -100 & -25 \\ -100 & 2.25 & 0.25 \\ -25 & 0.25 & 0.5 \end{pmatrix}$$

$\boxed{\text{а}}$ Проверим значимость каждого из коэффициентов. На наше счастье, в рамках предпосылок КЛММР статистика для этого теста все еще подчиняется распределению Стьюдента с $n - k$ степеней свободы.

Формулируем гипотезы:

$$H_0^k : \beta_k = 0$$

$$H_1^k : \beta_k \neq 0$$

Статистика имеет вид $t_{\text{расч}} = \dfrac{\hat{\beta}_k}{se(\hat{\beta}_k)}$. Возникает справедливый вопрос, откуда нам взять стандартные ошибки.

Как оказывается, $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}) = S^2(X'X)^{-1}$, где $S^2 = \dfrac{ESS}{n-k}$. Если подробнее, то ковариационная матрица имеет вид:

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}) = \begin{pmatrix} \widehat{\mathrm{Var}}(\hat{\beta}_1) & \widehat{\mathrm{Cov}}(\hat{\beta}_1,\hat{\beta}_2) & \widehat{\mathrm{Cov}}(\hat{\beta}_1,\hat{\beta}_3) \\ \widehat{\mathrm{Cov}}(\hat{\beta}_1,\hat{\beta}_2) & \widehat{\mathrm{Var}}(\hat{\beta}_2) & \widehat{\mathrm{Cov}}(\hat{\beta}_3,\hat{\beta}_2) \\ \widehat{\mathrm{Cov}}(\hat{\beta}_1,\hat{\beta}_3) & \widehat{\mathrm{Cov}}(\hat{\beta}_3,\hat{\beta}_2) & \widehat{\mathrm{Var}}(\hat{\beta}_3) \end{pmatrix} = \frac{ESS}{n-k}(X'X)^{-1}$$

Легко вычисляем оценку: $S^2 = \dfrac{40}{27-3} = \dfrac{40}{24}$. Начинаем тестировать, $t_{\text{кр}}(24{,}0.05) = 2.064$.

1. $t_{\text{расч}} = \dfrac{5}{2.855} = 1.75$. Значит, константа незначима.
2. $t_{\text{расч}} = \dfrac{0.5}{0.061} = 8.2$. Значит, первый коэффициент значим.
3. $t_{\text{расч}} = \dfrac{0.4}{0.029} = 13.79$. Значит, второй коэффициент значим.

$\boxed{\text{б}}$ Проверим, можно ли утверждать, что в отрасли есть постоянная отдача от масштаба, двумя способами.

Формулируем наши гипотезы:

$$H_0^k : \beta_2 + \beta_3 = 1$$
$$H_1^k : \beta_2 + \beta_3 \neq 1$$

Так как мы имеем всего одно линейное ограничение, тест на линейное ограничение можно проводить двумя способами.

Сначала сделаем через t-тест.

$$t_{\text{расч}} = \frac{\hat{\beta}_2 + \hat{\beta}_3 - 1}{se(\hat{\beta}_2 + \hat{\beta}_3)} = \frac{0.5 + 0.4 - 1}{\sqrt{\mathrm{Var}(\hat{\beta}_2) + \mathrm{Var}(\hat{\beta}_3) + 2\,\mathrm{Cov}(\hat{\beta}_2, \hat{\beta}_3)}} = \frac{-0.1}{0.074} = -1.36$$

Сравниваем с $t_{\text{кр}}(24{,}0.05) = 2.064$. Таким образом, мы принимаем нулевую гипотезу. Можно признать, что есть постоянная отдача от масштаба в отрасли.

Альтернативно можно пойти через общий вариант – сформулировать F-статистику.

Система линейных ограничений на коэффициенты имеет вид $H\beta = r$.

$$\begin{pmatrix} 0 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = 1$$

Тогда

$$F_{\text{расч}} = \frac{(H\hat{\beta} - r)'(H(X'X)^{-1}H')^{-1}(H\hat{\beta} - r)/q}{ESS/(n-k)}$$

q – количество ограничений. Совершая все манипуляции, мы получаем, что $F_{\text{расч}} = 1.849$. $F_{\text{кр}}(\alpha{,}q{,}n-k) = F_{\text{кр}}(0.05, 1, 24) = 4.25$. Значит, мы принимаем нулевую гипотезу.

6.1.3 Задача 3

Пусть истинная связь имеет вид:

$$y_i = \beta_1 + \beta_2 x_i^{(1)} + \beta_3 x_i^{(2)} + \varepsilon_i$$

При этом мы ошибочно не включили вторую переменную в нашу модель и получили оценку:

$$\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x_i.$$

Покажем смещенность коэффициента $\hat{\beta}_2$:

$$\mathrm{E}(\hat{\beta}_2) = \mathrm{E}\left(\frac{\mathrm{Cov}(x^{(1)}, y)}{\mathrm{Var}(x^{(1)})}\right) = \mathrm{E}\left(\frac{\mathrm{Cov}(x^{(1)}, \beta_1 + \beta_2 x^{(1)} + \beta_3 x^{(2)} + \varepsilon)}{\mathrm{Var}(x^{(1)})}\right) =$$

$$= \frac{\beta_2 \,\mathrm{Cov}(x^{(1)}{,}x^{(1)}) + \beta_3 \,\mathrm{Cov}(x^{(1)}{,}x^{(2)})}{\mathrm{Var}(x^{(1)})} = \beta_2 + \frac{\beta_3 \,\mathrm{Cov}(x^{(1)}{,}x^{(2)})}{\mathrm{Var}(x^{(1)})}.$$

$\boxed{\text{а}}$ Пусть $\beta_3 > 0, Cov(x^2{,}x^3) > 0$. Тогда, как мы видим, мы получаем завышенную оценку.

$\boxed{\text{б}}$ Если ровно один из этих значений отрицательно, а другое положительно, мы получаем заниженную оценку.

$\boxed{\text{в}}$ В случае нулевой ковариации между регрессорами мы получаем несмещенную оценку, как можно видеть.

6.1.4 Задача 4

Исследователь МНК-оценивал регрессию на константу и три регрессора по 26 наблюдениям. Результат:

$$\hat{y}_i = 5 + 3.5x_i^{(1)} - 0.7x_i^{(2)} + 2.0x_i^{(3)}, R^2 = 0.882$$

$\boxed{\text{а}}$ Исследователь оценил ту же модель при ограничении $\beta_2 = \beta_4$. Результат:

$$\hat{y}_i = 4.5 + 3.0(x_i^{(1)} + x_i^{(3)}) - 0.9x_i^{(2)}, R^2 = 0.876$$

Проверим гипотезу, что $\beta_2 = \beta_4$.

$$H_0 : \beta_2 - \beta_4 = 0$$
$$H_1 : \beta_2 \neq \beta_4$$

Выписываем F-статистику:

$$F_{\text{расч}} = \frac{R^2_{UR} - R^2_R}{1 - R^2_{UR}} \frac{n-k}{q} = \frac{0.882 - 0.876}{1 - 0.882} \frac{26-4}{1} = 1.12$$

$$F_{\text{кр}}(0.05, 26-4, 1) = 4{,}3$$

Таким образом, мы принимаем нулевую гипотезу и признаем, что коэффициенты равны с вероятностью 95%.

$\boxed{\text{б}}$ Дальше исследователь сделал тест Рамсея для первоначальной модели (с добавлением двух соответствующих переменных). В уравнении, которое он оценил для этого, R^2 оказался равен 0.901.

Проведем и мы этот тест и проверим результаты.

В ходе теста Рамсея он оценивал регрессию:

$$y_i = \beta_1 + \beta_2 x_i^{(1)} + \beta_3 x_i^{(2)} + \beta_4 x_i^{(3)} + \alpha_1 \hat{y}_i^2 + \alpha_2 \hat{y}_i^3 + \varepsilon_i$$

Формулируем сопутствующие гипотезы:

$$H_0 : \alpha_1 = \alpha_2 = 0$$

[1]

$$H_1 : \text{не } H_0$$

Выписываем F-статистику теста на короткую и длинную регрессию.

$$F_{\text{расч}} = \frac{R^2_{UR} - R^2_R}{1 - R^2_{UR}} \frac{n-k}{q} = \frac{0.901 - 0.882}{1 - 0.901} \frac{26-6}{2} = 1.91$$

$$F_{\text{кр}}(0.05, 2, 20) = 3.9$$

Таким образом, мы принимаем нулевую гипотезу: с вероятностью 95% наше уравнение специфицировано верно.

6.1.5 Задача 8

Исследователь МНК-оценивал следующую модель по 5000 наблюдений:

$$y_i = \beta_1 + \beta_2 x_i^{(2)} + \beta_3 x_i^{(3)} + \beta_4 x_i^{(4)} + \varepsilon_i$$

У него получилось, что:

$$\widehat{\mathrm{y}_\mathrm{i}} = 5 + \underset{(1{,}9)}{3{,}5}\,\mathrm{x}_i^{(2)} + \underset{(3{,}0)}{6{,}0}\,\mathrm{x}_i^{(3)} + \underset{(4{,}0)}{9{,}0}\,\mathrm{x}_i^{(4)} \tag{6.1}$$

При этом еще известно, что $\widehat{\mathrm{Cov}}(\hat\beta_3, \hat\beta_4) = 1.0$

Проверим на 5-процентном уровне значимости следующую гипотезу:

$$H_0 : 2\beta_3 - \beta_4 = 4$$
$$H_1 : 2\beta_3 - \beta_4 \neq 4$$

Так как мы имеем всего одно ограничение, это мы можем проверить с помощью t-статистики.

$$t_{\text{расч}} = \frac{(2\hat\beta_3 - \hat\beta_4 - (2\beta_3 - \beta_4))}{se(2\hat\beta_3 - \hat\beta_4)}$$

Вычислим стандартную ошибку.

[1] Эта гипотеза эквивалентна другой: H_0: спецификация уравнения верна

$$se(2\hat{\beta}_3 - \hat{\beta}_4) = \sqrt{\widehat{\mathrm{Var}}(2\hat{\beta}_3 - \hat{\beta}_4)} = \sqrt{4\,\widehat{\mathrm{Var}}(\hat{\beta}_3) + \widehat{\mathrm{Var}}(\hat{\beta}_4) - 4\,\widehat{\mathrm{Cov}}(\hat{\beta}_3,\hat{\beta}_4)} =$$

$$= \sqrt{4 \cdot 9 + 16 - 4} = \sqrt{48}$$

$$se(2\hat{\beta}_3 - \hat{\beta}_4) \approx 6{,}93$$

$$t_{\text{расч}} = \frac{(12-9)-4}{6.93} = \frac{-1}{6.93} = -0.144$$

Так как $t_{\text{кр}}(\infty, 0.05) = 1.96$, мы принимаем нулевую гипотезу.

6.2 Семинар 10-11. Гетероскедастичность

6.2.1 Задача 1

Исследователь при помощи МНК оценил коэффициенты в следующем уравнении регрессии:

$$y_i = \beta_1 + \beta_2 x_i^{(1)} + \beta_3 x_i^{(2)} + \beta_4 x_i^{(3)} + \varepsilon_i$$

Число наблюдений равно 180. После этого он решил провести тест Уайта (с перекрестными эффектами) на гетероскедастичность.

$\boxed{\text{а}}$ Выпишем явно гипотезы и уравнение регрессии, которое ему требуется оценить.

$$H_0 : \sigma_1^2 = ... = \sigma_{180}^2$$
$$H_1 : \text{В модели есть гетероскедастичность}$$

Чтобы провести тест Уайта, нужно построить вектор предсказанных значений $\hat{y}$ и найти остатки регрессии. После этого мы строим регрессию $\boldsymbol{e}$ на все регрессоры, их квадраты и попарные произведения[2].

Регрессия будет иметь вид:

$$e_i^2 = \alpha_1 + \alpha_2 x_i^{(1)} + \alpha_3 x_i^{(2)} + \alpha_4 x_i^{(3)} + \alpha_5 (x_i^{(1)})^2 + \alpha_6 (x_i^{(2)})^2 + \\ + \alpha_7 (x_i^{(3)})^2 + \alpha_8 x_i^{(1)} x_i^{(2)} + \alpha_9 x_i^{(1)} x_i^{(3)} + \alpha_{10} x_i^{(2)} x_i^{(3)} + u_i$$

$\boxed{\text{б}}$ Проведем теперь сам тест Уайта. $R^2 = 0{,}45$.
Его статистика имеет вид $\chi^2_{\text{расч}} = nR^2$, где коэффициент детерминации берется из вспомогательной регрессии. Ее мы сравниваем с $\chi^2_{\text{кр}}(p)$, где p – число регрессоров вспомогательной регрессии, не считая константу. В нашем случае их 9.

$$\chi^2_{\text{расч}} = 180 \cdot 0.45 = 81$$

$$\chi^2_{\text{кр}}(0.05, 9) = 16.92$$

Как мы видим, расчетное значение больше критического. Мы отклоняем нулевую гипотезу. С вероятностью 95% в нашей модели есть гетероскедастичность.

[2] При этом квадраты фиктивных переменных не берутся

6.2.2 Задача 2

Исследователь при помощи МНК оценил коэффициенты в следующем уравнении регрессии:

$$y_i = \beta_1 + \beta_2 x_i^{(1)} + \beta_3 x_i^{(2)} + \beta_4 x_i^{(3)} + \varepsilon_i$$

Число наблюдений равно 180. После этого он решил провести тест Голдфелда – Квандта. Он отсортировал выборку по возрастанию регрессора $x^{(1)}$ и разбил ее на три равных части. Проведем этот тест за него.

Выпишем гипотезы теста[3]:

$$H_0 : \sigma_1^2 = ... = \sigma_{180}^2$$

$$H_1 : \sigma_i^2 = const \cdot x_i^{(1)}$$

Как известно, чтобы провести тест Голдфелда-Квандта, надо расположить наблюдения по возрастанию подозреваемой переменной.

После этого выборка делится на три части[4]. Средная часть не рассматривается временно, а отдельно по левой (меньшая треть) и по правой (большая треть) оценивается заново наша регрессия.

Таким образом, сначала мы оцениваем регрессию по первым 60 наблюдениям. $TSS_1 = 1000, R^2 = 0.4, ESS = 600$.

После этого мы оцениваем ее по последним 60 наблюдениям. $TSS = 1000, R^2 = 0.3, ESS = 700$.

После этого мы выписываем F-статистику теста. $F_{\text{расч}} = \frac{ESS_{\text{больший}}}{ESS_{\text{меньший}}}$. Сравнивается с $F_{\text{кр}}(n' - k, n' - k)$, где n' – число наблюдений в каждой подвыборке, а k – число регрессоров.

Таким образом, $F_{\text{расч}} = \dfrac{7}{6}$, а $F_{\text{кр}}(60 - 4{,}60 - 4{,}0.05) = 1.56$. Значит, так как расчетное значение меньше критического, мы принимаем нулевую гипотезу. С вероятностью 95% гетероскедастичности такой формы в модели нет.

---

[3]Сама формулировка задания намекает нам, что он подозревает пропорциональность первому из регрессоров.

[4]Если число наблюдений не делится на три, мы делаем так, чтобы в крайних подвыборках было одинаковое число наблюдений

6.2.3 Задача 3

Исследователь рассматривает модель с гетероскедастичностью и детерминированными регрессорами:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$$

$$\sigma_i^2 = c^2 x_i^2$$

Данные имеют вид:

Y	1	1	1.5	1.5	1
X	1	1	0.5	0.5	0.25

$\boxed{\text{а}}$ Покажем, что переход к взвешенной модели решит проблему гетероскедастичности:

$$\frac{y_i}{x_i} = \beta_1 \frac{1}{x_i} + \beta_2 + \frac{\varepsilon_i}{x_i}$$

Переобозначая, получим модель:

$$y_i^* = \beta_1 x_i^* + \beta_2 + u_i$$

Очевидно, что, гетероскедастичности в этой модели нет:

$$\mathrm{Var}(u_i) = \mathrm{Var}\left(\frac{\varepsilon_i}{x_i}\right) = \frac{\mathrm{Var}(\varepsilon_i)}{x_i^2} = \frac{c^2 x_i^2}{x_i^2} = c^2$$

Аналогичным образом гетероскедастичность убивается всегда, когда есть знания о связи: мы просто делим на корень.

$\boxed{\text{б}}$ Оценим теперь саму модель. Как известно,

$$\hat{\beta}_1 = \frac{\overline{x^* y^*} - \overline{x^*} * \overline{y^*}}{\overline{(x^*)^2} - (\overline{x^*})^2}$$

Таким образом, находим, что модель имеет вид:

$$\hat{y}_i = x_i^* + 0.4$$

6.2.4 Задача 4

Исследователь рассматривает модель регрессии на константу.

$$y_i = \beta + \varepsilon_i$$

При этом случайные ошибки имеют несколько свойств: $\mathrm{E}(\varepsilon_i) = 0; \mathrm{E}(\varepsilon_i \varepsilon_j) = 0, i \neq j; \mathrm{E}(\varepsilon_i^2) = \sigma^2 x_i, x_i > 0$.

$\boxed{\text{а}}$ Вычислим для начала обычную МНК-оценку β и найдем ее дисперсию.

Как мы помним, $\hat{\beta} = \overline{y}$.

$$\mathrm{E}(\hat{\beta}) = \frac{1}{n}\mathrm{E}(\sum y_i) = \frac{1}{n}\sum \mathrm{E}(\beta + \varepsilon_i) = \beta$$

Теперь вычислим дисперсию:

$$\mathrm{Var}(\hat{\beta}) = \mathrm{E}(\hat{\beta} - E(\hat{\beta}))^2 = \mathrm{E}(\hat{\beta} - \beta)^2 = \mathrm{E}\left(\overline{y} - \beta\right)^2 = \mathrm{E}\left(\beta + \frac{\sum \varepsilon_i}{n} - \beta\right) = \mathrm{E}\left(\frac{\sum \varepsilon_i}{n}\right)^2$$

Легко осознать, что $(\sum \varepsilon_i)^2 = \sum \varepsilon_i^2 + \sum\limits_{i \neq j}\sum \varepsilon_i \varepsilon_j$.

Таким образом,

$$\mathrm{E}\left(\frac{\sum \varepsilon_i}{n}\right)^2 = \frac{\mathrm{E}(\sum \varepsilon_i^2 + \sum\limits_{i \neq j}\sum \varepsilon_i \varepsilon_j)}{n^2} = \frac{\mathrm{E}(\sum \varepsilon_i^2) + \mathrm{E}(\sum\limits_{i \neq j}\sum \varepsilon_i \varepsilon_j)}{n^2} = \frac{\sigma^2 \sum x_i}{n^2}$$

$\boxed{\text{б}}$ Теперь рассмотрим эффекты взвешенного МНК.
Перепишем нашу модель:

$$\frac{y_i}{\sqrt{x_i}} = \beta \frac{1}{\sqrt{x_i}} + \frac{\varepsilon_i}{\sqrt{x_i}}$$

В новых переменных,

$$y_i^* = \beta x_i^* + u_i$$

Как мы помним, в таком случае оценка имеет вид:

$$\hat{\beta}^{WLS} = \frac{\overline{x^* y^*}}{\overline{(x^*)^2}}$$

Покажем сначала несмещенность:

$$E(\hat{\beta}^{WLS}) = E\left(\frac{\sum x_i^*(\beta x_i^* + u_i)}{\sum (x_i^*)^2}\right) = E\left(\frac{\beta \sum (x_i^*)^2 + \sum x_i^* u_i)}{\sum (x_i^*)^2}\right) =$$
$$= \beta + E\left(\frac{\sum x_i^* u_i}{\sum (x_i^*)^2}\right) = \beta$$

Вычислим теперь дисперсию нашей оценки:

$$\mathrm{Var}(\hat{\beta}^{WLS}) = E(\hat{\beta}^{WLS} - E(\hat{\beta}^{WLS}))^2 = E(\hat{\beta}^{WLS} - \beta)^2 =$$
$$= E\left(\beta + \frac{\sum x_i^* u_i}{\sum (x_i^*)^2} - \beta\right)^2 = E\left(\frac{\sum x_i^* u_i}{\sum (x_i^*)^2}\right)^2$$

Как и в прошлый раз, помним, что:

$$E\left(\sum x_i^* u_i\right)^2 = E\left(\sum (x_i^* u_i)^2 + \sum_{i \neq j}\sum x_i^* u_i x_j^* u_j\right) = \sum E(x_i^{*2} u_i^2) +$$
$$+ \sum_{i \neq j}\sum x_i^* x_j^* E(u_i u_j) = \sum E(x_i^{*2} u_i^2)$$
$$E(u_i u_j) = E\left(\frac{\varepsilon_i}{\sqrt{x_i}} \frac{\varepsilon_j}{\sqrt{x_j}}\right) = \frac{E(\varepsilon_i \varepsilon_j)}{\sqrt{x_i^* x_j^*}} = 0$$

Продолжаем вычисление дисперсии:

$$E\left(\frac{\sum x_i^* u_i}{\sum (x_i^*)^2}\right)^2 = \frac{E\left(\sum (x_i^{*2} u_i^2)\right)}{(\sum (x_i^*)^2)^2} = \frac{\sum (x_i^*)^2 E(u_i)^2}{\left(\sum (x_i^*)^2\right)^2} = \frac{\sigma^2}{\sum (x_i^*)^2} = \frac{\sigma^2}{\sum \dfrac{1}{x_i}}$$

$\boxed{\text{в}}$ Выпишем и сравним дисперсии:

$$\mathrm{Var}(\hat{\beta}^{OLS}) = \frac{\sigma^2 \sum x_i}{n^2}$$

$$\mathrm{Var}(\hat{\beta}^{WLS}) = \frac{\sigma^2}{\sum \frac{1}{x_i}}$$

Сравнение:

$$\frac{\mathrm{Var}(\hat{\beta}^{OLS})}{\mathrm{Var}(\hat{\beta}^{WLS})}?1$$

Факт: в произвольном нормированном пространстве $\forall a, b \; ||ab|| \leq ||a||||b||$.

Воспользуемся неравенством Коши-Шварца для помощи себе: пусть $a_i = \sqrt{x_i}, b_i = \dfrac{1}{\sqrt{x_i}}$. Тогда $a_i b_i = 1, \sum a_i b_i = n$.

Соответственно, получаем,

$$\sum x_i \sum \frac{1}{x_i} \geq n^2$$

$$\frac{1}{n^2} \sum x_i \geq \left(\sum \frac{1}{x_i} \right)^{-1}$$

$$\frac{\sigma^2}{n^2} \sum x_i \geq \frac{\sigma^2}{\sum \dfrac{1}{x_i}}$$

Таким образом, дисперсия МНК-оценки больше дисперсии ВМНК-оценки.

6.2.5 Задача 5

Исследователь рассматривает модель регрессии без константы.

$$y_i = \beta x_i + \varepsilon_i$$

При этом случайные ошибки имеют несколько свойств: $E(\varepsilon_i) = 0; E(\varepsilon_i \varepsilon_j) = 0, i \neq j; E(\varepsilon_i^2) = a x_i^2, x_i > 0$. $\sum x_i^2 = n$.

$\boxed{\text{а}}$ Вычислим для начала обычную МНК-оценку β и найдем ее дисперсию.

$$\hat{\beta}^{OLS} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\overline{xy}}{\overline{x^2}}$$

Она, как известно, несмещенная,

$$E(\hat{\beta}^{OLS}) = E\left(\frac{\sum x_i y_i}{\sum x_i^2}\right) = \frac{1}{n}E(\sum x_i(\beta x_i + \varepsilon_i)) =$$
$$= \frac{1}{n}E(\sum \beta x_i^2 + \sum x_i\varepsilon_i) = \beta + \frac{\sum E(x_i\varepsilon_i)}{n} = \beta$$

Ищем теперь дисперсию:

$$\text{Var}(\hat{\beta}^{OLS}) = E\left(\hat{\beta}^{OLS} - \beta\right)^2 = E\left(\frac{\sum x_i y_i}{\sum x_i^2} - \beta\right)^2 = E\left(\frac{\sum x_i(\beta x_i + \varepsilon_i)}{n} - \beta\right)^2 =$$

$$= \frac{1}{n^2}E(\beta\sum x_i^2 + \sum x_i\varepsilon_i - n\beta)^2 = \frac{1}{n^2}E(\beta n - n\beta + \sum x_i\varepsilon_i)^2 =$$

$$= \frac{1}{n^2}E(\sum x_i\varepsilon_i)^2$$

Как мы понимаем:

$$E\left(\sum x_i\varepsilon_i\right)^2 = \left(\sum x_i^2\varepsilon_i^2 + \sum_{i\neq j}\sum x_i\varepsilon_i x_j\varepsilon_j\right) = E\left(\sum x_i^2\varepsilon_i^2\right)$$

Подставляем это назад:

$$\frac{1}{n^2}E(\sum x_i\varepsilon_i)^2 = \frac{E\left(\sum x_i^2\varepsilon_i^2\right)}{n^2} = \frac{\sum x_i^2 E(\varepsilon_i^2)}{n^2} = \frac{\sum x_i^2 a x_i^2}{n^2} = \frac{a\sum x_i^4}{n^2}$$

$\boxed{\text{б}}$ Теперь рассмотрим эффекты взвешенного МНК.
Перепишем нашу модель:

$$\frac{y_i}{x_i} = \beta\frac{x_i}{x_i} + \frac{\varepsilon_i}{x_i}$$

,

В новых переменных,

$$y_i^* = \beta + u_i$$

Оценка имеет вид:

$$\hat{\beta}^{WLS} = \overline{y_i^*}$$

Покажем несмещенность:

$$E(\hat{\beta}^{WLS}) = E(\overline{y^*}) = E\left(\frac{1}{n}\sum(\beta + u_i)\right) = \beta + \frac{1}{n}E\left(\sum u_i\right) = \beta$$

Вычислим дисперсию ВМНК-оценки:

$$\mathrm{Var}(\hat{\beta}^{WLS}) = E(\hat{\beta}^{WLS} - \beta)^2 = E(\overline{y^*} - \beta)^2 = E\left(\frac{1}{n}\sum(\beta + u_i) - \beta\right)^2 =$$

$$= \frac{1}{n^2}E\left(n\beta - n\beta + \sum(u_i)\right)^2 = \frac{1}{n^2}E\left(\sum u_i\right)^2 =$$

$$= \frac{1}{n^2}E\left(\sum u_i^2 + \sum_{i \neq j}\sum u_j u_i\right) = \frac{1}{n^2}\sum E(u_i^2)$$

$$E(u_i^2) = E\left(\frac{\varepsilon_i^2}{x_i^2}\right) = a$$

$$\frac{1}{n^2}\sum E(u_i^2) = \frac{a}{n}$$

$\boxed{\text{в}}$ Сравним наши дисперсии:

$$\mathrm{Var}(\hat{\beta}^{OLS}) = \frac{a\sum x_i^4}{n^2}$$

$$\mathrm{Var}(\hat{\beta}^{WLS}) = \frac{a}{n}$$

Снова воспользуемся неравенством Коши-Шварца.

$$||ab|| \leq ||a||||b||$$

У нас это $ab \leq a^2b^2$. Пусть наши векторы:

$$a = \begin{pmatrix} x_1^2 \\ \vdots \\ x_n^2 \end{pmatrix}; b = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

Тогда мы получим, что $ab = n, a^2 = \sum x_i^4, b^2 = n$.

$$n \sum x_i^4 \geq n$$

$$\sum x_i^4 \geq 1$$

$$\frac{a \sum x_i^2}{n} \geq \frac{a}{n}$$

Значит, мы получаем, что ВМНК-оценка имеет меньшую дисперсию.

6.2.6 Задача 6

По 540 наблюдениям исследовать МНК-оценил следующее уравнение:

$$\widehat{\text{EARNINGS}}_i = \underset{(4.12)}{-26.40} + \underset{(0.21)}{0.56}\,\text{EXP}_i + \underset{(1.1)}{2.67}\,\text{S}_i \tag{6.2}$$

При этом ESS=89496.58, SEE = 12.90970,

$\boxed{\text{а}}$ Проверим все коэффициенты на значимость.
Формулируем гипотезы:

$$H_0^k : \beta_k = 0$$
$$H_1^k : \beta_k \neq 0$$

Статистика имеет вид $t_{\text{расч}} = \dfrac{\hat{\beta}_k}{se(\beta_k)}$. $t_{\text{кр}}(537, 0.05) = 1{,}96$.

1. $t_{\text{расч}} = \dfrac{-26.4}{4.12} = -6.4$
2. $t_{\text{расч}} = \dfrac{0.56}{0.21} = 2.67$
3. $t_{\text{расч}} = \dfrac{2.67}{1.1} = 2.43$

Все коэффициенты значимы на 5-процентном уровне значимости.

$\boxed{\text{б}}$ Уважаемый исследователь сделал тест Уайта.

$$H_0 : \sigma_1^2 = ... = \sigma_{180}^2$$

$$H_1 : \text{В модели есть гетероскедастичность}$$

Расчетное значение статистики оказалось $\chi^2_{\text{расч}} = 19.609$. $\chi^2_{\text{кр}}(0.05, 5) = 11.07$. Наше расчетное значение превышает критическое значение, мы отклоняем нулевую гипотезу. В модели есть гетероскедастичность.

P-значение = 0,0014. Это меньше даже одной сотой. Значит, мы можем утверждать, что есть гетероскедастичность с уверенностью 99%.

$\boxed{\text{в}}$ После этого исследователь, используя те же самый данные, оценил такое же уравнение, но с использованием оценок стандартных ошибок в форме Уайта.

$$\widehat{\text{EARNINGS}}_i = \underset{(4.21)}{-26.40} + \underset{(0.22)}{0.56}\,\text{EXP}_i + \underset{(1.5)}{2.67}\,\text{S}_i \tag{6.3}$$

Как мы видим, используя робастные стандартные ошибки, мы получили более высокие значения оценок стандартных ошибок.

Однако мы можем сказать, что использование робастных ошибок позволяет нам корректно вычислять и использовать t-статистику.

Проверим коэффициенты на значимость:

1. $t_{\text{расч}} = \dfrac{-26.4}{4.21} = -6.27$
2. $t_{\text{расч}} = \dfrac{0.56}{0.22} = 2.54$
3. $t_{\text{расч}} = \dfrac{2.67}{1.5} = 1.78$

Как мы видим, теперь значимы на 5-процентном уровне значимости только константа и коэффициент при регрессоре EXP.

Доверять мы можем результатам, полученным с использованием робастных ошибок, так как в нашей модели есть гетероскедастичность.

6.3 Семинар 13-15. Асимптотический подход. IV регрессия

6.3.1 Задача 1

Пусть $x_1 ... x_n$ – выборка из генеральной совокупности независимых нормально распределенных случайных величин с математическим ожиданием a и дисперсией σ^2. Уважаемого исследователя интересует математическое ожидание a, и он рассматривает три возможных оценки этого параметра, которые можно вычислить на основе имеющихся данных:

1. $\hat{a} = \frac{1}{n} \sum x_i$;
2. $\hat{a} = \frac{\sum x_i}{n-1}$
3. $\hat{a} = \frac{0.5}{1-0.5^n} \sum\limits_{i=1..n} 0.5^{i-1} x_i$.

$\boxed{\text{a}}$ Рассмотрим сперва вопрос о несмещенности каждой из оценок.

1. Очевидно, что она несмещенная:
$$\mathrm{E}\left(\frac{1}{n}\sum x_i\right) = \frac{1}{n}\sum \mathrm{E}(x_i) = \frac{1}{n}\sum a = \frac{1}{n} na = a$$
2. Наблюдаем смещенность:
$$\mathrm{E}\left(\frac{\sum x_i}{n-1}\right) = \frac{\sum \mathrm{E}(x_i)}{n-1} = \frac{na}{n-1}$$
3. Вспомним одно тривиальное свойство – сумму членов геометрической прогрессии. Пусть b – первый член последовательности, m – число членов в ней, $r \neq 1$ – ее знаменатель.

   Тогда мы получаем, что:
$$br^0 + br + ... + br^{m-1} = \frac{b(1-r^m)}{1-r}$$

   Используя это, можем изучить свойства нашей третьей оценки:

$$\mathrm{E}\left(\frac{0.5}{1-0.5^n}\sum_{i=1}^{n}0.5^{i-1}x_i\right)=\frac{0.5}{1-0.5^n}\sum 0.5^{i-1}a=$$

$$=a\frac{0.5}{1-0.5^n}(1+0.5+...+0.5^{n-1})=a\frac{0.5}{1-0.5^n}\frac{1-0.5^n}{0.5}=a$$

Как много трудов ради несмещенной оценки.

$\boxed{\text{б}}$ Рассмотрим теперь вопрос об асимптотической несмещенности наших оценок:

1. Она уже и так несмещенная, на бесконечности такой и останется.
2. Так как предел $\lim \dfrac{n}{n-1}=1$ наша оценка асимптотически несмещенная.
3. И так несмещенная.

Рассмотрим теперь вопрос о состоятельности наших оценок. Все наши оценки хотя бы асимптотически несмещенные, поэтому не можем пока отбросить хоть одну из них.

Чтобы проверить состоятельность, надо вычислить дисперсии оценок.

1. Считаем дисперсию:

$$\mathrm{Var}(\hat{a})=\mathrm{Var}\left(\frac{\sum x}{n}\right)=\frac{1}{n^2}\mathrm{Var}(\sum x_i)=\frac{1}{n^2}n\,\mathrm{Var}(x_i)=\frac{\sigma^2}{n}\to 0$$

   Оценка состоятельна.

2. Поступаем аналогично:

$$\mathrm{Var}(\hat{a})=\mathrm{Var}\left(\frac{\sum x_i}{n-1}\right)=\frac{1}{(n-1)^2}\sum\mathrm{Var}(x_i)=\frac{n\sigma^2}{(n-1)^2}=\frac{n\sigma^2}{n^2-2n+1}\to 0$$

   Тоже чудесно состоятельная оценка.

3. Посчитаем дисперсию здесь:

$$\mathrm{Var}(\hat{a}) = \mathrm{Var}\left(\frac{0.5}{1-0.5^n}\sum_{i=1}^{n} 0.5^{i-1}x_i\right) = \frac{0.5^2}{(1-0.5^n)^2}\sum 0.5^{2i-2}\,\mathrm{Var}(x_i) =$$

$$= \sigma^2\frac{0.5^2}{(1-0.5^n)^2}(1+0.5^2+\ldots+0.5^{2n-2}) = \sigma^2\frac{0.25}{1-2\cdot 0.5^n+0.25^n}\frac{1-0.25^n}{1-0.25} =$$

$$= \frac{\sigma^2}{3}\frac{1-0.25^n}{1-2\cdot 0.5^n+0.25^n} = \frac{\sigma^2}{3}\frac{(4^n-1)}{4^n-2\cdot 2^n+1} \to \frac{\sigma^3}{3}$$

Магии не случилось. Если бы знали распределение, можно было бы даже числа подобрать для демонстрации несостоятельности.

6.3.2 Задача 2

Уважаемый исследователь использует данные размере класса (CS) и средних по каждому классу оценках за контрольную работу ($TestScore$) (всего 100 наблюдений), чтобы оценить следующую регрессию:

$$\widehat{TestScore}_i = \underset{(20.4)}{520.4} - \underset{(2.21)}{5.82}CS_i$$

При этом у него $R^2 = 0.08$, $SEE = 11.5$.

$\boxed{\text{а}}$ Построим 95-процентный доверительный интервал для коэффициента при переменной.

Мы помним формулу:

$$\left(\hat{\beta}_2 - t_{\text{кр}}se(\hat{\beta}_2); \hat{\beta}_2 + t_{\text{кр}}se(\hat{\beta}_2)\right)$$

$t_{\text{кр}}(0.05,\infty) = 1.96$. Подставляем в нашу формулу:

$$\beta_2 \in (-5.82 - 1.96\cdot 2.21; -5.82 + 1.96\cdot 2.21)$$

$$\beta_2 \in (-10.15; -1.49)$$

$\boxed{\text{б}}$ Вычислим p-значение для теста на значимость переменной. Вспомним определение:

Определение 6.3.1. p-значение – уровень значимости, при котором нулевая гипотеза теста находится на грани принятия/отклонения.

Таким образом, мы должны найти, какое α соответствует критическому значению, равному расчетному.

$$t_{\text{расч}} = \frac{-5.82}{2.21} = -2.63$$

.

Теперь можем просто найти, исходя из геометрических соображений. Получаем две формы для поиска:

- $p_{val} = 2\Phi(-|t_{\text{расч}}|)$
- $p_{val} = 2(1 - \Phi(|t_{\text{расч}}|))$.

Так как сейчас у нас отрицательное расчетное значение, используем первую формулу. В MS EXCEL:

$$= 2 * \text{НОРМ.СТ.РАСП}(2.63,\ 1)$$

В итоге получаем $p = 0.0085$.

$\boxed{\text{в}}$ Вычислим p-значение для гипотезы $H_0 : \beta_2 = -5.6$.
Считает расчетное значение:

$$t_{\text{расч}} = \frac{-5.82 - (-5.6)}{2.21} = -0.0993$$

Получаем, что p-значение равно 0.92.

$\boxed{\text{г}}$ Построим 99-процентный асимптотический доверительный интервал для коэффициента β_1.
Мы помним формулу:

$$\hat{\beta}_1 - t_{\text{кр}} se(\hat{\beta}_1); \hat{\beta}_1 + t_{\text{кр}} se(\hat{\beta}_1)$$

$t_{\text{кр}}(0.01, \infty) = 2.57$. Подставляем в нашу формулу:

$$\beta_1 \in (520.4 - 2.57 \cdot 20.4; 520.4 + 2.57 \cdot 20.4)$$

6.3.3 Задача 3

Решение в разделе 10.1.5.

6.3.4 Задача 4

Рассмотрим линейную регрессионую модель:

$$y_i = \beta_1 + \beta_2 x_i + u_i$$

Пусть выполнены все предпосылки модели со стохастическими регрессорами, кроме предпосылки некоррелированости регрессора со случайной ошибкой. Пусть $\mathrm{Cov}(x,u) > 0$.

$\boxed{\text{а}}$ Рассмотрим вопрос, состоятельна ли тогда МНК-оценка коэффициента при регрессоре.

Как мы помним,

$$\hat{\beta}_2^{OLS} = \frac{\widehat{\mathrm{Cov}}(x,y)}{\widehat{\mathrm{Var}}(x)} = \frac{\sum(x_i - \overline{x})(y_i - \overline{y}))}{\sum(x_i - \overline{x})^2} =$$

$$= \frac{\sum(x_i - \overline{x})(\beta_1 + \beta_2 x_i + u_i - \beta_1 - \beta_2 \overline{x} - \overline{u})}{n\,\widehat{\mathrm{Var}}(x)} =$$

$$= \frac{\sum(x_i - \overline{x})(\beta_2(x_i - \overline{x}) + u_i - \overline{u_i})}{n\,\widehat{\mathrm{Var}}(x)} = \beta_2 \frac{\sum(x_i - \overline{x})^2}{\sum(x_i - \overline{x})^2} + \frac{\sum(x_i - \overline{x})(u_i - \overline{u}))}{n\,\widehat{\mathrm{Var}}(x)} =$$

$$= \beta_2 + \frac{\widehat{\mathrm{Cov}}(x,u)}{\widehat{\mathrm{Var}}(x)}$$

Таким образом,

$$\mathrm{plim}\,\hat{\beta}_2^{OLS} = \beta_2 + \frac{\mathrm{Cov}(x,u)}{\mathrm{Var}(x)}$$

Мы имеем завышенную несостоятельную оценку.

$\boxed{\text{б}}$ Уважаемый исследователь пытается обойти эту проблему, используя альтернативную оценку. У него есть переменная z, отвечающая требованиям к инструменту.

Новый вариант оценки:

$$\hat{\beta}_2 = \frac{\widehat{\mathrm{Cov}}(z,y)}{\widehat{\mathrm{Cov}}(z,x)}$$

Покажем ее состоятельность:

$$\hat{\beta}_2 = \frac{\sum(z_i - \overline{z})(y_i - \overline{y})}{\sum(x_i - \overline{x})(z_i - \overline{z})} = \frac{\sum(z_i - \overline{z})(\beta_2(x_i - \overline{x}) + u_i - \overline{u})}{\sum(z_i - \overline{z})(x_i - \overline{x})} = \beta_2 + \frac{\widehat{\mathrm{Cov}}(z,u)}{\widehat{\mathrm{Cov}}(x,z)}$$

Согласно условию, $\mathrm{Cov}(z,u) = 0$.
Берем предел:

$$\operatorname{plim} \hat{\beta}_2 = \beta_2 + \frac{\mathrm{Cov}(z,u)}{\mathrm{Cov}(x,z)} = \beta_2$$

6.3.5 Задача 5

Пусть мы имеем модель:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$$

Пусть регрессор x эндогенен. Пусть у нас есть валидный инструмент z. Выведем 2МНК-оценки коэффициентов

Мы имеем пару уравнений:

$$x_i = a_1 + a_2 z_i + v_i$$
$$y_i = \beta_1 + \beta_2 \hat{x}_i + \varepsilon_i$$

$\boxed{1}$ Найдем оценки I шага:

$$\hat{a}_2 = \frac{\widehat{\mathrm{Cov}}(x,z)}{\widehat{\mathrm{Var}}(z)}$$

$\boxed{б}$ Найдем МНК-оценки второго шага:

$$\hat{\beta}_2^{TSLS} = \frac{\widehat{\mathrm{Cov}}(y,\hat{x})}{\widehat{\mathrm{Var}}(\hat{x})} = \frac{\widehat{\mathrm{Cov}}(y, a_0 + a_1 z + v_i)}{\widehat{\mathrm{Var}}(a_0 + a_1 z_i} = \frac{a_1 \widehat{\mathrm{Cov}}(y,z)}{a_1^2 \widehat{\mathrm{Var}}(z)} = \frac{\widehat{\mathrm{Cov}}(y,z)}{a_1 \widehat{\mathrm{Var}}(z)} =$$

$$= \frac{\widehat{\mathrm{Cov}}(y,z)}{\frac{\widehat{\mathrm{Cov}}(x,z)}{\widehat{\mathrm{Var}}(z)} \widehat{\mathrm{Var}}(z)} = \frac{\widehat{\mathrm{Cov}}(y,z)}{\widehat{\mathrm{Cov}}(x,z)} = \frac{\frac{1}{n}\sum(y_i - \overline{y})(z_i - \overline{z}))}{\frac{1}{n}\sum(x_i - \overline{x})(z_i - \overline{z})}$$

Значит, $\hat{\beta}_2^{TSLS} = \dfrac{\widehat{\mathrm{Cov}}(y,z)}{\widehat{\mathrm{Cov}}(x,z)}$.

$$\hat{\beta}_1^{TSLS} = \overline{y} - \hat{\beta}_2^{TSLS}\overline{\hat{x}}$$

.

6.3.6 Задача 6

На рынке сигарет в некоторой стране функция спроса в i-м регионе имеет вид:

$$\ln(Q_i) = \beta_1 + \beta_2 \ln(P_i) + \varepsilon_i$$

Функция предложения описывается соотношением:

$$\ln(Q_i) = \gamma_1 + \gamma_2 \ln(P_i) + \gamma_3 \ln(T_i) + u_i$$

Q_i — количество сигарет в i-м регионе, P_i — цена сигарет в i-м регионе, T_i — налог с продаж в i-м регионе. ε_i — независимые и одинаково распределенные случайные величины, характеризующие шоки спроса (не коррелированы с налогами), u_i — независимые и одинаково распределенные случайные величины, характеризующие шоки предложения.

$\boxed{\text{а}}$Покажем эндогенность цены и несостоятельность МНК-оценок для спроса.

Цены и объемы определяются пересечениями спросов и предложений:

$$\beta_1 + \beta_2 \ln(P_i) + \varepsilon_i = \gamma_1 + \gamma_2 \ln(P_i) + \gamma_3 \ln(T_i) + u_i$$

Выражаем логарифм цены:

$$\ln P_i = \frac{\gamma_1 - \beta_1 + \gamma_3 \ln T_i + u_i + \varepsilon_i}{\beta_2 - \gamma_2}$$

Найдем ковариацию логарифма цены и шоков спроса. Мы помним, что налоги некоррелированы с ними:

$$\mathrm{Cov}(lnP, \varepsilon) = \mathrm{Cov}\left(\frac{\gamma_1 - \beta_1 + \gamma_3 \ln T_i + u_i + \varepsilon_i}{\beta_2 - \gamma_2}, \varepsilon\right) = \frac{\mathrm{Var}(\varepsilon)}{\gamma_2 - \beta_2}$$

Теперь рассмотрим МНК-оценку:

$$\hat{\beta}_2^{OLS} = \frac{\widehat{\mathrm{Cov}}(\ln P, \ln Q)}{\mathrm{Var}(\ln P)} = \frac{\widehat{\mathrm{Cov}}(\ln P, \beta_1 + \beta_2 \ln P + \varepsilon_i)}{\widehat{\mathrm{Var}}(\ln P)} =$$

$$= \beta_2 + \frac{\widehat{\mathrm{Cov}}(\ln P, \varepsilon)}{\widehat{\mathrm{Var}}(\ln P)} \to \beta_2 + \frac{\mathrm{Var}(\varepsilon)}{(\beta_2 - \gamma_2)\,\mathrm{Var}(\ln P)}$$

Из экономического смысла подозревается ординарность сигарет $\gamma_2 > 0, \beta_2 < 0$. Дисперсии положительны. Тогда мы получаем смещенную завышенную несостоятельную оценку коэффициента при логарифме цены.

$\boxed{\text{б}}$ Состоятельными будут 2МНК: эндогенный регрессор – цена, инструмент – налоги.

На первом шаге оцениваем регрессию:

$$\ln P_i = a_1 + a_2 \ln T_i + v_i$$

На втором шаге оцениваем спрос:

$$\ln Q_i = \beta_1 + \beta_2 \widehat{\ln P}_i + \varepsilon_i$$

2МНК-оценка:

$$\hat{\beta}_2^{TSLS} = \frac{\widehat{\mathrm{Cov}}(\ln Q, \ln T)}{\widehat{\mathrm{Cov}}(\ln P, \ln T)} \xrightarrow{p} \frac{\mathrm{Cov}(\ln T_i, \ln Q_i)}{\mathrm{Cov}(\ln T_i, \ln P_i)} = \frac{\mathrm{Cov}(\ln T_i, \beta_1 + \beta_2 \ln P_i + \varepsilon_i)}{\mathrm{Cov}(\ln T_i, \ln P_i)} = \beta_2$$

Решительно состоятельная.

6.3.7 Задача 7

Уважаемый исследователь анализирует спрос на сигареты в 48 американских штатах. Он исследует зависимость величины спроса (Q) от цены (P), используя в качестве инструмента для цены ставку налога, взимаемого с производителей в соответствующем штате (T).

Уважаемый исследователь получил следующие результаты: на первом шаге $R^2 = 0.47$

$$\widehat{\ln P}_i = \underset{0.03}{4.63} + \underset{0.005}{0.03} \ln T_i$$

На втором шаге (с робастными к гетероскедастичности ошибками):

$$\widehat{\ln Q_i} = \underset{1.53}{9.72} - \underset{0.32}{1.08}\widehat{\ln P_i}$$

$\boxed{\text{а}}$ Проверим силу нашего инструмента. Так как у нас есть только один эндогенный регрессор, может использовать правило $F_{\text{расч}} > 10$. Проверяем уравнение первой стадии на значимость:

$$H_0 : a_1 = 0$$

$$H_1 : a_1 \neq 0$$

Можем просто посчитать квадрат t-статистики.

$$t_{\text{расч}} = \frac{0.03}{0.005} = 6$$

$$F_{\text{расч}} = t^2_{\text{расч}} = 36 > 10$$

Значит, мы имеем сильные инструменты.

$\boxed{\text{б}}$ Построим теперь асимптотический 95-процентных доверительный интервал для эластичности спроса по цене.

$$\beta_2 \in (\hat{\beta}_2 - 1.96se(\hat{\beta}_2); \hat{\beta}_2 + 1.96se(\hat{\beta}_2)$$

$$\beta_2 \in (-1.71, -0.45)$$

Цена влияет значимо на величину спроса, но мы не можем определить категорию товара, так как интервал захватывает единицу.

6.3.8 Задача 8

Пусть все тот же уважаемый исследователь теперь считает, что на спрос влияет не только цены сигарет, но и уровень доходов потребителей. Таким образом, его модель теперь:

$$\ln(Q_i) = \beta_0 + \beta_1 \ln(P_i) + \beta_2 \ln(Income_i) + \varepsilon_i$$

Цена все еще, конечно, эндогенная, а доход экзогенен. Кроме этого он нашел еще пару инструментов:

- $\ln(rtaxso_i)$ – логарифм налога с продаж, общего для всех товаров в i-м регионе;
- $\ln(rtax_i)$ – логарифм акциза на продажу сигарет в i-м регионе;

$\boxed{\text{а}}$ Он будет использовать 2МНК для оценки модели. На первом шаге он будет работать со следующей моделью:

$$\ln P_i = a_1 + a_2 \ln(rtaxso_i) + a_3 \ln(rtax_i) + a_4 \ln Income_i + u_i$$

$\boxed{\text{б}}$ Дополнительно проверим силу инструментов: нам известно, что расчетное значение статистики равно 900.

Формулируем гипотезы:

$$H_0 : a_2 = a_3 = 0$$
$$H_1 : \exists j \in (2,3) : a_j \neq 0$$

Мы получаем просто тест на короткую и длинную регрессию при $q = 2$, $k + q = 3$.

$$F_{\text{расч}} = \frac{(R^2_{UR} - R^2_R)/q}{(1 - R^2_{UR})/(n - (k+q))} = 900 > 10$$

Таким образом, наши инструменты в целом сильные.

$\boxed{\text{в}}$ Уважаемый исследователь получил следующие результаты на втором шаге:

$$\widehat{\ln Q}_i = \underset{(1.20)}{8.40} - \underset{(0.20)}{1.20}\widehat{\ln P}_i + \underset{(0.31)}{0.46} \ln Income_i$$

Проверим всем переменные на значимость и дадим содержательную интерпретацию.

- $t^1_{\text{расч}} = \dfrac{8.40}{1.20}$. Константа значима на уровне 1%.
- $t^2_{\text{расч}} = -\dfrac{1.20}{0.20} = -6$. Логарифм цены значим на уровне 1%.
- $t^3_{\text{расч}} = \dfrac{0.46}{0.31} = 1.48$. Доход не влияет значимо даже на уровне 5%.

При росте цены на 1% величина спроса падает на 1.2%.

6.3.9 Задача 11

Уважаемого исследователя интересует ответ на вопрос: как переменная x влияет на переменную y? Исследователь предполагает, что переменная x является эндогенной.

Он располагает данными о переменной x, об экзогенной переменной w, а также о нескольких переменных, которые могут использоваться как инструменты для переменной x: $z^{(1)}$, $z^{(2)}$, $z^{(3)}$ и $z^{(4)}$.

Исследователь оценил пять моделей, используя двухшаговый МНК и различные комбинации инструментов. Результаты оценивания моделей представлены в таблице. Поможем ему выбрать ему выбрать лучший вариант.

Dependent variable: *Y*			Number of observations: 964		
Regressor	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
X	7,1 (0,8)	7,0 (1,2)	3,4 (0,1)	3,2 (0,2)	9,9 (2,1)
W	4,2 (0,9)	4,0 (2,3)	3,3 (0,8)	4,1 (1,3)	2,9 (2,1)
Intercept	1,4 (0,2)	1,5 (0,2)	1,3 (0,3)	1,2 (0,3)	1,8 (0,4)
Instrumental variable(s)	$Z^{(1)}$	$Z^{(1)}, Z^{(2)}$	$Z^{(2)}, Z^{(3)}$	$Z^{(1)}, Z^{(3)}$	$Z^{(1)}, Z^{(4)}$
First-stage F-statistic	2.1	14.2	36.7	16.8	6.2
P-value for overidentifying restrictions J-test	—	0.127	0.001	0.009	0.181

Сразу отвергаем первую и последнюю спецификации – в них слабые инструменты.

Тест Саргана для третьей и четвертой спецификаций демонстрирует очень маленькое p-значение. В этих спецификациях есть эндогенные инструменты.

Значит, остается только второй вариант.

6.3.10 Задача 12

В некоторой отрасли промышленности фирмы определяют запасы готовой продукции в зависимости от ожидаемых годовых объемов продаж:

$$y_i = \beta_1 + \beta_2 x_i^e$$

Однако фактический объем продаж для фирмы отличается от плана на случайную величину: $x_i = x_i^e + u_i$, где u_i – i.i.d., $E(u_i) = 0$, $\mathrm{Var}(u_i) = \sigma^2$, $\mathrm{Cov}(x^e, u) = 0$.

Информация об ожиданиях фирм недоступна уважаемому исследователю. Конечно, он мог бы провести опрос, но у него нет уверенности в том, что фирмы предоставят ему честные ответы о своих ожиданиях. Зато в распоряжнии уважаемого исследователя есть данные о фактических объемах продаж.

Он МНК-оценил следующих вопрос.

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

$\boxed{\text{а}}$ Рассмотрим вопрос о свойствах ошибок и состоятельности оценок. Что есть эти ε_i?

Подставим в исходную модель:

$$y_i = \beta_1 + \beta_2 x_i^*$$

$$y_i = \beta_1 + \beta_2 (x_i - u_i)$$

$$y_i = \beta_1 + \beta_2 x_i - \beta_2 u_i$$

Значит, $\varepsilon_i = -\beta_2 u_i$.

$\boxed{\text{б}}$Теперь очевидна коррелированность регрессоров в наблюдаемой модели со случайными ошибками:

$$\mathrm{Cov}(x,\varepsilon) = \mathrm{Cov}(x^e + u_i, \beta_2 u_i) = -\beta_2 \mathrm{Cov}(x^e + u_i, u_i) =$$
$$= -\beta_2 \mathrm{Cov}(x^e, u_i) - \beta_2 \mathrm{Var}(u_i,u_i) = -\beta_2 \sigma^2$$

$$\mathrm{Var}(x) = \mathrm{Var}(x^e + u_i) = \mathrm{Var}(x^e) + \sigma^2$$

На этом состоятельность закончилась:

$$\hat{\beta}_2 \xrightarrow{p} \beta_2 - \frac{\beta_2 \sigma^2}{\mathrm{Var}(x)} = \beta \frac{\mathrm{Var}(x^e)}{\mathrm{Var}(x^e) + \sigma^2}$$

Мы видим, что наша оценка смещена к нулю.

$\boxed{\text{в}}$ Очевидно, что нужно пожелать релевантности и экзогенности инструментальной переменной. Как вариант, объем продаж прошлого периода может сойти.

$$\hat{\beta}^{TSLS} = \frac{\widehat{\mathrm{Cov}}(z,y)}{\widehat{\mathrm{Cov}}(z,x)} = \frac{\widehat{\mathrm{Cov}}(z, \beta_1 + \beta_2 x_i + \varepsilon_i)}{\widehat{\mathrm{Cov}}(z,x)} =$$

$$= \frac{\beta_2 \widehat{\mathrm{Cov}}(z,x) + \widehat{\mathrm{Cov}}(z,\varepsilon)}{\widehat{\mathrm{Cov}}(z,x)} = \beta_2 + \frac{\widehat{\mathrm{Cov}}(z,\varepsilon)}{\widehat{\mathrm{Cov}}(z,x)} \xrightarrow{p} \beta_2$$

6.3.11 Задача 13

Страна Кейнсиания является закрытой экономикой без непосредственного государственного вмешательства, поэтому основное макроэкономическое тождество в ней имеет следующий вид:

$$Y_t = C_t + I_t$$

Потребление выглядит следующим образом:

$$C_t = C_a + mpcY_t + \varepsilon_t$$

ε_t – случайные шоки потребления, которые представляют собой независимые одинаково распределенные случайные величины с нулевым математическим ожиданием и постоянной дисперсией σ^2, случайные шоки потребления не коррелированы с инвестициями. Структурных сдвигов в Кейнсиании не происходит, поэтому автономное потребление и предельная склонность к потреблению неизменны.

Нам известно, что $\widehat{\mathrm{Cov}}(Y,C) = 150$, $\widehat{\mathrm{Var}}(C) = 100$, $\widehat{\mathrm{Var}}(Y) = 300$.

$\boxed{\text{a}}$ Как мы понимаем, $\widehat{mpc}^{OLS} = 0.5 = \dfrac{150}{300}$. Но здесь лежит несостоятельность из-за эндогенности дохода.

Легко покажем это. Вставляем потребление в доход:

$$Y_t = \frac{C_a + I_t + \varepsilon_t}{1 - mpc}$$

Тогда ковариация:

$$\mathrm{Cov}(Y,\varepsilon) = \frac{\sigma^2}{1 - mpc}$$

Значит, получаем, что:

$$\widehat{mpc}^{OLS} = mpc + \frac{\widehat{\mathrm{Cov}}(Y,\varepsilon)}{\widehat{\mathrm{Var}}(Y)} \xrightarrow{p} mpc + \frac{\sigma^2}{(1 - mpc)\,\mathrm{Var}(Y)}$$

$\boxed{\text{б}}$ Найдем теперь 2МНК-оценку. В качестве инструмента используем инвестиции. Тогда:

$$\widehat{mpc}^{TSLS} = \frac{\widehat{\mathrm{Cov}}(I,C)}{\widehat{\mathrm{Cov}}(Y,I)}$$

$$\widehat{mpc}^{TSLS} = \frac{\widehat{\mathrm{Cov}}(I,C)}{\widehat{\mathrm{Cov}}(Y,I)} = \frac{\widehat{\mathrm{Cov}}(I_t, C_a + mpcY_t + \varepsilon_t)}{\widehat{\mathrm{Cov}}(I_t, Y_t)} =$$

$$= \frac{mpc\,\widehat{\mathrm{Cov}}(Y_t, I_t) + \widehat{\mathrm{Cov}}(I_t, \varepsilon_t)}{\widehat{\mathrm{Cov}}(Y_t, I_t)} \xrightarrow{p} mpc$$

Найдем нужное выражение.

$$\widehat{\mathrm{Cov}}(Y,C) = \widehat{\mathrm{Cov}}(C,I) + \widehat{\mathrm{Var}}(C) = \widehat{\mathrm{Cov}}(C,I) + 100 = 150$$

$$\begin{cases} \widehat{\mathrm{Var}}(Y) = \widehat{\mathrm{Var}}(C+I) = \widehat{\mathrm{Var}}(C) + \widehat{\mathrm{Var}}(I) + 2\,\widehat{\mathrm{Cov}}(C,I) = 300 \\ \widehat{\mathrm{Var}}(C) = \widehat{\mathrm{Var}}(Y-I) = \widehat{\mathrm{Var}}(I) + \widehat{\mathrm{Var}}(Y) - 2\,\widehat{\mathrm{Cov}}(Y,I) = 100 \end{cases}$$

$$\begin{cases} \widehat{\mathrm{Var}}(I) + 2\,\widehat{\mathrm{Cov}}(C,I) = 200 \\ \widehat{\mathrm{Var}}(I) - 2\,\widehat{\mathrm{Cov}}(Y,I) = -200 \end{cases}$$

$$\begin{cases} \widehat{\mathrm{Var}}(I) = 100 \\ 100 - 2\,\widehat{\mathrm{Cov}}(Y,I) = -200 \end{cases}$$

$$\begin{cases} \widehat{\mathrm{Var}}(I) = 100 \\ \widehat{\mathrm{Cov}}(Y,I) = 150 \end{cases}$$

$$\widehat{mpc}^{TSLS} = \frac{50}{150} = \frac{1}{3}$$

Глава 7
Семинары 16-21

7.1 Семинар 16-17. Модели бинарного выбора

Вспомним модели бинарного выбора. У нас зависимая переменная имеет простой вид:

$$y_i = \begin{cases} 1, & \text{тогда-то} \\ 0, & \text{иначе} \end{cases}$$

Эта простота игрека говорит, что использование линейной вероятностной модели $y_i = \alpha_0 + \alpha_1 x_i$ быстро кончится проблемой. Мы вместо этого живем с пробитом и логитом:

$$y_i = \Phi(\alpha + \beta x_i)$$

В нормальном случае мы получаем пробит. Тогда:

$$\Phi(x) = \int_{\infty}^{x} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} dt$$

В логистическом случае будет логит. Тогда:

$$\Lambda(x) = \frac{1}{1 + e^{-x}}$$

Конечно, можно любую функцию распределения засунуть, выйдет своя радость.

Здесь МНК не пройдет, надо чуть более умно. Давайте порассуждаем о логите. Мы хотим понять:

$$P(y_i = 1) = \frac{1}{1 + e^{-(\alpha + \beta \cdot x)}}$$

Оценивать, конечно, хочется методом максимального правдоподобия:

$$L = \prod P(y_i = 1) \cdot (1 - P(y_i = 1)) \to \max$$

Дальше берем логарифм, если ручками считать, все получается линейное.

Еще помним связь ошибок: $\varepsilon_i \sim N(0,1)$ превращается в логите в $N(0,1.6^2)$. В итоге возможен переход:

$$\frac{y_i}{1.6} = \beta_1 \frac{1}{1.6} + \beta \frac{x_i}{1.6} + \frac{\varepsilon}{1.6}$$

Еще в мире есть предельные эффекты:

$$\frac{\partial P}{\partial x} = \hat{\beta}_2 f(\hat{\beta}_1 + \hat{\beta}_2 x)$$

7.1.1 Задача 1

Вопросы этого задания основаны на следующем эксперименте: 400 водителей, выбранных случайным образом, попросили пройти специальный тест на вождение автомобилем. Для каждого водителя были собраны следующие данные: Pass — фиктивная переменная, равная единице, если водитель сдал тест, Male — фиктивная переменная, равная единице, если водитель мужчи-на, и равная 0, если водитель женщина, Experience — опыт вождения автомобилем (в годах). В таблице представлены результаты семи моделей, оцененных на основе имеющихся данных.

Dependent Variable: Pass							
	Probit (1)	Logit (2)	Linear Probability (3)	Probit (4)	Logit (5)	Linear Probability (6)	Probit (7)
Experience	0.031 (0.009)	0.040 (0.016)	0.006 (0.002)				0.041 (0.156)
Male				−0.333 (0.161)	−0.622 (0.303)	−0.071 (0.034)	−0.174 (0.259)
Male*Experience							−0.015 (0.019)
Constant	0.712 (0.126)	1.059 (0.221)	0.774 (0.034)	1.282 (0.124)	2.197 (0.242)	0.900 (0.022)	0.806 (0.200)

Вопросы 1 $\boxed{\text{a}}$ Чтобы проверить, зависит ли в первой модели вероятность сдать тест от опыта вождения, надо проверить коэффициент на значимость:

$$t_{\text{расч}} = \frac{0.031}{0.009} = 3.44$$

Это, конечно, значимо на уровне значимости 1%.

$\boxed{\text{б}}$ У Мэттью 10-летний водительский стаж. Какова вероятность того, что он сдаст тест?

У нас пробит-модель: $\Phi(z)$.

$$z = 0.712 + 0.031 \cdot 10 = 1.022$$

$$\Phi(1.022) = 0.847$$

$\boxed{\text{в}}$ Кристофер новичок, его водительский опыт равен нулю. Какова вероятность того, что он сдаст тест? Очевидно, что:

$$\Phi(0.712) = 0.762$$

$\boxed{\text{г}}$ Выборка включает значения переменной Experience от 0 до 40, при этом только 4 наблюдения превышают 30. Джеду сейчас 95 лет и он начал водить с 15 лет. Какова прогнозируемая моделью вероятность сдачи теста Джедом? Этот прогноз адекватен?

$$z = 0.712 + 0.031 \cdot 80 = 3.192$$

$$\Phi(3.192) = 0.999$$

Не очень, монотонное возрастание – сомнительная идея.

Вопросы 2 $\boxed{\text{а}}$ Чтобы проверить, зависит ли в первой модели вероятность сдать тест от опыта вождения в логит-модели, надо проверить коэффициент на значимость:

$$t_{\text{расч}} = \frac{0.040}{0.016} = 2.5$$

Это очевидно значимо на уровне значимости 5%.

$\boxed{\text{б}}$ У Мэттью 10-летний водительский стаж. Какова вероятность того, что он сдаст тест?

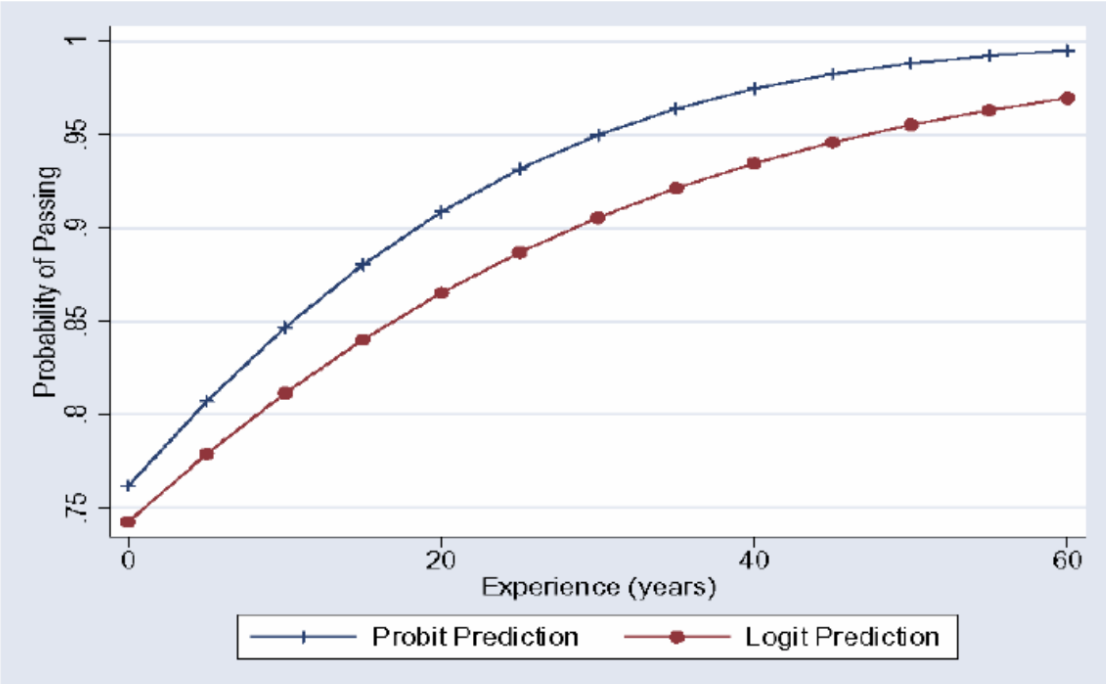
У нас логит-модель:

$$P = \frac{1}{1 + e^{-(1.59+0.040\cdot 10)}} = \frac{1}{1 + e^{-1.459}} = 0.811$$

$\boxed{\text{в}}$ Кристофер новичок, его водительский опыт равен нулю. Какова вероятность того, что он сдаст тест? Очевидно, что:

$$P = \frac{1}{1 + e^{1.059+0.040\cdot 0}} = 0.742$$

$\boxed{\text{г}}$ Нарисуем на похищенном графике:



Вопросы 3 $\boxed{\text{а}}$ Теперь ЛВМ:

$$t_{\text{расч}} = \frac{0.006}{0.002} = 3$$

Это очевидно значимо на уровне значимости 1%.

$\boxed{\text{б}}$ У Мэттью 10-летний водительский стаж. Какова вероятность того, что он сдаст тест?

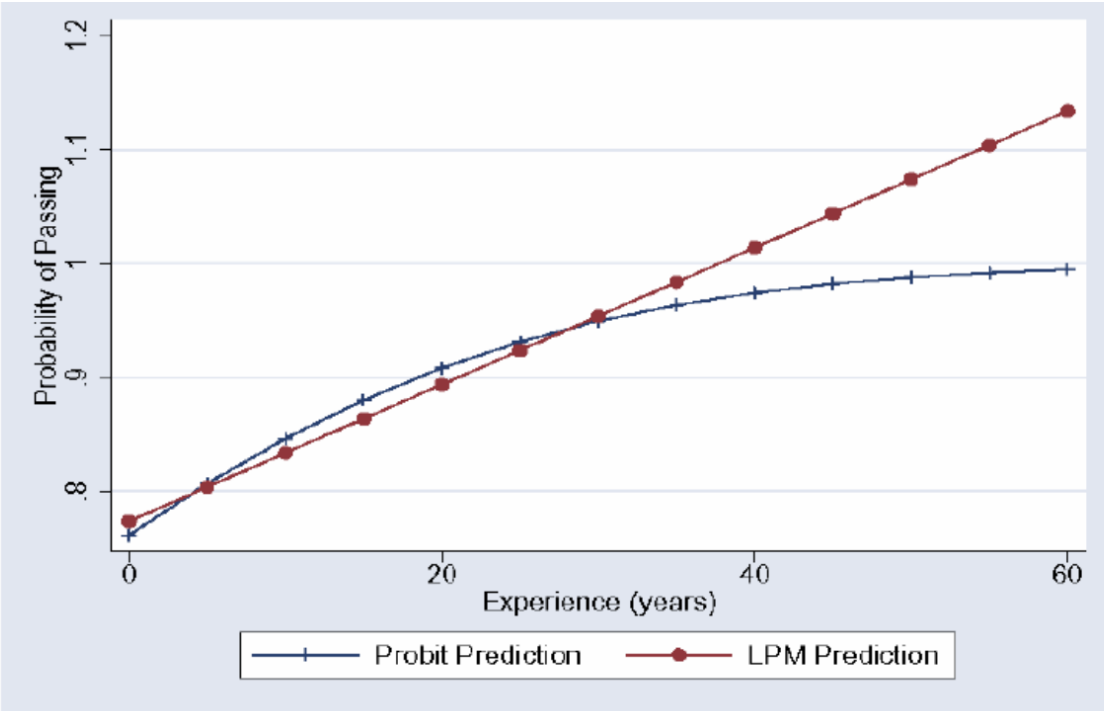
У нас ЛВ-модель:

$$P = 0.774 + 0.006 \cdot 10 = 0.836$$

$\boxed{\text{в}}$ Кристофер новичок, его водительский опыт равен нулю. Какова вероятность того, что он сдаст тест? Очевидно, что:

$$P = 0.774 + 0.006 \cdot 0 = 0.774$$

$\boxed{\text{г}}$ Нарисуем на похищенном графике:



Вопросы 4 $\boxed{\text{а}}$ Используем каждую из моделей, вычислим оцененную вероятность сдать экзамен для мужчины и для женщины.

Сначала считаем женщин:

- Пробит: $\Phi(1.282) = 0.9$
- Логит:
$$\frac{e^{2.197}}{1 + e^{2.197}} = 0.9$$
- ЛВМ: Ясно, что 0.9.

Сначала считаем мужчин:

- Пробит: $\Phi(1.282 - 0.3333) = 0.829$

- Логит:

$$\frac{e^{2.197-0.622}}{1+e^{2.197-0.622}} = 0.829$$

- ЛВМ: Ясно, что 0.829.

Вопросы 5 $\boxed{\text{а}}$ Акира — мужчина с 10-летним водительским опытом. Какова вероятность того, что он сдаст экзамен?

Просто подставляем везде сразу:

$$\Phi(0.806 + 0.041 \cdot 10 \cdot 0.174 - 0.015 \cdot 1 \cdot 10) = 0.814$$

$\boxed{\text{б}}$ Джейн — женщина с 2-летним водительским опытом. Какова вероятность того, что она сдаст экзамен?

$$\Phi(0.806 + 0.041 \cdot 2 - 0.174 \cdot 0 - 0) = 0.813$$

$\boxed{\text{в}}$ Влияние опыта вождения на успешность сдачи экзамена зависит от пола? Обоснуйте свой ответ

$$t_{\text{расч}} = -\frac{0.015}{0.019} = -0.79$$

Решительно нет значимости.

7.1.2 Задача 2

Уважаемого исследователя интересует, как вероятность сдать зачет зависит от времени, затраченного на подготовку. Пусть y_i – бинарная переменная, которая равна единице, если i-ый студент сдал зачет, и равна нулю в противном случае, а x_i – количество часов, затраченное на подготовку к зачету i-ым студентом. Исследователь располагает данными о 200 студентах.

Первые сто не сдали, вторые сто сдали. При этом суммарно первые сто готовились 500 часов, а вторые – 1500 часов. Сумма квадратов часов подготовки этих т.н. студентов равно 26 250.

$\boxed{\text{а}}$ МНК-оценим ЛВМ:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$\hat{\beta}_1 = \frac{\overline{xy} - \overline{x}\overline{y}}{\overline{x^2} - \overline{x}^2}$$

Подставляем числа:

$$\hat{\beta}_1 = \frac{7.5 - 10 \cdot 0.5}{131.25 - 100} = \frac{2.5}{31.25} = 0.08$$

Находим и константу:

$$\hat{\beta}_0 = 0.5 - 0.08 \cdot 10 = -0.3$$

Можно долго и тоскливо считать ESS, раскладывая все суммы на куски, но легче через корреляцию подойти:

$$R^2 = \widehat{\mathrm{Corr}}^2(x,y) = \frac{2.5^2}{31.25 \cdot 0.25} = 0.8$$

Осталось понять, каков шанс сдать зачет у Пети Лалочкина, готовившегося к нему жалкие 12 часов:

$$-0.3 + 12 \cdot 0.08 = 0.66$$

$\boxed{\text{б}}$ Теперь уважаемый исследователь, используя те же самые данные, оценил параметры логит-модели. Результаты оценивания представлены в таблице. Какова в соответствии с новой моделью вероятность сдать зачет для Пети Лалочкина?

У нас есть оцененная логит-модель:

$$\hat{p}(y_i = 1) = \frac{1}{1 + e^{15 - 1.5x}}$$

Модель 1: Логит, использованы наблюдения 1-100

Зависимая переменная: y

Стандартные ошибки рассчитаны на основе Гессиана

	Коэффициент	*Ст. ошибка*
const	−15,0	2,1
x	1,50	0,1

В нее осталось подставить значение для Пети:

$$\hat{p} = \frac{1}{1 + e^{-3}} = 0.95$$

7.1.3 Задача 3

100 кандидатов проходили собеседование о приеме на работу в крупную компанию. Известны следующие данные о кандидатах:

1. x – стаж работы кандидата;
2. $gender$ – бинарная переменная, равная единице для кандидатов-мужчин и равная нулю для кандидатов-женщин;
3. $black$ – переменная с аналогичным принципом;
4. y – бинарная переменная, равна единице для кандидатов, принятых на работу.

Результаты оценивания трех моделей:

Зависимая переменная: y

Метод оценивания: логит-модель

	Модель 1	Модель 2	Модель 3
x	—	0,49 (0,30)	0,49 (0,15)
gender	—	—	0,15 (0,43)
black	—	—	−0,32 (0,43)
constant	−0,32 (0,20)	−1,02 (0,15)	−0,90 (0,42)
Логарифм функции правдоподобия	−68,0	−62,0	−61,0
R-квадрат Макфаддена			

Об опыте и итогах собеседования:

Стаж работника	Сколько кандидатов имеет такой стаж работы	Сколько из них было принято на работу в компанию
0 лет	40	10
1 год	25	10
2 года	10	5
3 года	10	6
4 года	10	7
5 лет	5	4

$\boxed{\text{а}}$ Посчитаем сначала R^2 Макфаддена для всех моделей:

$$R_M^2 = pseudoR^2 = 1 - \frac{\ln L}{\ln L_0}$$

Здесь L_0 – модель только с константой. Чем показатель выше, тем лучше модель.

Просто используем данные из таблицы.

$\boxed{\text{б}}$ Теперь проверим значимость всего уравнения в целом третьей модели, используя тест правдопободия.

Запишем общую гипотезу:

$$H_0 : H\beta = r$$

$$H_1 : \text{ не } H_0$$

Тестовая статистика при q ограничениях:

$$LR = -2(\ln L_R - \ln L_{UR}) \sim \chi^2(q)$$

В нашем случае это все имеет вид:

$$H_0 : \forall i \quad \beta_i = 0$$

$$H_1 : \text{ не } H_0$$

Считаем:

$$LR = -2(61 - 68) = 14$$

$\chi^2_{\text{кр}}(3) = 7.81$. Значит, наше уравнение значимо.

$\boxed{\text{в}}$ Сравним теперь модели (2) и (3), используя тест отношения правдопободия.

$$H_0 : \beta_2 = \beta_3 = 0$$
$$H_1 : \text{ не } H_0$$

Считаем:

$$LR = -2(62 - 61) = 2$$

$\chi^2_{\text{кр}}(2) = 6.99$. Значит, мы принимаем нулевую гипотезу.

$\boxed{\text{г}}$ Найдем теперь предельный эффект стажа работы для среднего по выборке работника, используя вторую модель.

Можно же в лоб посчитать средний опыт:

$$\bar{x} = \frac{1}{100}(0 \cdot 40 + 25 \cdot 1 + 10 \cdot 2 + 10 \cdot 3 + 10 \cdot 4 + 5 \cdot 5) = 1.4$$

Считаем для второй модели:

$$p_i = \frac{1}{1 + e^{1.02 - 0.49x_i}}$$

Наш предельный эффект:

$$\frac{\partial p}{\partial x} = \frac{e^{-z}}{(1 + e^{-z})^2}\beta_1$$

Находим теперь $\overline{z} = -1.02 + 0.49 \cdot 1.4 = -0.334$. Наконец, просто считаем:

$$\frac{\partial p}{\partial x} = \frac{e^{0.334}}{(1 + e^{0.334})^2} \cdot 0.49 = 0.12$$

В среднем для кандидата с таким опытом работы вероятности найти работу с ростом опыта на 1 год увеличивается на 12%.

$\boxed{\text{д}}$ А теперь вычислим средний предельный эффект стажа работы.

Считаем для $x = 0, 1 \ldots 5$.

$$x = 0,\ \frac{\partial p}{\partial x} = 0.0954$$
$$x = 1,\ \frac{\partial p}{\partial x} = 0.1143$$
$$x = 2,\ \frac{\partial p}{\partial x} = 0.1225$$

$$x = 3,\ \frac{\partial p}{\partial x} = 0.1165$$
$$x = 4,\ \frac{\partial p}{\partial x} = 0.0990$$
$$x = 5,\ \frac{\partial p}{\partial x} = 0.0764$$

Легко посчитать, что средний предельный эффект – 0.104 – это средний по жизни. У среднего кандидата 0.119.

Это значит, что в среднем 1 дополнительный год стажа увеличивает вероятность принятия на работу на 10.43%.

7.1.4 Задача 4

Оценим логит по данным

	$y = 0$	$y = 1$
$x = 0$	20	32
$x = 1$	36	12

Все просто в мире. Берем и оцениваем.

Нас интересует модель вида $p(y_i = 1) = F(\alpha+\beta x_i)$. В нее можно решительно подставить x и разложить на кусочки:

$$L(y_1 \cdot y_{100}) = p(y_1 \dots y_{100}|x_1 \dots x_{100}, \alpha, \beta) = \prod p(y_i) =$$
$$= \prod p(y_i = 1) \cdot \prod p(y_i = 0) = F(\alpha + \beta)^{12}(1 - F(\alpha + \beta))^{36} F(\alpha)^{32}(1 - F(\alpha))^{20}$$

Берем теперь логарифм:

$$\ln L = 32 \ln F(\alpha) + 12 \ln F(\alpha + \beta) + 20 \ln(1 - F(\alpha)) + 36 \ln(1 - F(\alpha + \beta))$$

Выписываем условия первого порядка:

$$\frac{\partial \ln L}{\partial \alpha} = 32 \frac{f(\alpha)}{F(\alpha)} - 20 \frac{f(\alpha)}{1 - F(\alpha)} = 0$$
$$\frac{\partial \ln L}{\partial (\alpha + \beta)} = 12 \frac{f(\alpha + \beta)}{F(\alpha + \beta)} - 36 \frac{f(\alpha + \beta)}{1 - F(\alpha + \beta)} = 0$$

Помним, что $F'(\cdot) = f(\cdot)$. Чтобы решить эту систему, просто убираем числители. В итоге выходит, что $F(\alpha) = \frac{8}{13}$, а $F(\alpha + \beta) = \frac{1}{4}$.

Сделаем теперь до конца для логит-модели:

$$\frac{1}{1+e^{-\alpha}} = \frac{8}{13}$$

$$\alpha = -\ln\left(\frac{5}{8}\right) = 0.47$$

$$\frac{1}{1+e^{-\alpha-\beta}} = \frac{1}{4}$$

$$\beta = \ln\left(\frac{5}{8}\right) - \ln(3) = -1.5686$$

Для пробит-модели все аналогично, но руками считать становится тяжко.

7.1.5 Задача 5

Уважаемый исследователь использует логит-модель бинарного выбора для того, чтобы выяснить, как вероятность оказаться безработным зависит от опыта работы и образования. Исследователь опросил 1000 экономических активных граждан в возрасте от 21 до 28 лет и получил данные о следующих переменных: Unemployed — бинарная переменная равная единице, если респондент является безработным; Experience — опыт работы респондента (в годах); Education — продолжительность обучения респондента (в годах).

В таблице представлены результаты оценивания модели:

Dependent Variable: Unemployed

	Logit
Experience	– 0.20 (0.03)
Education	– 0.10 (0.02)
Constant	– 0.60 (0.12)

$\boxed{\text{а}}$ Аристрах Татьянович Мышь 10 лет учился в школе и еще 4 года в бакалавриате. Опыта работы у него пока нет. С какой вероятностью он окажется безработным?

Элементарно подставим его в логит:

$$\hat{p} = \frac{1}{1 + e^{0.6+0.2\cdot 0+0.1\cdot 14}} = 0.12$$

$\boxed{\text{б}}$ Посмотрим, на сколько процентных пунктов снизит для Аристраха вероятность оказаться безработным один дополнительный год опыта работы.

$$\frac{\partial p}{\partial \beta_{exp}} = -0.2 \frac{e^{0.6+0.2\cdot 0+0.1\cdot 14}}{(1 + e^{0.6+0.2\cdot 0+0.1\cdot 14})^2} = 0.0209$$

7.2 Семинар 18-20. Панельные данные

Вспомним суть панельных данных:

- Статические модели
  - Pooled: здесь сразу возникает гора наблюдений без особых размышлений:

$$y_i = \beta x_i + \varepsilon_i$$

  - Модели с фиксированными эффектами: есть какие-то фиксированные про времени или по объектам эффекты:

$$y_{it} = \beta x_{it} + \mu_i + \varepsilon_{it}$$

$$y_{it} = \beta x_{it} + \zeta_t + \varepsilon_{it}$$

    Бывают и двунаправленные версии.

  - Модели со случайными эффектами. Главное разница в том, что отличия здесь случайны.

$$y_{it} = \beta x_{it} + u_i + \varepsilon_{it}$$

$$y_{it} = \beta x_{it} + v_t + \varepsilon_{it}$$

    Здесь тоже есть двунаправленные варианты есть, где эффекты фиксируются и по времени, и по пространству.

- Динамические модели

О фиксированных эффектах Можно работать в первых разностях. Рассмотрим для двух периодов:

$$y_{it} = \beta_0 + \sum \beta_j x_{it}^{(j)} + \gamma z_i + \varepsilon_{it}$$

$$y_{it-1} = \beta_0 + \sum \beta_j x_{it-1}^{(j)} + \gamma z_i + \varepsilon_{it-1}$$

Берем разницу:

$$\Delta y_{it} = \sum \beta_j \Delta x_{it}^{(j)} + \Delta\varepsilon_{it}$$

Важный факт – в иксах только то, что меняется.

LSDV-модель Можно поступить иначе. Запишем исходную регрессию:

$$y_{it} = \beta_0 + \sum \beta_j x_{it}^{(j)} + \gamma z_i + \varepsilon_{it}$$

Создаем n фиктивных переменных:

$$D1i = \begin{cases} 1, & \text{для 1-ого объекта} \\ 0, else \end{cases}$$

$$Dni = \begin{cases} 1, & \text{для n-ого объекта} \\ 0, else \end{cases}$$

В итоге модель размазывается:

$$y_{it} = \sum \beta_j x_{it}^{(j)} + \alpha_1 D1_i + \ldots \alpha_n Dn_i + \varepsilon_{it}$$

Тогда мы выкинули константу заменой: $\alpha_1 = \beta_0 + \gamma z_1$. Можно и вернуть β_0, но тогда надо выкинуть одну фиктивную.

Внутригрупповое преобразование Еще можно третьим способом.

$$y_{it} = \beta_0 + \beta_1 x_{it} + \gamma z_i + \varepsilon_{it}$$

Просто складываем все по времени, а потом берем среднее:

$$\sum_{t=1}^{T} y_{it} = T\beta_0 + \beta_1 \sum_{t=1}^{T} x_{it} + T\gamma z_i + \sum_{t=1}^{T} \varepsilon_{it}$$

В итоге просто вычитаем из каждого наблюдедение среднее и переобозначаем:

$$y_{it} - \overline{y}_i = \beta_1 (x_{it} - \overline{x}_i) + (\varepsilon_{it} - \overline{\varepsilon}_i)$$

$$b_{it} = \beta_1 + c_{it} + u_{it}$$

Сравнение pooled и фиксированных эффектов

$$H_0: \text{ Индивидуальных эффектов нет}$$

Чтобы сравнить, оцениваем сначала LSDV-модель с $D1_i \dots Dn_i$. Делаем тест на короткую и длинную регрессию, проверяя, равны ли нулю коэффициенты. И все.

Сравнение FE и RE Чтобы сравнить модели с фиксированными и случайными эффектами, делаем тест Хаусмана. Если $\text{Cov}(u_{i_0}, x_{i_0 t_0}) = 0$, то оценка со случайными эффектами состоятельна.

$$H_0 : \text{Cov}(u_{i_0}, x_{i_0 t_0}) = 0$$
$$H_1 : \text{Cov}(u_{i_0}, x_{i_0 t_0}) \neq 0$$

Если мы приняли нулевую гипотезу, то используем модель со случайными эффектами, потому что оценки становятся состоятельными и эффективными. Иначе берем фиксированные.

Сравнение RE и pooled

$$H_0: \text{ Случайных эффектов нет: } \text{Var}(u_i) = \sigma_u^2 = 0$$

Если мы приняли нулевую гипотезу, то используем pooled регрессию. Для тестирования это гипотезы используем тест множителей Лагранжа.

7.2.1 Задача 1

Мы имеем модель уважаемого исследователя Э. Старка:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \gamma z_i + \varepsilon_{it}$$

Уважаемый исследователь изучает влияние располагаемого дохода на потребление. При этом мы предполагаем фиксированные эффекты.

$\boxed{\text{а}}$ С помощью внутригруппового преобразования найдем оценку для β_1:

$$\sum y_{it} = 2\beta_0 + \beta_1 \sum x_{it} + 2\gamma z_i + \sum \varepsilon_{it}$$

$$\overline{y}_i = \beta_0 + \beta_1 \overline{x}_i + \gamma z_i + \overline{\varepsilon_i}$$

Значит,

$$(y_{it} - \overline{y}_i) = \beta_1(x_{it} - \overline{x}_i) + (\varepsilon_{it} - \overline{\varepsilon}_i)$$

В итоге:

$$\tilde{y}_{it} = \beta_1 \tilde{x}_{it} + \tilde{\varepsilon}_{it}$$

Оценка будет:

$$\hat{\beta}_1 = \frac{\sum \tilde{x}_{it}\tilde{y}_{it}}{\sum (\tilde{x}_{it})^2}$$

Рисуем табличку для первого года:

Регион	$\tilde{y}_{it}$	$\tilde{x}_{it}$	$\tilde{x}_{it}\tilde{y}_{it}$	$(\tilde{x}_{it})^2$
Винтерфелл	0	0	0	0
Риверран	2	0	0	0
Королевская гавань	3	5	15	25
Дорн	0	0	0	0
Хайгарден	6	10	60	100
$\sum$			75	125

Для второго года:

Регион	$\tilde{y}_{it}$	$\tilde{x}_{it}$	$\tilde{x}_{it}\tilde{y}_{it}$	$(\tilde{x}_{it})^2$
Винтерфелл	0	0	0	0
Риверран	-2	0	0	0
Королевская гавань	-3	-5	15	25
Дорн	0	0	0	0
Хайгарден	-6	-10	60	100
$\sum$			75	125

Считаем чудесную оценку:

$$\hat{\beta}_1 = \frac{2 \cdot 75}{2 \cdot 125} = 0.6$$

$\boxed{\text{б}}$ Найдем теперь оценку в первых разностях. Пишем модели:

$$y_{i2} = \beta_0 + \beta x_{i2} + \gamma z_i + \varepsilon_{i2}$$
$$y_{i1} = \beta_0 + \beta x_{i1} + \gamma z_i + \varepsilon_{i1}$$

Вычитаем и получаем:

$$\Delta y_{i2} = \beta_1 \Delta x_{i2} + \Delta \varepsilon_{i2}$$

Значит,

$$\hat{\beta}_1 = \frac{\sum \Delta x_{i2} \cdot \Delta y_{i2}}{\sum (\Delta x_{i2})^2}$$

Регион	Δy_{i2}	Δx_{i2}	$\Delta x_{i2} \Delta y_{i2}$	$(\Delta x_{i2})^2$
Винтерфелл	0	0	0	0
Риверран	-4	0	0	0
Королевская гавань	-6	-10	60	100
Дорн	0	0	0	0
Хайгарден	-12	-20	240	400
$\sum$			300	500

Значит,

$$\hat{\beta}_1 = \frac{\sum \Delta x_{i2} \cdot y_{i2}}{\sum (\Delta x_{i2})^2} = \frac{300}{500} = 0.6$$

$\boxed{\text{в}}$ Уважаемый исследователь С. Баратеон предлагает нам включить в модель еще тренд, потому что он верит, что автономное потребление тоже меняется:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 t + \gamma z_i + \varepsilon_{it}$$

Оценим это тоже через первые разности:

$$y_{i2} = \beta_0 + \beta x_{i2} + 2\beta_2 + \gamma z_i + \varepsilon_{i2}$$
$$y_{i1} = \beta_0 + \beta x_{i1} + \beta_2 + \gamma z_i + \varepsilon_{i1}$$

Вычитаем и получаем:

$$\Delta y_i 2 = \beta_1 \Delta x_{i2} + \beta_2 + \Delta \varepsilon_{i2}$$

Тут уже совсем очевидная парная регрессия:

$$\Delta y_{i2} = \beta_2 + \beta_1 \Delta x_{i2} + \Delta\varepsilon_{i2}$$

Получаем оценки:

$$\hat{\beta}_1 = \frac{\overline{\Delta y_{i2} \cdot \Delta x_{i2}}}{\overline{\Delta x_{i2}^2} - \left(\overline{\Delta x_{i2}}\right)^2}$$

Регион	Δy_{i2}	Δx_{i2}	$\Delta x_{i2}\Delta y_{i2}$	$(\Delta x_{i2})^2$
Винтерфелл	0	0	0	0
Риверран	-4	0	0	0
Королевская гавань	-6	-10	60	100
Дорн	0	0	0	0
Хайгарден	-12	-20	240	400
$\sum$	-22	-30	300	500

Подставляем все в формулу:

$$\hat{\beta}_1 = \frac{\frac{300}{5} - \frac{22}{5} \cdot \frac{30}{5}}{\frac{500}{5} - \left(\frac{30}{5}\right)^2} = 0.525$$

$$\hat{\beta}_2 = \overline{\Delta y_{it}} - \hat{\beta}_1 \overline{\Delta x_{i2}} = -\frac{22}{5} + 0.525 \cdot \frac{30}{5} = -1.25$$

В итоге модель:

$$\Delta_{i2} = -1.25 + 0.525 \Delta x_{i2}$$

$\boxed{\text{г}}$ Уважаемый исследователь Т. Ланнистер решил сделать pooled регрессию:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \varepsilon_{it}$$

Фиксированные эффекты представлены LSDV-моделью:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \alpha_2 du2_i + \dots + \alpha_5 du5_i + \varepsilon_{it}$$

Сравним два подхода по данным:

$$H_0 : \boldsymbol{\alpha} = 0$$

$$H_1: \text{ иначе}$$

Пишем нашу $F_{\text{расч}}$.

$$F_{\text{расч}} = \frac{(R^2_{UR} - R^2_R)/4}{(1 - R^2_{UR})/(10 - 6)}$$

$$F_{\text{расч}} = \frac{(0.981802 - 0.964390)/4}{(1 - 0.981802)/4} = 0.96$$

Помним, что $F_{\text{кр}}(4, 4{,}0.05) = 6.388$. Значит, принимаем нулевую гипотезу. Можно спокойно пользоваться pooled моделью, а значимых различий нет.

7.2.2 Задача 2

У уважаемых нас есть данные о переменных x и y в трех различных странах за 2011–2013 гг.:

2011 год		
	y	**x**
Страна A	−3	0
Страна B	0	3
Страна C	27	12

2012 год		
	y	**x**
Страна A	−3	0
Страна B	12	3
Страна C	27	12

2013 год		
	y	**x**
Страна A	3	3
Страна B	18	9
Страна C	33	15

Надо оценить модель с фиксированными эффектами. Делаем простейшеее внутригрупповое преобразование.

i	t	y	x	$\tilde{y}$	$\tilde{x}$	$\tilde{x}\tilde{y}$	$\tilde{x}^2$
1	1	-3	0	-2	-1	2	1
1	2	-3	0	-2	-1	2	1
1	3	3	3	4	2	8	4
2	1	0	3	-10	-2	20	4
2	2	12	3	2	-2	-4	4
2	3	18	9	8	4	32	16
3	1	27	12	-2	-1	2	1
3	2	27	12	-2	-1	2	1
3	3	33	15	4	2	8	4
						72	36

Ясно, что наше преобразование убивает константу. Остается просто посчитать простым МНК:

$$\hat{\beta}_2 = \frac{\sum\sum \tilde{x}_i\tilde{y}_i}{\sum\sum \tilde{x}^2} = \frac{72}{36}$$

7.3 Семинар 21. Динамические панели. Эксперименты

Иногда в жизни становится грустно от эндогенностей и прочего. Тогда на помощь нам приходят динамические панели.

$$y_{it} = \beta_0 + \beta_1 y_{it-1} + \gamma z_i + \varepsilon_{it}$$

Если мы вычтем из y_{it} y_{it-1}, то заметим эндогенность. Поэтому приходится использовать инструменты вида y_0 для разности $y_{i2} - y_{i1}$, y_0, y_1 для разности $y_{i3} - y_{i2}$, т.д. Очевидно из таких разностей:

$$y_{it} - y_{it-1} = \beta_1(y_{it-1} - y_{it-2}) + (\varepsilon_{it} - \varepsilon_{it-1})$$

Здесь будет уже $t-1$ инструмент. Релевантность и экзогенность в рамках предпосылки вроде понятны. Остается только следить за экзогенностью.

7.3.1 Задача 1

Сразу отвергаем первую модель в пользу второй, основываясь на тесте на наличие индивидуальных эффектов.

Осталось понять, что в третьей модели все лучше. Тест Саргана там хорош, нулевую гипотезу мы не отклоняем. Смотрим теперь на автокорреляцию первого порядка:

$$\varepsilon_{it} = \rho \varepsilon_{it} + u_{it}$$

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

Принимаем альтернативную гипотезу, автокорреляция первого порядка.

Смотрим теперь второго порядка:

$$H_0 : \text{ автокорреляции второго порядка нет}$$

Здесь $p > 0.01$. Значит, нет у нас автокорреляции второго порядка. Верно специфицировали.

Регрессоры	(1) Обычный МНК	(2) Модель с фиксированными эффектами	(3) Обобщенный метод моментов (Аррелано-Бонд)
Темп прироста реального ВВП с лагом в 1 год	0.552*** (0.059)	0.272*** (0.067)	0.304*** (0.065)
Таргетирование инфляции с лагом в 1 год	1.076 (2.448)	0.901 (2.166)	2.637 (3.725)
Таргетирование инфляции с лагом в 2 года	-1.375 (2.510)	-0.624 (1.783)	-0.559 (1.523)
Таргетирование инфляции с лагом в 3 года	0.358 (0.384)	0.803* (0.424)	1.699*** (0.516)
Смена главы ЦБ	-0.178 (0.351)	-0.0679 (0.215)	-0.292 (0.242)
Прирост населения	-7.017 (17.98)	9.866 (38.52)	28.44 (80.53)
Константа	1.485*** (0.391)	1.535*** (0.413)	
Страновые эффекты	Нет	Да	Да
Фиктивные переменные времени	Да	Да	Да
Количество наблюдений	422	422	418
Число инструментов	—	—	31
R-квадрат	0.383	0.248	—
P-value теста на отсутствие индивидуальных эффектов	—	0,00	—
P-value теста Ареллано-Бонда для первых разностей переменных AR(1)			0.00
AR(2)			0.03
P-value теста Саргана на сверхидентифицирующие ограничения			0.32

7.3.2 Задача 2

Подумаем о treatment эффекте. В хорошем случае мы бы хотели знать $y_i(1) - y_i(0)$, из этого получить $E(y(1) - y(0) = ATE$. Но в жизни часто случаются всякие неблагоприятные отборы, искривления: здоровые не пьют лекарства, больные пьют.

С этим как-то бороться. Пусть $D_i = 1$ для принимавших, $D_i = 0$ для не принимавших здоровых.

$$\mathrm{E}(y_i(1)|D_i = 1) - \mathrm{E}(y_i(0)|D_i = 0) =$$

$$= \mathrm{E}(y_i(1)|D_i = 1) - \mathrm{E}(y_i(0)|D_i = 1) + \mathrm{E}(y_i(0)|D_i = 1) - \mathrm{E}(y_i(0)|D_i = 0) =$$

$$= \mathrm{E}(y_i(1) - y_i(0)|D_i = 1) + \mathrm{E}(y_i(0)|D = 1) - \mathrm{E}(y_i(0)|D = 0)$$

Мы разложили на ATE и два члена смещения. Все подходы идут к тому, что члены смещения должны быть равными. Это куда-то на тему экспериментов еще.

Все это было преамбулой к задаче. Мы делим наши регионы на treatment и control group.

$$y_{it} = \mu_i + \lambda_t + \delta D_{it} + \varepsilon_{it}$$

Ясно, что:

$$\mathrm{E}(y_{it}|treatm{,}14) = \mu_{tr} + \lambda_{14} + \delta$$

$$\mathrm{E}(y_{it}|treatm{,}13) = \mu_{tr} + \lambda_{14}$$

Для контрольной группы:

$$\mathrm{E}(y_{it}|contr{,}14) = \mu_{cont} + \lambda_{14}$$

$$\mathrm{E}(y_{it}|cont{,}13) = \mu_{cont} + \lambda_{13}$$

Дальше вычитаем:

$$\Delta_{tr} = \mathrm{E}(y_{it}|treatm{,}14) - \mathrm{E}(y_{it}|treatm{,}13) = \lambda_{14} - \lambda_{13} + \delta$$

$$\Delta_{cont} = \mathrm{E}(y_{it}|contr{,}14) - \mathrm{E}(y_{it}|cont{,}13) = \lambda_{14} - \lambda_{13}$$

В итоге:

$$\Delta_{tr} = \Delta_{contr} + \delta$$

$$\delta = \Delta_{tr} - \Delta_{contr}$$

Методом разности разностей мы нашли нашу дельту. Осталось посчитать. Вместо ожиданий просто берем выборочные средние.

$$\delta = \overline{y}_{tr,14} - \overline{y}_{tr,13} - \overline{y}_{cont,14} + \overline{y}_{cont,13}$$

В итоге:

$$\hat{\delta} = 8 - 10 - 7.6 + 12 = 2.4\%$$

В результате введения закона уровень безработацы вырос на 2.4 п.п.

$\boxed{\text{б}}$ Теперь надо бы МНК-оценить в жизни:

$$y_{it} = \beta_0 + \beta_1 x_i + \beta_2 z_t + \delta x_i z_t + \varepsilon_{it}$$

В этом случае ясно, что:

$$E(y_{it}|treatm{,}14) = \beta_0 + \beta_1 + \beta_2 + \delta$$

$$E(y_{it}|treatm{,}13) = \beta_0 + \beta_1$$

Для контрольной группы:

$$E(y_{it}|contr{,}14) = \beta_0 + \beta_2$$

$$E(y_{it}|cont{,}13) = \beta_0$$

Помним свои средние:

$$\overline{y}_{tr,14} = 8, \overline{y}_{tr,13} = 10, \overline{y}_{cont,14} = 7.6, \overline{y}_{cont,13} = 12$$

Довольно очевидно, что $\hat{\beta}_0 = 12$. Теперь ищем $\hat{\beta}_1$.

$$\beta_1 = E(y_{it}|treatm{,}13) - E(y_{it}|cont{,}13)$$

$$\hat{\beta}_1 = 10 - 12 = -2$$

Аналогично:

$$\beta_2 = E(y_{it}|contr{,}14) - E(y_{it}|cont{,}13)$$

$$\hat{\beta}_2 = -4.4$$

Осталась δ:

$$\delta = E(y_{it}|treatm{,}14) - E(y_{it}|treatm{,}13) - E(y_{it}|contr{,}14) + E(y_{it}|cont{,}13)$$

$$\hat{\delta} = 2.4$$

В итоге:

$$\hat{y}_{it} = 12 - 2x_i - 4.4z_t + 2.4x_i z_t$$

$\boxed{\text{в}}$ Проверим теперь значимость коэффициентов:

- $\hat{\beta}_0$:

$$t_{\text{расч}} = \frac{12}{0.4} = 30$$

- $\hat{\beta}_1$:

$$t_{\text{расч}} = \frac{-2}{0.53} = -3.8$$

- $\hat{\beta}_2$:

$$t_{\text{расч}} = \frac{-4.4}{0.61} = -7.21$$

- δ:

$$t_{\text{расч}} = \frac{2.4}{0.93} = 2.6$$

Посмотрим критическое значение: $t_{\text{кр}} = t(n \cdot T - k, \alpha) = t(9 \cdot 2 - 4{,}0.05) = t(14, 0.05) = 1.76$.

В среднем у регионов, которые воплотили закон о зарплате, уровень безработицы был ниже на 2 п.п.. Для применивших закон уровень безработицы выше на 2.4 п.п.. Со сменой периода безработица у тех, кто не применил, снизилась на 4.4 п.п. Все значимо.

Глава 8
Семинары 22-25

8.1 Семинар 22. Стационарность, автокорреляции

Вспомним, когда стационарен процесс $AR(p)$:

$$y_t = \delta + \theta y_{t-1} + \dots + \theta_p y_{t-p} + \varepsilon_t$$

$$(1 - \theta_1 L - \dots \theta_p L^p) y_t = \delta + \varepsilon$$

Получаем характеристическое уравнение:

$$1 - \theta + 1z - \dots - \theta_p z^p = 0$$

Если все корни по модулю больше единицы, то процесс стационарен.

Процесс $MA(q)$ конечного порядка стационарен по построению, все моменты у него постоянные:

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

$ARMA(p,q)$ тоже надо проверять, но только AR часть:

$$y_t = \delta + \theta y_{t-1} + \dots + \theta_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

$ARIMA(p,k,q)$ чуть хитрее. Если $k = 0$, то y_t стационарен; если $k = 1$, то первые разности Δy_t стационарны.

8.1.1 Задача 1

Посмотрим процесс:

$$y_t = 0.8 y_{t-1} + \varepsilon_t$$

Чтобы построить коррелограмму, нужно считать коэффициенты вида:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\mathrm{Cov}(y_t, y_{t-k})}{\mathrm{Var}(y_t)}$$

$\boxed{\text{а}}$ Сначала проверим стационарность:

$$y_t - 0.8y_{t-1} = \varepsilon_t$$

$$y_t(1 - 0.8L) = \varepsilon_t$$

$$1 - 0.8z = 0$$

$$z = \frac{10}{8} > 1$$

Получаем победу.

$\boxed{\text{б}}$ Считаем теперь дисперсию:

$$\gamma_0 = \mathrm{Var}(y_t) = \mathrm{Var}(0.8y_{t-1} + \varepsilon_t) = 0.64\,\mathrm{Var}(y_{t-1}) + \mathrm{Var}(\varepsilon_t) + 2 \cdot 0 =$$
$$= 0.64\,\mathrm{Var}(y_{t-1}) + \sigma^2$$

При этом фокус. Стационарность значит, что у всех моменты сохраняются. Значит, у предыдущего игрека же такая же дисперсия. В итоге магия:

$$\gamma_0 = 0.64\gamma_0 + \sigma^2$$

$$\gamma_0 = \frac{\sigma^2}{0.36}$$

$\boxed{\text{в}}$ Теперь надо считать чудесные автоковариации:

$$\gamma_1 = \mathrm{Cov}(y_t, y_{t-1}) = \mathrm{Cov}(0.8y_{t-1} + \varepsilon_t, y_{t-1}) = 0.8\gamma_0$$

$$\gamma_2 = \mathrm{Cov}(y_t, y_{t-2}) = \mathrm{Cov}(0.8y_{t-1} + \varepsilon_t, y_{t-2}) =$$
$$= \mathrm{Cov}(0.8y_{t-1}, y_{t-2}) + \mathrm{Cov}(\varepsilon_t, y_{t-2}) = 0.8\,\mathrm{Cov}(y_{t-1}, y_{t-2}) =$$
$$= 0.8\gamma_1 = 0.64\gamma_0$$

$$\gamma_3 = \operatorname{Cov}(y_t, y_{t-3}) = \operatorname{Cov}(0.8y_{t-1} + \varepsilon_t, y_{t-3}) =$$
$$= 0.8 \operatorname{Cov}(y_{t-1}, y_{t-3}) = 0.8\gamma_2 = 0.8^3\gamma_0$$

Подставляем в автокорреляцию, получим:

$$\rho_1 = 0.8$$
$$\rho_2 = 0.8^2$$
$$\rho_3 = 0.8^3$$

Суть происходящего совершенно ясна.

8.1.2 Задача 2

У нас есть процесс $MA(1)$:

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

$\boxed{\text{а}}$ Считаем просто дисперсию, помня о некоррелированности эпсилонов:

$$\gamma_0 = \operatorname{Var}(y_t) = \operatorname{Var}(\varepsilon_t + 0.8\varepsilon_{t-1}) = 1.64\sigma^2$$

$\boxed{\text{б}}$ Можем теперь всякую тоску расписать:

$$\gamma_1 = \operatorname{Cov}(y_t,y_{t-1}) = \operatorname{Cov}(\varepsilon_t + 0.8\varepsilon_{t-1},\varepsilon_{t-1} + 0.8\varepsilon_{t-2}) =$$
$$= 0.8 \operatorname{Cov}(\varepsilon_{t-1},\varepsilon_{t-1}) = 0.8\sigma^2$$

Очевидно, что дальше пересечения индексов не будет:

$$\gamma_2 = \operatorname{Cov}(y_t,y_{t-2}) = 0$$

$$\rho_1 = 0.49$$
$$\rho_{2,\dots,k} = 0$$

8.1.3 Задача 3

Теперь у нас $AR(2)$. Все то же со стандартными нормальными эпсилонами:

$$y_t = 0.6y_{t-1} + 0.3y_{t-2} + \varepsilon_t$$

$\boxed{\text{а}}$ Проверяем на стационарность:

$$y_t(1 - 0.6L - 0.3L^2) = \varepsilon$$

$$1 - 0.6z - 0.3z^2 = 0$$

$$z_1 = 1.0817, z_2 = -3.08$$

Все стационарно.

$\boxed{\text{б}}$ Считаем дисперсию:

$$\begin{gathered}\gamma_0 = \mathrm{Var}(y_t) = \mathrm{Var}(0.6y_{t-1} + 0.3y_{t-2} + \varepsilon_t) = \\ = 0.36\,\mathrm{Var}(y_{t-1}) + 0.09\,\mathrm{Var}(y_{t-2}) + \mathrm{Var}(\varepsilon_t) + 2\,\mathrm{Cov}(0.6y_{t-1}, 0.3y_{t-2}) + \\ +2\,\mathrm{Cov}(0.6y_{t-1}, \varepsilon_t) + 2\,\mathrm{Cov}(0.3y_{t-2}, \varepsilon_t) = 0.36\gamma_0 + 0.09\gamma_0 + 1 + 0.36\gamma_1\end{gathered}$$

$$0.55\gamma_0 = 0.36\gamma_1 + 1$$

$\boxed{\text{в}}$ Считаем теперь γ_1:

$$\gamma_1 = \mathrm{Cov}(y_t, y_{t-1}) = \mathrm{Cov}(0.6y_{t-1} + 0.3y_{t-2} + \varepsilon, y_{t-1}) = 0.6\gamma_0 + 0.3\gamma_1 + 0$$

$$\gamma_1 = 0.6\gamma_0 + 0.3\gamma_1$$

$$\gamma_1 = \frac{6}{7}\gamma_0$$

Теперь дальше этот бесконечный процесс:

$$\begin{gathered}\gamma_2 = \mathrm{Cov}(y_t, y_{t-2}) = \mathrm{Cov}(0.6y_{t-1} + 0.3y_{t-2} + \varepsilon_t, y_{t-2}) = \\ = 0.6\gamma_1 + 0.3\gamma_0 + 0 = 0.6\gamma_1 + 0.3\gamma_0\end{gathered}$$

$$\gamma_3 = \mathrm{Cov}(y_t, y_{t-3}) = \mathrm{Cov}(0.6y_{t-1} + 0.3y_{t-2} + \varepsilon_t, y_{t-3}) = 0.6\gamma_2 + 0.3\gamma_1$$

$$\gamma_3 = 0.6\gamma_2 + 0.3\gamma_1$$

Подставляем в наши коэффициенты автокорреляции:

$$\rho_1 = 0.86$$

$$\rho_2 = 0.81$$

$$\rho_3 = 0.744$$

Осталось только понять, что $\gamma_0 = 4.17$.

8.1.4 Задача 4

Теперь у нас $MA(3)$:

$$y_t = 311 + \varepsilon_t + 0.8\varepsilon_{t-1} - 0.5\varepsilon_{t-2} - 0.3\varepsilon_{t-3}$$

Все стационарно, да. Ищем теперь дисперсию с опорой на некоррелированность ошибок:

$$\gamma_0 = \mathrm{Var}(y_t) =$$
$$= \mathrm{Cov}(311+\varepsilon_t+0.8\varepsilon_{t-1}-0.5\varepsilon_{t-2}-0.3\varepsilon_{t-3}, 311+\varepsilon_t+0.8\varepsilon_{t-1}-0.5\varepsilon_{t-2}-0.3\varepsilon_{t-3}) =$$
$$= 0 + \mathrm{Var}(\varepsilon_t) + 0.64\,\mathrm{Var}(\varepsilon_{t-1}) + 0.25\,\mathrm{Var}(\varepsilon_{t-2}) + 0.09\,\mathrm{Var}(\varepsilon_{t-3}) = 1.98\sigma^2$$

Коэффициенты автокорреляции:

$$\gamma_1 = \mathrm{Cov}(y_t,y_{t-1}) = 0.8\,\mathrm{Var}(\varepsilon_{t-1}) - 0.4\,\mathrm{Var}(\varepsilon_{t-2}) + 0.15\,\mathrm{Var}(\varepsilon_{t-3}) = 0.55\sigma^2$$

$$\gamma_2 = \mathrm{Cov}(y_t,y_{t-2}) = -0.74\sigma^2$$

$$\gamma_3 = \mathrm{Cov}(-0.3\varepsilon_{t-3},\varepsilon_{t-3}) = -0.3\sigma^2$$

После всех этих чудовищных манипуляций:

$$\rho_1 = 0.28$$
$$\rho_2 = -0.37$$
$$\rho_3 = -0.15$$

8.2 Семинар 23. Стационарность, прогнозирование

8.2.1 Задача 5

Теперь процесс:

$$y_t = 100 + 0.3y_{t-1} + 0.2y_{t-2} + \varepsilon_t$$

$\boxed{\text{а}}$ Проверяем стационарность:

$$1 - 0.3z - 0.2z^2 = 0$$

Один корень оказывается близким к 1.6, другой к -3.1. Все оказывается стационарным.

$\boxed{\text{б}}$ Считаем теперь безусловное математическое ожидание:

$$\mu = \mathrm{E}(y_t) = \mathrm{E}(100 + 0.3y_{t-1} + 0.2y_{t-2} + \varepsilon_t) = 100 + 0.3\mu + 0.2\mu$$

$$\mu = 200$$

$\boxed{\text{в}}$ Считаем теперь дисперсию:

$$\begin{gathered}\mathrm{Var}(y_t) = \mathrm{Var}(100 + 0.3y_{t-1} + 0.2y_{t-2} + \varepsilon_t) = \\ = 0.09\,\mathrm{Var}(y_{t-1}) + 0.04\,\mathrm{Var}(y_{t-2}) + \sigma^2 + 2\,\mathrm{Cov}(0.3y_{t-1}, 0.2y_{t-2}) = \\ = 0.09\gamma_0 + 0.04\gamma_0 + \sigma^2 + 0.12\gamma_1\end{gathered}$$

$$0.87\gamma_0 = 0.12\gamma_1 + \sigma^2$$

$$\gamma_1 = \mathrm{Cov}(y_t, y_{t-1}) = 0.3\gamma_0 + 0.2\gamma_1$$

В итоге:

$$\gamma_1 = \frac{3}{8}\gamma_0$$

$$\gamma_0 = \frac{\sigma^2}{0.825}$$

$\boxed{\text{г}}$ Теперь мы знаем, что $y_{t-1} = 400$, $y_{t-2} = 400$. Надо найти себе условное матожидание игрека.

$$\begin{gathered}\mathrm{E}(y_t|y_{t-1} = 400, y_{t-2} = 400) = \\ = \mathrm{E}(100 + 0.3y_{t-1} + 0.2y_{t-2} + \varepsilon_t | y_{t-1} = 400, y_{t-2} = 400) = \\ = 100 + 0.3 \cdot 400 + 0.2 \cdot 400 = 300\end{gathered}$$

$\boxed{\text{д}}$ Сделаем прогноз:

$$\begin{gathered}\hat{y}_{t+1} = \mathrm{E}(100 + 0.3y_t + 0.2y_{t-1} + \varepsilon_{t+1} | y_{t-1} = 400, y_{t-2} = 400) = \\ = 100 + 0.3\,\mathrm{E}(y_t|y_{t-1} = 400, y_{t-2} = 400) + 0.2 \cdot 400 = 270\end{gathered}$$

8.2.2 Задача 6

$$y_t = \alpha + \theta y_{t-1} + \varepsilon_t, |\theta| < 1$$

Построим прогноз. Достаточно очевидно, что лучший прогноз – условное матожидание вида:

$$\mathrm{E}(y_{t+1}|y_t) = \mathrm{E}(\alpha + \theta y_t + \varepsilon_t) = \alpha + \theta y_t$$

Если мы хотим получить еще на шаг вперед, получается так:

$$\mathrm{E}(y_{t+2}|y_t) = \mathrm{E}(\alpha + \theta y_{t+1} + \varepsilon_{t+1}|y_t) = \alpha + \theta(\alpha + \theta y_t)$$

Этот тренд достаточно ясен. Если мы назовем $f(z) = \alpha + \theta z$, то:

$$\mathrm{E}(y_{t+s}|y_t) = f^s(y_t)$$

Здесь степень функции берется в плане композиции.

Достаточно легко вычислить и дисперсию такой конструкции[1]:

$$\mathrm{Var}(\hat{y}_{t+1} - y_{t+1}) = \mathrm{Var}(\varepsilon_t) = \sigma^2_\varepsilon$$

Дальше уже чуть менее хорошо:

$$\mathrm{Var}(\hat{y}_{t+2} - y_{t+2}) = \mathrm{Var}(\alpha + \theta(\alpha + \theta y_t + \varepsilon_{t+1}) - \alpha + \theta(\alpha + \theta y_t) - \varepsilon_{t+2}) =$$

[1] Или возможна какая-то форма гетероскедастичности.

$$= \mathrm{Var}(\theta\varepsilon_{t+1} - \varepsilon_{t+2}) = \theta^2\sigma^2 + \sigma^2 = (1+\theta^2)\sigma^2$$

Продолжая в том же не очень вкусном духе, мы получим:

$$\mathrm{Var}(\hat{y}_{t+s} - y_{t+s}) = (1 + \theta^2 + \cdots + \theta^{2(s-1)})\sigma^2$$

8.2.3 Задача 7

Пусть у нас теперь есть процесс $AR(1)$:

$$y_t = 0.5y_{t-1} + \varepsilon_t$$

При этом здесь $\varepsilon \sim N(0,4)$, а $y_1 = 20$.

$\boxed{\text{а}}$ Построим прогноз для $\hat{y}_3$. Известно, что лучше условного матожидание не уйти:

$$\hat{y}_3 = \theta\theta y_1 = 5$$

$\boxed{\text{б}}$ Построим теперь 95% доверительный интервал для прогноза:

$$y_3 \in (\hat{y}_3 - 1.96\sqrt{C_3}, \hat{y}_3 + 1.96\sqrt{C_3})$$

Считаем ошибку[2]:

$$C_3 = \mathrm{Var}(y_3 - \hat{y}_3 | y_1) = \sigma^2 \frac{1-\theta^4}{1-\theta^2} = 5$$

Тогда выходит, что:

$$y_3 \in (-3.88, 4.88)$$

[2]Используем, что

$$1 + \theta^2 + \cdots + \theta^{2(s-1)} = \frac{1-\theta^{2s}}{1-\theta^2}$$

8.2.4 Задача 8

Теперь у нас процесс $MA(1)$:

$$y_t = \varepsilon_t + \alpha\varepsilon_{t-1}$$

Достаточно очевидно, что здесь с любым прогнозом совершенно просто:

$$\hat{y}_{t+s} = \mathrm{E}(\varepsilon_{t+s} + \alpha\varepsilon_{t+s-1}|y_t) = 0$$

Дисперсия ошибки такого прогноза тоже очевидна:

$$\mathrm{Var}(\hat{y}_{t+s} - y_{t+s}) = \mathrm{Var}(0 - \varepsilon_{t+s} - \alpha\varepsilon_{t+s-1}) = (1+\alpha^2)\sigma^2$$

8.2.5 Задача 9

Пусть есть $ARIMA(1,1,0)$:

$$\Delta y_t = \delta + \theta\Delta y_{t-1} + \varepsilon_t$$

Можно ее поразвинчивать:

$$y_t - y_{t-1} = \delta + \theta y_{t-1} + \theta y_{t-2} + \varepsilon_t$$

$$y_t = \delta + (\theta - 1)y_{t-1} + \theta y_{t-2} + \varepsilon_t$$

Легким движением руки превратили процесс в $AR(2)$. Дальше уже можно занудно, но концептуально просто прогнозировать.

8.2.6 Задача 10

Посмотрим процесс $ARMA$(2,1):

$$y_t = 0.6y_{t-1} - 0.3_{t-2} + \varepsilon_t + 0.9\varepsilon_{t-1}$$

$\boxed{\text{а}}$ Смотрим на стационарность:

$$y(1 - 0.6L + 0.3L^2) = \varepsilon_t(1 + 0.9L)$$

$$3z^3 - 6z + 10 = 0$$

Все снова стационарно.

$\boxed{\text{б}}$ Ищем дисперсию:

$$\gamma_0 = \operatorname{Var}(0.6y_{t-1} - 0.3y_{t-2} + \varepsilon_t + 0.9\varepsilon_{t-1}) =$$

$$= 0.36\gamma_0 + 0.09\gamma_0 + \sigma^2 + 0.81\sigma^2 - 2 \cdot 0.6 \cdot 0.3\gamma_1 + 2 \cdot 0.6 \cdot 0.9 \operatorname{Cov}(y_{t-1}, \varepsilon_{t-1})$$

$$\operatorname{Cov}(y_{t-1}, \varepsilon_{t-1}) = \operatorname{Cov}(0.6y_{t-2} - 0.3y_{t-3} + \varepsilon_{t-1} + 0.9\varepsilon_{t-2}, \varepsilon_{t-1}) = \sigma^2$$

В итоге:

$$\gamma_0 = 0.45\gamma_0 - 0.36\gamma_1 + 2.89\sigma^2$$

$$0.55\gamma_0 = 2.89\sigma^2 - 0.36\gamma_1$$

$$\gamma_1 = \operatorname{Cov}(y_t, y_{t-1}) =$$

$$= \operatorname{Cov}(0.6y_{t-1} - 0.3y_{t-2} + \varepsilon_t + 0.9\varepsilon_{t-1}, y_{t-1}) =$$

$$= 0.6\gamma_0 - 0.3\gamma_1 + 0.9\sigma^2$$

$$1.3\gamma_1 = 0.6\gamma_0 + 0.9\sigma^2$$

$$\gamma_2 = \operatorname{Cov}(y_t, y_{t-2}) =$$

$$= \operatorname{Cov}(0.6y_{t-1} - 0.3y_{t-2} + \varepsilon_t + 0.9\varepsilon_{t-1}, y_{t-2}) =$$

$$= 0.6 \operatorname{Cov}(y_{t-1}, y_{t-2}) - 0.3\gamma_0$$

$$\gamma_2 = 0.6\gamma_1 - 0.3\gamma_0$$

8.2.7 Задача 11

Посмотрим еще $ARMA(2,1)$:

$$y_t = 0.7y_{t-1} - 0.1y_{t-2} + \varepsilon_t - 0.2\varepsilon_{t-1}$$

$$y_t(1 - 0.7L + 0.1L^2) = \varepsilon_t(1 - 0.2L)$$

Оказывается, что происходит магия корней:

$$y_t(0.2L - 1)(1 - 0.5L) = \varepsilon_t(1 - 0.2L)$$

$$y_t = 0.5y_{t-1} + \varepsilon_t$$

Мы свели занудную задачу к $AR(1)$. Все до сих пор занудно, но хоть теперь попроще.

Глава 9
Семинары 26-30

9.1 Семинар 26-27. Многомерные модели временных рядов

9.1.1 Задача 1

При оценивании уважаемым исследователем модели распределенных лагов по 100 наблюдениям получилось следующее:

$$\hat{y}_t = 2.3 + 0.5x_t + 0.8x_{t-1} + 0.6x_{t-2} + 0.4x_{t-3}$$

В текущем периоде x увеличен на 1. Посмотрим итог:

$$\Delta y_t = 2.3 + 0.5\Delta x_t + 0.8\Delta x_{t-1} + 0.6\Delta x_{t-2} + 0.4\Delta x_{t-3}$$

$\boxed{\text{а}}$ В текущем периоде изменится только один член, значит, $\Delta y_t = 0.5$.

$\boxed{\text{б}}$ В следующем периоде:

$$\Delta y_{t+1} = 2.3 + 0.5\Delta x_{t+1} + 0.8\Delta x_t + 0.6\Delta x_{t-1} + 0.4\Delta x_{t-2}$$

Значит, здесь уже $\Delta y_{t+1} = 0.8$. При этом ясно, что возникает динамический мультипликатор второго периода:

$$\Delta y_t + \Delta y_{t+1} = 1.3$$

Это будет суммарное изменение.

$\boxed{\text{в}}$ Найдем теперь долгосрочный динамический эффект. Для этого надо найти оставшиеся куски:

$$\Delta y_{t+2} = 2.3 + 0.5\Delta x_{t+2} + 0.8\Delta x_{t+1} + 0.6\Delta x_t + 0.4\Delta x_t$$

$$\Delta y_{t+3} = 2.3 + 0.5\Delta x_{t+3} + 0.8\Delta x_{t+2} + 0.6\Delta x_{t+1} + 0.4\Delta x_t$$

Отсюда ясно, что $\Delta y_{t+2} = 0.6$, $\Delta y_{t+3} = 0.4$, дальше уже нули.
Значит, мы получаем долгосрочный динамический мультипликатор:

$$m = 0.5 + 0.8 + 0.6 + 0.4 = 2.3$$

9.1.2 Задача 2

Рассмотрим теперь авторегрессионную модель распределенных лагов ($ADL(p,q)$) о связи расходов фирму на рекламу и объем продаж фирмы. По 100 наблюдениям уважаемый исследователь получил это:

$$\hat{y}_t = 1.3 + 0.7y_{t-1} + 0.3x_t + 0.4x_{t-1} + 0.3x_{t-2} + 0.2x_{t-3}$$

Это очевидная $ADL(1,3)$. Пусть везде $\Delta x_t = 1$
$\boxed{\text{а}}$ Текущий эффект найти элементарно:

$$\Delta y_t = 0.3\Delta x_t = 0.3$$

$\boxed{\text{б}}$ Дальше в мире становится чуть печальнее:

$$\Delta y_{t+1} = 0.7\Delta y_t + 0.3\Delta x_{t+1} + 0.4\Delta x_t + 0.3\Delta x_{t-1} + 0.2\Delta x_{t-2}$$

Это значит, что:

$$\Delta y_{t+1} = 0.21 + 0.4 = 0.61$$

Тогда уже надо складывать и тот период, и этот:

$$\Delta y_t + \Delta y_{t+1} = 0.3 + 0.61 = 0.91$$

$\boxed{\text{в}}$ Долгосрочный мультипликатор чуть тоскливее. Понимаем, что модель у нас такая:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \delta_0 x_t + \delta_1 x_{t-1} + \delta_2 x_{t-2} + \delta x_{t-3} + \varepsilon$$

Отсюда строится горка эффектов:

$$\Delta y_t = \delta_0$$

$$\Delta y_{t+1} = \delta_0\beta_1 + \delta_1$$

$$\Delta y_{t+2} = \beta_1(\beta_1\delta_0 + \delta_1) + \delta_2$$

$$\Delta y_{t+3} = \beta_1(\beta_1(\beta_1\delta_0 + \delta_1) + \delta_2) + \delta_3$$

$$\Delta y_{t+4} = \beta_1((\beta_1(\beta_1\delta_0 + \delta_1) + \delta_2) + \delta_3)$$

Раскроем скобки:

$$\Delta y_{t+3} = \beta_1^2(\beta_1\delta_0 + \delta_1) + \beta_1\delta_2 + \delta_3$$

$$\Delta y_{t+4} = \beta_1^3(\beta_1\delta_0 + \delta_1) + \beta_1^2\delta_2 + \beta_1\delta_3$$

Очевидно, что возникает геометрическая прогрессия. $\hat{\beta}_1 < 1$, поэтому может просто просуммировать:

$$\delta_0+(\delta_0\beta_1+\delta_1)+\beta_1(\delta_0\beta_1+\delta_1)+\beta_1^2(\delta_0\beta_1+\delta_1)+\cdots = \delta_0+\frac{\delta_0\beta_1+\delta_1}{1-\beta_1}+\frac{\delta_2}{1-\beta-1}+\frac{\delta_3}{1-\beta_1} =$$

$$= \frac{\delta_0(1-\beta_1)+\delta_0\beta_1+\delta_1+\delta_2+\delta_3}{1-\beta_1} = \frac{\delta_0+\delta_1+\delta_2+\delta_3}{1-\beta_1}$$

Значит,

$$LRP = \frac{\delta_0+\delta_1+\delta_2+\delta_3}{1-\beta_1}$$

9.1.3 Задача CTB-15-1

У нас есть уравнение, связывающее темпы роста ВВП и O – *неотрицательное* отклонение цен на нефть от их максимума за прошлый год:

$$\hat{y}_t = \underset{(0.1)}{1.0} - \underset{(0.054)}{0.055}O_t - \underset{(0.057)}{0.026}O_{t-1} - \underset{(0.048)}{0.031}O_{t-2} - \underset{(0.042)}{0.109}O_{t-3} - \underset{(0.053)}{0.128}O_{t-4} +$$

$$+\underset{(0.025)}{0.008}O_{t-5} + \underset{(0.048)}{0.025}O_{t-6} - \underset{(0.039)}{0.019}O_{t-7} + \underset{(0.042)}{0.067}O_{t-8}$$

$\boxed{\text{а}}$ Внезапно случилось $\Delta O_t = 25$[1]. Выясним эффект за два года, учитывая, что у нас квартальне данные с 1955 по 2000 годы.

[1] При этом $O_{t+h} = 0$ для любого $h \geqslant 1$.

Возникает следующая конструкция:

$$\hat{y}_t = -0.055 \cdot 25$$

$$\hat{y}_{t+1} = -0.026 \cdot 25$$

И так далее.

$\boxed{\text{б}}$ Построим доверительный интервал для первого эффекта. Так как у нас здесь временное изменение, но ошибка просто берется от оценки коэффициента:

$$\Delta y_{t+h} \in 25 \cdot (\hat{\beta}_h - 1.96 e(\hat{\beta}_h), \hat{\beta}_h + 1.96 e(\hat{\beta}_h))$$

$$\Delta y_t \in (-4.021, 1.271)$$

Интервал включает ноль, значит, эффект не значим.

$\boxed{\text{в}}$ Накопленное влияние получается просто суммированием:

$$m \cdot 25 = 25 \sum \hat{\beta}_h = 8.375\%$$

Проверим на значимость:

$$H_0 : \delta_1 = \delta_2 = \cdots = \delta_8 = 0$$

$F_{\text{расч}} = 3.49$. $F_{\text{кр}}(0.01, 8, 184 - 8) = 2.51$. Вполне отклонили нулевую гипотезу, ура.

9.1.4 Задача 6

Есть модель:

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$$

У нас 18 периодов. Надо провести тест Дарбина-Уотсона, в рамках которого идет размышление о $\varepsilon_t = \rho \varepsilon_{t-1}$. Тестируется следующая нулевая гипотеза:

$$H_0 : \rho = 0$$

Статистика имеет следующий вид:

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

Мы знаем, что $\sum_{t=1}^{18} e_t^2 = 126$, $\sum_{t=2}^{18} e_t e_{t-1} = 45$, $e_1 = 1$, $e_{18} = -5$. Преобразуем статистику, чтобы подставить данные:

$$DW = \frac{\sum_{t=2}^{18} e_t^2 + \sum_{t=2}^{18} e_{t-1}^2 - 2\sum_{t=2}^{18} e_t e_{t-1}}{\sum_{t=1}^{18} e_t^2}$$

При этом мы понимаем, что:

$$\sum_{t=2}^{18} e_{t-1}^2 = e_1^2 + \cdot + e_{17}^2 + e_{18}^2 - e_{18}^2 = \sum_{t=1}^{18} e_t^2 - e_{18}^2$$

$$\sum_{t=2}^{18} e_t^2 = \sum_{t=1}^{18} e_t^2 - e_1^2$$

В итоге $DW = 1.087$. Помним, что там три интервальчика. Если от нуля до d_L, то $\rho > 0$, если от d_u до $4 - d_u$, то $\rho = 0$, если от $4 - d_L$ до 4, то $\rho < 0$. У нас $d_L = 1.16, d_u = 1.39$. Довольно ясно, что $\rho > 0$.

9.2 Семинар 27-28. VAR, ARCH, GARCH

9.2.1 Задача 1

Уважаемый исследователь анализирует так называемый эффект переноса валютного курса на цены: ослабление курса национальной валюты приводит к росту цен импортных товаров и, следовательно, к росту общего уровня цен, то есть к инфляции.

У нас есть приведенная VAR(1) модель:

$$\begin{cases} \hat{\pi}_t = 0.05 + 0.5\pi_{t-1} + 0.4x_{t-1}, R^2 = 0.8 \\ \hat{x}_t = 0.14 + 0.2\pi_{t-1} + 0.2x_{t-1}, R^2 = 0.4 \end{cases}$$

Здесь π – инфляция, x – темп прироста обратного курса национальной валюты в процентах к предыдущему месяцу. У него есть ежемесячные данные за 15 лет.

$\boxed{\text{а}}$ Чтобы найти средний уровень, надо найти безусловное матожидание:

$$E(\pi_t) = \mu_\pi$$

Берем матожидание от правых и левых частей:

$$\begin{cases} \mu_\pi = 0.05 + 0.5\mu_\pi + 0.4\mu_x \\ \mu_x = 0.14 + 0.2\mu_\pi + 0.2\mu_x \end{cases}$$

Из системы следует $\mu_\pi = 3\%$. Средний уровень инфляции равен 3%.

$\boxed{\text{б}}$ Чтобы построить прогноз, надо уже условные матожидания делать.

$$\mathrm{E}(\pi_{t+1}|\pi_t,x_t) = 0.05 + 0.5 \cdot 2 + 0.4 \cdot 8 = 4.25$$

$$\mathrm{E}(x_{t+1}|\pi_t,x_t) = 0.14 + 0.2 \cdot 2 + 0.2 \cdot 8 = 2.14$$

$$\mathrm{E}(\pi_{t+2}|\pi_t,x_t) = 0.05 + 0.5\,\mathrm{E}(\pi_{t+1}|\pi_t,x_t) + 0.4\,\mathrm{E}(x_{t+1}|\pi_t,x_t) = 3.031$$

$$\mathrm{E}(x_{t+2}|\pi_t,x_t) = 1.418$$

$\boxed{\text{в}}$ Осуществим теперь тест Грейнджера. Пусть в первом уравнении коэффициент при x_{t-1} – β_2. Во втором коэффициенты при π_{t-1} – α_1.

$$H_0^{(1)} : \beta_2 = 0$$

$$H_0^{(2)} : \alpha_1 = 0$$

Проведем тесты:

$$F^1_{\text{расч}} = \frac{(R^2_{UR} - R^2_R)/q}{(1 - R^2_{UR})/(T - R)} = 177$$

$$F^2_{\text{расч}} = 2.95$$

Довольно ясно, что $F_{\text{кр}} = 3.89$. Отклоняем первую нулевую гипотезу. Это означает, что темп прироста обратного курса – причина по Грейнджеру для инфляции, но обратно связи нет.

9.2.2 Задача 2

Уважаемый исследователь оценил VAR(1) модель с двумя переменными:

$$\boldsymbol{y}_t = \boldsymbol{\theta} \boldsymbol{y}_{t-1} + \boldsymbol{u}_t$$

При этом:

$$\hat{\boldsymbol{\theta}} = \begin{pmatrix} 0.5 & 0.3 \\ 0.2 & 0.7 \end{pmatrix}$$

$\boxed{\text{а}}$ Запишем просто в скалярной форме:

$$\begin{cases} \hat{y}_t = 0.5y_{t-1} + 0.3x_{t-1} \\ \hat{x}_t = 0.2y_{t-1} + 0.7x_{t-1} \end{cases}$$

$\boxed{\text{б}}$ Чтобы построить прогноз на два периода вперед, надо находить условное матожидание.

$$\mathrm{E}(y_{t+1}|y_t) = \hat{\theta}\,\mathrm{E}(y_t) = \hat{\theta} y_t$$

$$\mathrm{E}(y_{t+2}|y_t) = \hat{\theta}\,\mathrm{E}(y_{t+1}|y_t) = \hat{\theta}^2 y_t$$

У нас есть $y_t = (2,3)$. Тогда быстро и просто подставляем в модель:

$$\mathrm{E}(y_{t+2}|y_t) = 0.1^2 \begin{pmatrix} 31 & 36 \\ 24 & 55 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1.72 \\ 2.13 \end{pmatrix}$$

9.2.3 Задача 5

В наших моделях можем быть и условная гетероскедастичность $ARCH(p)$:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t_1}^2 + \dots + \alpha u_{t-p}^2$$

Или GARCH(p,q):

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \dots + \alpha_p u_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_q \sigma_{t-q}^2$$

Рядом проживает условная дисперсия для ARCH:

$$u_t = \varepsilon_t \sqrt{\alpha_0 + \alpha_1 u_{t_1}^2 + \dots + \alpha u_{t-p}^2}$$

Тогда выходит:

$$\mathrm{Var}(u_t | u_{t-1} \dots u_{t-p}) = \mathrm{Var}(\varepsilon_t)(\alpha_0 + \alpha_1 u_{t_1}^2 + \dots + \alpha u_{t-p}^2)$$

Аналогично можно для GARCH.
Если процесс стационарен, то все тоже может быть хорошо:

$$\mathrm{E}(\sigma_t^2) = \mathrm{E}(\alpha_0 + \alpha_1 u_{t_1}^2 + \dots + \alpha u_{t-p}^2) = \alpha_0 + \alpha_1 \mathrm{E}(u_{t-1}^2) + \dots + \alpha_p \mathrm{E}(u_{t-p}^2)$$
$$\sigma_u^2 = \alpha_0 + \alpha_1 \sigma_u^2 + \dots + \alpha_p \sigma_u^2$$
$$\sigma_u^2 = \frac{\alpha_0}{1 - \alpha_1 - \dots - \alpha_p}$$

У нас есть процесс:

$$y_t = \theta_0 + \theta_1 x_t + \varepsilon_t$$

ARCH(1): $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$

Надо найти безусловную дисперсию случайной ошибки $\mathrm{Var}(\varepsilon_t)$. В предпосылке стационарности:

$$\sigma_\varepsilon^2 = E(\sigma_t^2) = E(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2) = \alpha_0 + \alpha_1 E(\varepsilon_{t-1}^2)$$

Совершенно ясно, пусть $|\alpha_1| < 1$, $\mathrm{Var}(\varepsilon_t) = \dfrac{\alpha_0}{1 - \alpha_1}$.

9.2.4 Задача 6

Рассмотрим модель:

$$y_t = \theta_0 + \theta_1 x_t + \varepsilon_t$$

При этом у нас $GARCH(1{,}1)$:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

Найти здесь в разумных условиях безусловную дисперсию тоже элементарно:

$$\mathrm{E}(\sigma_t^2) = \mathrm{E}(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2) = \alpha_0 + \alpha_1 \mathrm{E}(\varepsilon_{t-1}^2) + \beta_1 \mathrm{E}(\sigma_{t-1}^2)$$

$$\sigma^2 = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$$

9.2.5 Задача 3

Рассмотрим некоторый процесс:

$$y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$$

Здесь есть $ARCH(1)$:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2$$

От нас требуется помочь уважаемому исследователю найти в этом стационарном процессе дисперсию $\mathrm{Var}(y_t)$. Сперва распишем выражение:

$$\mathrm{Var}(y_t) = \mathrm{Var}(\beta_0 + \beta_1 y_{t-1} + u_t) = \beta_1^2 \mathrm{Var}(y_{t-1}) + \mathrm{Var}(u_{t-1})$$

Дальше начинаем активно использовать стационарность:

$$\mathrm{Var}(y_t) = \frac{\mathrm{Var}(u_t)}{1 - \beta_1^2} = \frac{\frac{\alpha_0}{1-\alpha_1}}{1 - \beta_1^2}$$

Замену на втором равенстве мы сделали с помощью этого выражения и стационарности:

$$\sigma_u^2 = \mathrm{E}(\sigma_t^2) = \alpha_0 + \alpha_1 \mathrm{E}(u_{t-1}^2) = \alpha_0 + \alpha_1 \sigma_u^2$$

$$\sigma_u^2 = \frac{\alpha_0}{1 - \alpha_1}$$

9.2.6 Задача 4

У нас в этот раз редкий гость – процесс $MA(1)$:

$$y_t = \beta_0 + \beta_1 u_{t-1} + u_t$$

В духе сезона $ARCH(2)$:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2$$

Здесь уважаемому исследователю тоже надо помочь найти дисперсию y_t, учитывая, что процесс стационарен. Для этого запишем дисперсию и воспользуемся предпосылкой независимости:

$$\mathrm{Var}(y_t) = \mathrm{Var}(\beta_0 + \beta_1 u_{t-1} + u_t) = \beta_1^2 \mathrm{Var}(u_{t-1}) + \mathrm{Var}(u_t) = (1 + \beta_1^2)\sigma_u^2$$

Значит, теперь надо найти дисперсию для ошибок:

$$\mathrm{E}(\sigma_t^2) = \mathrm{E}(\alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2)$$

$$\sigma_u^2 = \frac{\alpha_0}{1 - \alpha_1 - \alpha_2}$$

Тогда очевидно:

$$\mathrm{Var}(y_t) = \frac{(1 + \beta_1^2)\alpha_0}{1 - \alpha_1 - \alpha_2}$$

9.2.7 Задача 7

Динамика ряда описывается процессом $AR(1)$:

$$y_t = 4 + 0.8y_{t-1} + u_t$$

При этом случайные ошибки описываются моделью $ARCH(1)$, $\varepsilon \sim N(0, 1)$:

$$u_t = \varepsilon_t \sqrt{0.01 + 0.06u_{t-1}^2}$$

$\boxed{\text{а}}$ Пусть нам известно, что при этом $y_1 = 20$, $y_2 = 22$. Построим прогноз на период вперед:

$$\hat{y}_3 = \mathrm{E}(y_3|y_1,y_2) = \mathrm{E}(4 + 0.8y_2 + u_3|y_2 = 22) = 4 + 0.8 \cdot 22 = 21.6$$

$\boxed{\text{б}}$ Построим доверительный интервал для оценки точности прогноза:

$$y_3 \in \left(\hat{y}_3 - 1.96\sqrt{C_1},\, \hat{y}_3 + 1.96\sqrt{C_1}\right)$$

Найдем же:

$$C_1 = \mathrm{Var}(e_3) = \mathrm{Var}(y_3 - \hat{y}_3) = \mathrm{Var}(5 + 0.8 \cdot 22 + u_3 - 21.6) = \mathrm{Var}(u_3)$$

Но при этом мы знаем, что:

$$y_2 = 4 + 0.8y_1 + u_2$$

$$u_2 = 2$$

Тогда можно вычислить и условную дисперсию:

$$\mathrm{Var}(u_3|u_2) = \mathrm{Var}(\varepsilon_3\sqrt{0.01 + 0.06u_2^2}) = 0.01 + 0.06 \cdot 4 = 0.25$$

Значит,

$$y_3 \in (21.6 - 1.96 \cdot 0.5, 21.6 + 1.96 \cdot 0.5)$$

Глава 10
Прочие задания

10.1 Листок 2

10.1.1 Задача 1.1

Для оценки зависимости урожайности пшеницы y от количества внесенных удобрений x проводится эксперимент по выращиванию пшеницы на n полях одинаковой площади, причем на первом поле используется одна единица удобрения, а на каждом следующем в два раза меньшее количество, чем на предыдущем поле. По собранным данным, исследователь оценивает модель регрессии $y_i = \theta \cdot x_i + \varepsilon_i$, предполагая, что

1. Выполняются все предпосылки теоремы Гаусса-Маркова;
2. Случайные ошибки нормальны.

Требуется проверить состоятельность полученной по МНК оценки для θ.

$\boxed{1}$ Найдем сперва саму МНК-оценку.

$$e_1^2 + ... + e_n^2 = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{\theta} x_i)^2 \to \min$$

$$\frac{\partial S}{\partial \hat{\theta}} = 0$$

$$\frac{\partial S}{\partial \hat{\theta}} = \sum_{i=1}^{n} -2x_i(y_i - \hat{\theta} x_i) = 0$$

$$\sum x_i y_i - \hat{\theta} \sum x_i^2 = 0$$

$$\hat{\theta} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\overline{xy}}{\overline{x^2}}$$

$\boxed{2}$ Проверим несмещенность:

$$\hat{\theta} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i(\theta x_i + \varepsilon_i)}{\sum x_i^2} = \frac{\theta \sum x_i^2 + \sum x_i \varepsilon_i}{\sum x_i^2} =$$

$$= \theta + \frac{\sum x_i \varepsilon_i}{\sum x_i^2}.$$

$$E(\hat{\theta}) = E\left(\theta + \frac{\sum x_i \varepsilon_i}{\sum x_i^2}\right) = \theta + E\left(\frac{\sum x_i \varepsilon_i}{\sum x_i^2}\right) = \theta + 0 = \theta$$

$\boxed{3}$ Так как мы хотим осуществлять проверку через неравенство Чебышева, найдем дисперсию:

$$\operatorname{Var}(\theta + \frac{\sum x_i \varepsilon_i}{\sum x_i^2}) = Var(\frac{\sum x_i \varepsilon_i}{\sum x_i^2}) = \frac{\operatorname{Var}(\sum x_i \varepsilon_i)}{(\sum x_i^2)^2} =$$

Используем предпосылки модели о независимости и дисперсии ошибок:

$$= \frac{\sum x_i^2 \operatorname{Var}(\varepsilon_i)}{(\sum x_i^2)^2} = \frac{\sigma^2}{\sum x_i^2}.$$

Значит, мы получаем, что

$$\operatorname{Var}(\hat{\theta}) = \frac{\sigma^2}{\sum x_i^2}.$$

$\boxed{4}$ Чтобы продвинуться дальше, нам существенно знать x.
Согласно условию, $x_1 = 1, x_2 = \frac{1}{2} \dots x_n = \frac{1}{n}$.
Тогда $\sum x_i^2 = 1 + \frac{1}{2^2} + \frac{1}{3^2} +$.
При $n \to \infty$ мы получаем эталонный сходящийся ряд:

$$\sum_{n=1}^{\infty} \frac{1}{n^2}$$

Как известно, на бесконечности мы получаем, что сумма этого ряда равна $\zeta(2) = \frac{\pi^2}{6}$.[1]

Так как каждый член суммы положителен, последовательность частичных сумм стремится к своему пределу – сумме ряда – снизу.

[1]Значение дзета-функции Римана в точке 2.

Согласно неравенству Чебышева,

$$P\left\{|\hat{\theta}-\theta| \geq \varepsilon\right\} \leq \frac{Var(\hat{\theta})}{\varepsilon^2}$$

$$\lim_{n\to\infty} \frac{\sigma^2}{S_n\varepsilon^2} = \frac{6\sigma^2}{\pi^2\varepsilon^2}$$

Достаточное условие не задалось, да. Неформально это значит, что после какого-то момента каждое новое поле дает настолько малое количество информации, что наша оценка не увеличивается

На этом обычно заканчивалось все, но теперь закрадываются сомнения и хочется через определение явно продемонстрировать.

Определение 10.1.1. Оценка параметра θ $\hat{\theta}$ называется **состоятельной**, если $\lim\limits_{n\to\infty} P\left\{|\hat{\theta}-\theta| \geq \varepsilon\right\} = 0$

Языком Вейерштрасса: $\forall\varepsilon > 0\,\forall\delta > 0\,\exists n_0 : \forall\, N > n_0\, P_n\left\{|\hat{\theta}-\theta| \geq \varepsilon\right\} < \delta$

Очевидно, чтобы опровергнуть состоятельность, нужно найти такую пару ε, δ, что не найдется нужного объема выборки. Т.е. предъявить такую комбинацию что у разности всегда строго положительная вероятность оказаться больше некоторого ε.

Пусть $\varepsilon = \sqrt{12}\sigma/\pi$. Даже на бесконечности $(\hat{\theta}-\theta) \sim N(0, \frac{6\sigma^2}{\pi^2})$[2]. Мы запросто можем вычислить искомую вероятность. Например, при единичной дисперсии ошибок вероятность, что оценка не дотянет до настоящего значения 0.5, будет равна 0.2. Таким образом, нет состоятельности.

10.1.2 Задача 1.2

В условиях эксперимента предыдущей задачи исследователь решил, что случайные ошибки гетероскедастичны с дисперсиями $\mathrm{Var}(\varepsilon) = \delta x_i^2, \delta > 0$. Требуется вычислить эффективную оценку.

Модель $y_i = \theta \cdot x_i + \varepsilon_i$. Зададим к новым переменным: $w_i = y_i/x_i, 1 = z_i = x_i/x_i$. Тогда получается модель регрессии на константу без гетероскедастичности. $u_i \sim N(0,\delta)$.

$$w_i = \theta + u_i$$

[2]Обычно при стремлении дисперсии к нулю, которая выходит у состоятельных оценок, нормальное распределение разности вырождается в дельта-функцию Дирака, что и обеспечивает сходимость.

$\boxed{1}$ Найдем МНК-оценку для теты:

$$e_1^2 + ... + e_n^2 = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{\theta})^2 \to \min$$

$$\frac{\partial S}{\partial \hat{\theta}} = 0$$

$$\frac{\partial S}{\partial \hat{\theta}} = \sum_{i=1}^{n} -2(w_i - \hat{\theta}) = 0$$

$$-2 \sum w_i = n\hat{\theta} = 0$$

$$\hat{\theta} = \overline{w}$$

$\boxed{2}$ Проверяем несмещенность:

$$E(\hat{\theta}) = E(\overline{w}) = \frac{1}{n} \sum E(w) = \frac{1}{n} E(n\theta + nu_i) = \theta + E(u_i) = \theta$$

$\boxed{3}$ Вычислим дисперсию:

$$\mathrm{Var}(\hat{\theta}) = \mathrm{Var}(\overline{w}) = \left(\frac{1}{n}\right)^2 \mathrm{Var}(\sum w_i) = \left(\frac{1}{n}\right)^2 \mathrm{Var}(n\theta + \sum u_i) =$$

$$= \left(\frac{1}{n}\right)^2 \mathrm{Var}(\sum u_i) = \frac{n\delta}{n^2} = \frac{\delta}{n}$$

$\boxed{4}$ Проверим эффективность: сделаем это самым дуболомным способом.

Определение 10.1.2. Оценка $\hat{\theta}$ называется **эффективной** в некотором классе оценок, если она имеет минимальную дисперсию в этом классе.

Определение 10.1.3. Информация Фишера - $I(\theta) = E\left(\frac{\partial \ln p(\xi,\theta)}{\partial \theta}\right)^2$

Сведения об эффективности мы получаем из неравенства Рао-Фреше-Крамера. Отсюда можно получить другое определение при выполнении условий теоремы.

Определение 10.1.4. Несмещенная оценка $\hat{\theta}$ называется **эффективной**, если $\mathrm{Var}(\hat{\theta}) = \dfrac{1}{nI(\theta)}$

Из-за нормальности ошибок выходит, что $w_i \sim N(\theta, \delta)$.

Плотность распределения:

$$p_\xi(w) = \frac{1}{\sqrt{2\pi\delta}} e^{-\dfrac{(y-\theta)^2}{2\delta}}$$

$$\ln\left(\frac{1}{\sqrt{2\pi\delta}} e^{-\dfrac{(y-\theta)^2}{2\delta}}\right) = -\ln(\sqrt{2\pi\delta}) - \frac{(y-\theta)^2}{2\delta}$$

$$\frac{\partial \ln p_\xi}{\partial\theta} = \frac{(y-\theta)}{\delta}$$

$$E\left(\left(\frac{(y-\theta)}{\delta}\right)^2\right) = \frac{1}{\delta^2} E(y-\theta)^2 = \frac{1}{\delta}$$

$$\mathrm{Var}(\hat{\theta}) = \frac{\delta}{n} = \frac{1}{nI(\theta)} = \frac{\delta}{n}$$

Таким образом, оценка эффективная по определению.

$\boxed{5}$ Проверим теперь состоятельность.

Как мы нашли ранее, $\mathrm{Var}(\hat{\theta}) = \dfrac{\delta}{n}$, а $E(\hat{\theta}) = \theta$).

Согласно неравенству Чебышева,

$$P\left\{|\hat{\theta} - \theta| \geq \varepsilon\right\} \leq \frac{Var(\hat{\theta})}{\varepsilon^2}$$

$$P\left\{|\hat{\theta} - \theta| \geq \varepsilon\right\} \leq \frac{\delta}{n\varepsilon^2}$$

Таким образом, $\mathrm{Var}(\hat{\theta}) \to 0$ при росте выборки. Значит, это состоятельная оценка.

10.1.3 Stock-Watson. Задача 17.3

Пусть дана модель:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$$

$\boxed{\text{а}}$ Воспользуемся альтернативным представлением МНК оценки для коэффициента при переменной: пусть при этом μ_x – матожидание иксов.

$$\hat{\beta}_2 = \frac{\frac{1}{n}\sum(x_i - \overline{x})(y_i - \overline{y})}{\frac{1}{n}\sum(x_i - \overline{x})^2} =$$

$$= \frac{\sum\left((x_i - \overline{x})(\beta_1 + \beta_2 * x_i + \varepsilon_i - \beta_1 - \beta_2 * \overline{x} - \overline{\varepsilon_i})\right)}{\sum(x_i - \overline{x})^2} =$$

$$= \frac{\sum(x_i - \overline{x})(\beta_2(x_i - \overline{x}) + \varepsilon_i - \overline{\varepsilon_i})}{\sum(x_i - \overline{x})^2} =$$

$$= \frac{\sum \beta_2 \sum(x_i - \overline{x})^2 + \sum(x_i - \overline{x})(\varepsilon_i - \overline{\varepsilon})}{\sum(x_i - \overline{x})^2} =$$

$$= \beta_2 + \frac{\sum(x_i - \overline{x})\varepsilon_i - \overline{\varepsilon}\sum(x_i - \overline{x})}{\sum(x_i - \overline{x})^2} =$$

$$= \beta_2 + \frac{\sum(x_i - \overline{x})\varepsilon_i - \overline{\varepsilon}0}{\sum(x_i - \overline{x})^2} =$$

$$= \beta_2 + \frac{\sum(x_i - \overline{x})\varepsilon_i}{\sum(x_i - \overline{x})^2}$$

Вычтем и домножим:

$$\sqrt{n}(\hat{\beta}_2 - \beta_2) = \sqrt{n}\frac{\sum(x_i - \overline{x})\varepsilon_i}{\sum(x_i - \overline{x})^2} = \sqrt{n}\frac{(\sum(x_i - \mu_x) - (\overline{x} - \mu_x))\varepsilon_i}{\sum(x_i - \overline{x})^2} =$$

$$= \sqrt{n}\frac{\sum(x_i - \mu_x)\varepsilon_i}{\sum(x_i - \overline{x})^2} - \sqrt{n}\frac{\sum(\overline{x} - \mu_x)\varepsilon_i}{\sum(x_i - \overline{x})^2} =$$

$$= \frac{\frac{1}{\sqrt{n}}\sum v_i}{\frac{1}{n}\sum(x_i - \overline{x})^2} - \frac{(\overline{x} - \mu_x)\frac{1}{\sqrt{n}}\sum \varepsilon_i}{\frac{1}{n}\sum(x_i - \overline{x})^2}$$

Здесь $v_i = (x_i - \mu_x)\varepsilon_i$.

$\boxed{\text{b}}$ Рассмотрим более детально второе выражение:

$$\frac{(\overline{x}-\mu_x)\frac{1}{\sqrt{n}}\sum\varepsilon_i}{\frac{1}{n}\sum(x_i-\overline{x})^2}$$

- При достаточно большой выборке начинает работать центральная предельная теорема. Таким образом, если $\varepsilon_i \sim N(0,\sigma^2)$[3]

  Таким образом, величина $a = \dfrac{\frac{1}{\sqrt{n}}\sum\varepsilon_i}{\sigma} = \dfrac{\sqrt{n}(\overline{\varepsilon}-0)}{\sigma} \sim N(0{,}1)$

- Оценка дисперсии $\frac{1}{n}\sum(x_i-\overline{x})^2$ смещена, но состоятельна.

$$b = \frac{1}{n}\sum(x_i-\overline{x})^2 = \frac{1}{n}\sum x_i^2 - \overline{x}^2 = \hat{\alpha}_2 - \hat{\alpha_1}^2$$

  Начальные моменты сходятся по вероятности к своим истинным значениям. Значит, по теореме Слуцкого, оценка дисперсии сходится по вероятности к теоретический дисперсии.

- Согласно закону больших чисел, $c = \overline{x} \xrightarrow{p} \mu_x$. Это значит, что скобка в выражении стремится к нулю.

Согласно теореме Слуцкого, наше выражение $\dfrac{a*c}{b} \xrightarrow{d} 0$. Как известно, сходимость по распределению означает сходимость по вероятности, если предельная величина – константа.

$\boxed{\text{c}}$ Работаем с предпосылкой, что наши регрессоры и случайные ошибки имеют ненулевые конечные начальные моменты четвертого порядка[4].

Утверждение 10.1.1. $\mathrm{Var}(v_i) = (x_i-\mu_x)\varepsilon_i$ конечна.

Доказательство. Распишем подробнее это выражение:

$$\mathrm{Var}((x_i-\mu_x)\varepsilon_i) = E(((x_i-\mu_x)\varepsilon_i)^2) - E(((x_i-\mu_x)\varepsilon_i))^2 \leq$$

[3]В случае гетероскедастичности можно было бы использовать ЦПТ с условием Линдеберга, но уважаемые авторы используют гомоскедастичность как предпосылку.

[4]Как-то это не совсем так. Если мы рассмотрим распределение Парето с a=1,b=1. Тогда $\alpha_4 = -\frac{1}{3}$, а вот дисперсия бесконечна

$$\leq E(((x_i - \mu_x)\varepsilon_i)^2) =$$

Делаем переход по неравенству Коши-Буняковского.

$$= E((x_i - \mu_x)^2)\varepsilon_i^2) \leq \sqrt{E((x_i - \mu_x)^4)E(\varepsilon_i^4)} < \infty$$

□

10.1.4 Задача 2

Пусть дана модель:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$$

Найдем асимптотическое распределение для $\hat{\beta}_2$.

Как мы помним,

$$\hat{\beta}_2 = \beta_2 + \frac{\sum(x_i - \overline{x})\varepsilon_i}{\sum(x_i - \overline{x})^2}$$

$\boxed{1}$ Рассмотрим сначала вопрос состоятельности оценки.

Сделаем замену. Пусть векторы $\boldsymbol{x} = (x_i - \overline{x}), \boldsymbol{\varepsilon} = (\varepsilon_i)$. Тогда получаем запись:

$$\hat{\beta}_2 = \beta_2 + \frac{\boldsymbol{x'\varepsilon}}{\boldsymbol{x'x}}$$

Вернемся назад к средним:

$$\hat{\beta}_2 = \beta_2 + \frac{\frac{1}{n}\boldsymbol{x'\varepsilon}}{\frac{1}{n}\boldsymbol{x'x}}$$

Как известно, $\hat{\sigma}^2$ – состоятельная оценка для теоретической дисперсии. Таким образом, $\frac{1}{n}\boldsymbol{x'x} \xrightarrow{p} \sigma_x^2$. Мы предполагаем, что x обладает конечной ненулевой дисперсией, поэтому это какое-то число.

Рассмотрим теперь числитель нашего выражения.

$$\frac{1}{n}\sum(x_i - \overline{x})\varepsilon_i = \overline{x\varepsilon} - \overline{x} * \overline{\varepsilon}$$

Согласно закону больших чисел и теореме Слуцкого:

$$\overline{x\varepsilon} - \overline{x} * \overline{\varepsilon} \xrightarrow{p} E(x_i\varepsilon_i) - E(x_i)E(\varepsilon_i)$$

Но наши предпосылки (Key Concept 17.1) требуют независимости каждого наблюдения и случайной ошибки. Значит, мы можем разорвать произведение под матожиданием:

$$\overline{x\varepsilon} - \overline{x} * \overline{\varepsilon} \xrightarrow{p} E(x_i\varepsilon_i) - E(x_i)E(\varepsilon_i) = E(x_i)E(\varepsilon_i) - E(x_i)E(\varepsilon_i) = 0$$

Значит,

$$\overline{x\varepsilon} - \overline{x} * \overline{\varepsilon} \xrightarrow{p} 0$$

Таким образом, мы получаем, согласно теореме Слуцкого, состоятельность:

$$\hat{\beta}_2 \xrightarrow{p} \beta_2$$

$\boxed{2}$ Покажем асимптотическую нормальность.

Вернемся к нашему числителю, разделив теперь выражение на корень из числа наблюдений:

Опираясь на предпосылки, согласно Центральной Предельной Теореме[5], эта величина распределена асимптотически нормально:

$$\frac{1}{\sqrt{n}}\boldsymbol{x\varepsilon} \xrightarrow{d} N(0, \mathrm{Var})$$

Найдем дисперсию этого выражения:

$$\mathrm{Var}(\boldsymbol{x'\varepsilon}) = E(\boldsymbol{x'\varepsilon} - E(\boldsymbol{x'\varepsilon})^2) = E(\boldsymbol{x'\varepsilon})^2 =$$

Что есть квадрат такого выражения? Надо просто умножить еще раз на эти векторы так, что общее выражение осталось скаляром.

$$= E(\boldsymbol{x'\varepsilon})^2 = E(\boldsymbol{xx'\varepsilon\varepsilon'}) = E(\boldsymbol{xx'\varepsilon^2})) =$$

Так как, согласно предпосылкам, мы работаем в случае с гомоскедастичностью, применяем теорему о сглаживании[6]:

$$= E(\boldsymbol{xx'}E(\boldsymbol{\varepsilon}|\boldsymbol{x})) = n\sigma_x^2\sigma^2$$

Значит,

$$\frac{1}{\sqrt{n}}\boldsymbol{x\varepsilon} \xrightarrow{d} N(0, \sigma_x^2\sigma^2)$$

[5] В формулировке Линдеберга-Леви.

[6] Закон повторного матожидания.

Преобразуем теперь выражение для коэффицента, вычтя истинное значение и домножив на корень:

$$\sqrt{n}(\hat{\beta}_2 - \beta_2) = \frac{\frac{1}{\sqrt{n}}\boldsymbol{x'\varepsilon}}{\frac{1}{n}\boldsymbol{x'x}}$$

По ЦПТ и теореме Слуцкого:

$$\sqrt{n}(\hat{\beta}_2 - \beta_2) \xrightarrow{d} \frac{1}{\sigma_x^2} N(0, \sigma_x^2\sigma^2) = N(0, \frac{\sigma^2}{\sigma_x^2})$$

10.1.5 Stock-Watson. Задача 18.7

18.7 Consider the regression model, $Y_i = \beta_1 X_i + \beta_2 W_i + u_i$, where for simplicity the intercept is omitted and all variables are assumed to have a mean of zero. Suppose that X_i is distributed independently of (W_i, u_i) but W_i and u_i might be correlated and let $\hat{\beta}_1$ and $\hat{\beta}_2$ be the OLS estimators for this model. Show that

a. Whether or not W_i and u_i are correlated, $\hat{\beta}_1 \xrightarrow{p} \beta_1$.

b. If W_i and u_i are correlated, then $\hat{\beta}_2$ is inconsistent.

c. Let $\hat{\beta}_1^r$ be the OLS estimator from the regression of Y on X (the restricted regression that excludes W). Provide conditions under which $\hat{\beta}_1$ has a smaller asymptotic variance than $\hat{\beta}_1^r$, allowing for the possibility that W_i and u_i are correlated.

$\boxed{\text{a}}$ Покажем состоятельность коэффициента при первом регрессоре. Запишем нашу модель в векторном виде:

$$\boldsymbol{y} = \beta_1 \boldsymbol{X} + \beta_2 \boldsymbol{W} + \boldsymbol{\varepsilon}$$

Найдем МНК-оценки по известной формуле:

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'y$$

Подставим назад, чтобы получить альтернативное выражение для нашего вектора:

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}$$

Распишем подробнее:

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \sum x_i^2 & \sum x_i w_i \\ \sum x_i w_i & \sum w_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum x_i \varepsilon_i \\ \sum w_i \varepsilon_i \end{pmatrix}$$

Используя независимость между x и ε, x и w, равенство нулю матожиданий переменных, рассмотрим матрицы в правой части выражения, вынесем по $\dfrac{1}{n}$ из каждого выражения. Мы можем это сделать, так как из обратной к квадратной как раз вынесется n. Согласно закону больших чисел:

$$\begin{pmatrix} \frac{1}{n}\sum x_i^2 & \frac{1}{n}\sum x_i w_i \\ \frac{1}{n}\sum x_i w_i & \frac{1}{n}\sum w_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{n}\sum x_i \varepsilon_i \\ \frac{1}{n}\sum w_i \varepsilon_i \end{pmatrix} \xrightarrow{p} \begin{pmatrix} E(x^2) & 0 \\ 0 & E(w_i^2) \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ E(w_i \varepsilon_i) \end{pmatrix}$$

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} E(x^2) & 0 \\ 0 & E(w_i^2) \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ E(w_i \varepsilon_i) \end{pmatrix} \tag{10.1}$$

Таким образом, $\hat{\beta}_1 \xrightarrow{p} \beta_1$.

$\boxed{\text{b}}$ Покажем несостоятельность $\hat{\beta}_2$ при коррелированности второго регрессора и ошибок.

Из 10.1 видно, что:

$$\hat{\beta}_2 \xrightarrow{p} \beta_2 + \frac{E(w\varepsilon)}{E(w^2)}$$

Если корреляция между этими величинами не ноль, то мы не можем разорвать матожидание в числителе на два нулевых члена. Таким образом, получается, что, вообще говоря, $\hat{\beta}_2$ несостоятельна.

10.2 Листок 3

10.2.1 Stock-Watson. Задача 12.3

12.3 A classmate is interested in estimating the variance of the error term in Equation (12.1).

a. Suppose that she uses the estimator from the second-stage regression of TSLS: $\hat{\sigma}_a^2 = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \hat{\beta}_0^{TSLS} - \hat{\beta}_1^{TSLS}\hat{X}_i)^2$, where $\hat{X}_i$ is the fitted value from the first-stage regression. Is this estimator consistent? (For the purposes of this question suppose that the sample is very large and the TSLS estimators are essentially identical to β_0 and β_1.)

b. Is $\hat{\sigma}_b^2 = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \hat{\beta}_0^{TSLS} - \hat{\beta}_1^{TSLS}X_i)^2$ consistent?

Уравнение (12.1) имеет вид:

$$y_i = \beta_o + \beta_1 x_i + \varepsilon_i$$

При этом x и u коррелированы.

$\boxed{\text{а}}$ Рассмотрим оценку $\hat{\sigma}_a^2 = \dfrac{1}{n-2}\sum(y_i - \hat{\beta}_0^{TSLS} - \hat{\beta}_1^{TSLS}\hat{x}_i)^2$. Заменим оценки на истинные значения и приведем к виду с остатками регрессии:

$$\hat{\sigma}_a^2 = \frac{1}{n-2}\sum(y_i - \hat{\beta}_0^{TSLS} - \hat{\beta}_1^{TSLS}\hat{x}_i)^2 = \frac{1}{n-2}\sum(y_i - \beta_0 - \beta_1\hat{x}_i)^2 =$$

$$= \frac{1}{n-2}\sum(y_i - \beta_0 - \beta_1 x_i + \beta_1 x_i - \beta_1\hat{x}_i)^2 = \frac{1}{n-2}\sum(e_i + \beta_1(x_i - \hat{x}_i))^2 =$$

$$= \frac{1}{n-2}\left(\sum e_i^2 + 2\beta_1\sum e_i(x_i - \hat{x}_i) + \beta_1^2\sum(x_i - \hat{x}_i)^2\right)$$

Рассмотрим это выражение:

- $\dfrac{\sum e_i}{n-2} \xrightarrow{p} \sigma_\varepsilon^2$.
- $\dfrac{2\beta_1\sum e_i(x_i - \hat{x}_i)}{n-2}$ кажется подозрительным.

$$\sum e_i(x_i - \hat{x}_i) = \sum e_i x_i - \sum e_i\hat{x}_i = \sum e_i x_i$$

Если мы покажем, что $\sum e_i x_i$ не ноль (не сходится к нулю), то мы покажем несостоятельность оценки.

Это выражение можно рассматривать как многочлен от иксов. Как мы доказали раньше, многочлен $\sum e_i \hat{x}_i \equiv 0$ из-за МНК.

$\sum e_i x_i \equiv 0 \Leftrightarrow \forall i\ x_i = \hat{x}_i$. Но это невозможно, потому что тогда, так как сам икс был коррелирован со случайной ошибкой, то и оценка первого шага сохранит эту корреляцию. Значит, этот компонент перманентно не равен нулю.

Значит, оценка пункта $\boxed{\text{a}}$ несостоятельна.

$\boxed{\text{b}}$ Рассмотрим оценку $\hat{\sigma}_b^2 = \dfrac{1}{n-2}\sum(y_i - \hat{\beta}_0^{TSLS} - \hat{\beta}_1^{TSLS}x_i)^2$.
Проведем те же преобразования и легко покажем состоятельность:

$$\hat{\sigma}_b^2 = \frac{1}{n-2}\sum(y_i - \hat{\beta}_0^{TSLS} - \hat{\beta}_1^{TSLS}x_i)^2 = \frac{1}{n-2}\sum e_i^2 \xrightarrow{p} \sigma_\varepsilon^2$$

10.2.2 Stock-Watson. Задача 12.5

12.5 Consider the instrumental variable regression model

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + u_i,$$

where X_i is correlated with u_i and Z_i is an instrument. Suppose that the first three assumptions in Key Concept 12.4 are satisfied. Which IV assumption is not satisfied when:

a. Z_i is independent of (Y_i, X_i, W_i)?

b. $Z_i = W_i$?

c. $W_i = 1$ for all i?

d. $Z_i = X_i$?

Продублируем первые три предположения из 12.4:

1. $E(u_i|w_{1i}...w_{ri}) = 0$

2. $\boldsymbol{x}$,$\boldsymbol{w}$,$\boldsymbol{z}$,$\boldsymbol{y}$ i.i.d. из их совместного распределения
3. Маловероятны далеко удаленные от основной массы точки: $\boldsymbol{x}$,$\boldsymbol{w}$,$\boldsymbol{z}$,$\boldsymbol{y}$ имеют ненулевые конечные четвертые начальные моменты.

$\boxed{\text{a}}$ Наш инструмент становится нерелевантен: он некоррелирован с регрессором, он никак не улучшит нашу жизнь. Нарушается предпосылка релевантности, инструмент невалиден.

$\boxed{\text{b}}$ Случится катастрофа полной мультиколлинеарности, так как мы оцениванием на первом шаге регрессию:

$$x_i = \alpha_0 + \alpha_1 z_i + \alpha_2 w_i + \varepsilon_i$$

Таким образом, получается, что мы не сможем оценить такую регрессию.

$\boxed{\text{c}}$ Снова происходит полная мультиколлинеарность, но уже с константой.

$\boxed{\text{d}}$ Нарушается предпосылка экзогенности: так как x был коррелирован со случайной ошибкой, то и z, таким образом, с ней коррелирован.

10.2.3 Stock-Watson. Задача 12.7

12.7 In an instrumental variable regression model with one regressor, X_i, and two instruments, Z_{1i} and Z_{2i}, the value of the J-statistic is $J = 18.2$.

a. Does this suggest that $E(u_i|Z_{1i}, Z_{2i}) \neq 0$? Explain.

b. Does this suggest that $E(u_i|Z_{1i}) \neq 0$? Explain.

Авторы задачи намекают нам, что мы должны сделать тест Саргана-Хансена.

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$$

$$x_i = \alpha_0 + \alpha_1 z_i^{(1)} + \alpha_2 z_i^{(2)} + v_i$$

Он доступен нам, так как 2=m>k=1. Есть сверхидентификация. Формулируем гипотезы:

$$H_0 : \text{все инструменты экзогенны}$$

$$H_1 : \text{есть эндогенные инструменты}$$

В условиях нулевой гипотезы $\chi^2_{\text{расч}} = m * F_{\text{расч}} \sim \chi(m-k)$. $F_{\text{расч}}$ берется из теста с нулевой гипотезой $H_0 : \alpha_1 = \alpha_2$.

У нас $\chi^2_{\text{расч}} = 18.2$. $\chi^2_{\text{кр}}(1, 0.005) = 7.879$. Значит, на тончайшем уровне значимости мы отвергаем нулевую гипотезу. С вероятностью 99.5% у нас проблемы с эндогенностью инструментов. Однако ответ на вопрос пункта $\boxed{\text{b}}$ отрицателен, мы не знаем точно, где именно проблема.

10.2.4 Stock-Watson. Задача 12.9

12.9 A researcher is interested in the effect of military service on human capital. He collects data from a random sample of 4000 workers aged 40 and runs the OLS regression $Y_i = \beta_0 + \beta_1 X_i + u_i$, where Y_i is the worker's annual earnings and X_i is a binary variable that is equal to 1 if the person served in the military and is equal to 0 otherwise.

a. Explain why the OLS estimates are likely to be unreliable. (*Hint:* Which variables are omitted from the regression? Are they correlated with military service?)

b. During the Vietnam War there was a draft, where priority for the draft was determined by a national lottery. (Birthdates were randomly selected and ordered 1 through 365. Those with birthdates ordered first were drafted before those with birthdates ordered second, and so forth.) Explain how the lottery might be used as an instrument to estimate the effect of military service on earnings. (For more about this issue, see Joshua D. Angrist, "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administration Records," *American Economic Review*, June 1990: 313–336.)

$\boxed{\text{a}}$ Легко придумать целые горы потерянных переменных, которые нужно было бы включать как контрольные переменные, например.

- Очевидна коррелированная с решением служить переменная, измеряющая каким-то образом состояние здоровья. Состояние здоровье влияет и на его способность работать и зарабатывать деньги

  Весьма сложно придумать какую-то полноценную переменную, измеряющую эффективно здоровье для всех, можно использовать какие-то другие. Пример: обращение к врачу за последний год, количество хронических болезней, категории здоровья, выдаваемые военной медицинской комиссией[7][8]

- Классический пример: измеримое образование и неизмеримый талант. Сюда можно добавить часто изучаемые дополнительные показатели вроде знания дополнительных языков. Скорее всего, они коррелированы с решением о службе, но сложно указать направление связи.

$\boxed{\text{b}}$ Рассмотрим вопрос с лотерееей.

NB: Вовсе не все «победители» лотереи служили в армии (часть из них не была отобрана в силу различных ограничений, например, из-за здоровья), и в тоже время в армии служили не только «победители» лотереи (были, например, добровольцы).

Это примечание означает, что коэффициент корреляции между службой и исходом лотерии не равен единице.

Мы можем использовать исход лотереи как инструмент. Он валиден:

- Релевантность: исход лотерии заметно влиял на шансы попасть под призыв. Таким образом, служба и исход лотереи коррелированы.

- Экзогенность: согласно условию, выпадение чисел было случайно. Таким образом, исход лотерии не зависит от случайных ошибок нашей регрессии.

Используя его, мы можем получить TSLS-оценки коэффициентов.

[7] Все эти вариант страдают от сильных проблем. К первой, например, потребуются контрольные переменные, а третья может страдать от selection bias, ошибок, неполноты и прочих проблем.

[8] Можно наплодить еще связанных с армией переменных: возможно добавить какие-нибудь переменные, измеряющие тесноту контакта с армией до решения, служить ли: расстояние до ближайшей военной базы, участие в каких-то армейских программах, семейная история.

10.2.5 Stock-Watson. Задача 12.10

12.10 Consider the instrumental variable regression model $Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + u_i$, where Z_i is an instrument. Suppose that data on W_i are not available and the model is estimated omitting W_i from the regression.

a. Suppose that Z_i and W_i are uncorrelated. Is the IV estimator consistent?

b. Suppose that Z_i and W_i are correlated. Is the IV estimator consistent?

Имеем регрессию:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 w_i + \varepsilon_i$$

Чудовищным образом вынуждены оценить простую форму:

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$$

Так как мы используем 2МНК, то регрессия для оценки имеет вид:

$$\begin{cases} x_i = a_0 + a_1 z_i + v_i \\ y_i = \beta_0 + \beta_1 \hat{x}_i + \varepsilon_i \end{cases}$$

Найдем оценки:

$$\hat{a}_1 = \frac{\widehat{\mathrm{Cov}}(x,z)}{\widehat{\mathrm{Var}}(z)}$$

$$\hat{\beta}_1^{TSLS} = \frac{\widehat{\mathrm{Cov}}(y,\hat{x})}{\widehat{\mathrm{Var}}(\hat{x})} = \frac{\widehat{\mathrm{Cov}}(y, a_0 + a_1 z + v_i)}{\widehat{\mathrm{Var}}(a_0 + a_1 z_i} = \frac{a_1 \widehat{\mathrm{Cov}}(y,z)}{a_1^2 \widehat{\mathrm{Var}}(z)} = \frac{\widehat{\mathrm{Cov}}(y,z)}{a_1 \widehat{\mathrm{Var}}(z)} =$$

$$= \frac{\widehat{\mathrm{Cov}}(y,z)}{\frac{\widehat{\mathrm{Cov}}(x,z)}{\widehat{\mathrm{Var}}(z)} \widehat{\mathrm{Var}}(z)} = \frac{\widehat{\mathrm{Cov}}(y,z)}{\widehat{\mathrm{Cov}}(x,z)} = \frac{\widehat{\mathrm{Cov}}(z, \beta_0 + \beta_1 x_i + \beta_2 w_i + \varepsilon_i)}{\widehat{\mathrm{Cov}}(x,z)} =$$

$$= \frac{\beta_1 \widehat{\mathrm{Cov}}(z,x) + \beta_2 \widehat{\mathrm{Cov}}(z,w)}{\widehat{\mathrm{Cov}}(x,z)}$$

$\boxed{\text{a}}$ В таком случае второй член в числителе обнуляется, так как выборочная ковариация стремится к своему теоретическому значению. Значит, $\hat{\beta}_1^{TSLS} \xrightarrow{p} \beta_1$. Случилась состоятельность.

$\boxed{\text{b}}$ В таком случае, следуя аналогичной логике, мы получаем, что наша оценка не будет состоятельной: второй член в числителе не обнулится.

10.3 Не вошедшее в семинары 1-15

10.3.1 МКП. Задача 3.19

Утверждение 10.3.1. При добавлении в модель регрессора скорректированный коэффициент детерминации R^2_{adj} увеличивается тогда и только тогда, когда t-статистика оценки коэффициента при этом регрессоре по модулю превосходит единицу

Доказательство. $\boxed{\text{а}}$ Как мы знаем, квадрат t-статистики оценки коэффициента совпадает с F-статистикой для проверки гипотезы о равенстве этого коэффициента нулю, которая в свою очередь равна:

$$F_{\text{расч}} = \frac{R^2_{UR} - R^2_R}{(1 - R^2_{UR})/(n-k)}$$

k – число регрессоров в длинной регрессии.

$\boxed{\text{б}}$ Преобразуем выражение для R^2_{adj}:

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

$$= \frac{TSS(n-k) - ESS(n-k) - ESS(k-1)}{(n-k)TSS} = 1 - \frac{(n-1)}{TSS}\frac{ESS}{n-k} =$$

$$= 1 - \frac{n-1}{n-k}(1-R^2)$$

Таким образом, сразу ясно, что есть увеличение R^2_{adj}. Запишем это условие при добавлении еще одного регрессора:

$$\left(1 - (1 - R^2_{UR})\frac{n-1}{n-k}\right) - \left(1 - (1-R^2_R)\frac{n-1}{n-k+1}\right) > 0$$

Это превращается в:

$$(1-R^2_R)\frac{n-1}{n-k+1} - (1-R^2_{UR})\frac{n-1}{n-k} =$$

$$= \frac{(n-1)(1-R^2_R)(n-k) - (1-R^2_{UR})(n-k+1)}{(n-k+1)(n-k)} =$$

$$= \frac{n-1}{(n-k+1)(n-k)}((n-k)(R^2_{UR} - R^2_R - (1 - R^2_{UR})) =$$
$$= \frac{(n-1)(1-R^2_{UR})}{(n-k+1)(n-k)}\left(\frac{R^2_{UR} - R^2_R}{(1-R^2_{UR})/(n-k)} - 1\right) =$$
$$= \frac{(n-1)(1-R^2_{UR})}{(n-k+1)(n-k)}(F_{\text{расч}} - 1)$$

Отсюда немедленно видно, что нужно, что $t^2_{\text{расч}}$ была больше единицы, что и дает нам наше утверждение.

$\square$

МКП. Задача 3.23

Рассматривается классическая линейная нормальная модель:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

При этом $\mathrm{Var}(\boldsymbol{\varepsilon}) = 2I$, $\hat{\beta}_1 = 3$, $\hat{\beta}_2 = 2$, и:

$$X'X = \begin{pmatrix} 5 & 2 \\ 2 & 4 \end{pmatrix}$$

$\boxed{\text{а}}$ Построим 95% доверительный интервал для $\beta_1 + \beta_2$.

Для этого нам требуется знать дисперсии о коварициии наших оценок. Как мы понимаем,

$$(X'X)^{-1} = \begin{pmatrix} \frac{1}{4} & -\frac{1}{8} \\ -\frac{1}{8} & \frac{5}{16} \end{pmatrix}$$

Значит, раз $\sigma^2 = 2$, то получаем:

$$\mathrm{Var}(\hat{\beta}_1 + \hat{\beta}_2) = 2 * \left(\frac{1}{4} + \frac{5}{16} - \frac{1}{8}\right) = -\frac{1}{4}$$

Получаем 95%-доверительный интервал:

$$\beta_1 + \beta_2 \in \left(\hat{\beta}_1 + \hat{\beta}_2 - 1.96 * \sqrt{0.625}; betah1 + \hat{\beta}_2 + 1.96 * \sqrt{0.625}\right)$$

$$\beta_1 + \beta_2 \in (3.45; 6.55)$$

$\boxed{\text{б}}$ Построим 95%-доверительный интервал для вектора $(\beta_1 \quad \beta_2)$.

Как мы осознаем, наши коэффициенты распределены нормально, как и ошибки. Мы знаем о том, что они несмещенные, и знаем их дисперсию. Иными словами:

$$\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix} \sim N\left(\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}, \begin{pmatrix}0.5 & -0.25\\ -0.25 & 0.625\end{pmatrix}\right)$$

Если мы сделаем из каждой компоненты стандартную нормальную, потом перемножим и сложим, то у нас получится сумма нормальных квадратов – $\chi^2(2)$.

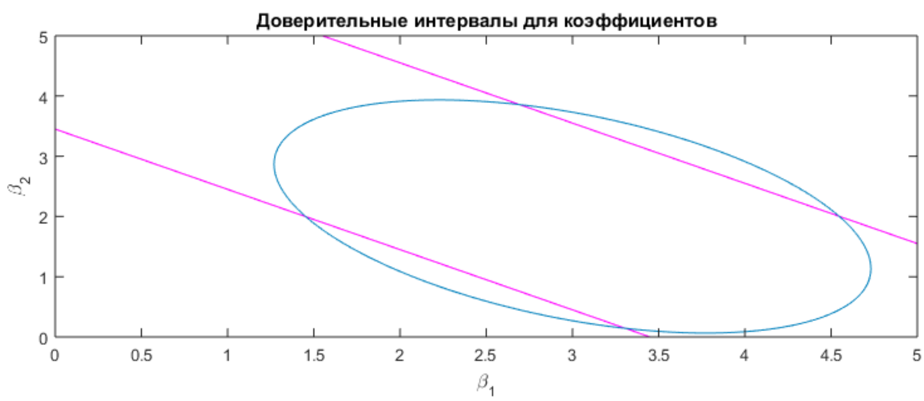
Сделаем это:

$$\begin{pmatrix}\hat{\beta}_1-\beta_1 & \hat{\beta}_2-\beta_2\end{pmatrix}\begin{pmatrix}2.5 & 1\\ 1 & 2\end{pmatrix}\begin{pmatrix}\hat{\beta}_1-\beta_1\\ \hat{\beta}_2-\beta_2\end{pmatrix} \sim \chi^2(2)$$

Раз мы знаем точную дисперсию, то коррекция Бесселя не требуется. Две степени свободы. $\chi^2_{\text{кр}}(0.05{,}2) = 5.99$ Значит, мы получаем доверительный эллипс:

$$2.5(\beta_1-3)^2+2(\beta_1-3)(\beta_2-2)+2(\beta_2-2)^2<5.99$$

Ниже дан график, совмещающий эти два множества. Между розовыми линиями лежит интервал из первого вопроса, внутри эллипса - из второго. Вероятность попасть в каждую из областей – 95%.



10.3.2 МКП. Задача 3.24

Для города и для деревни рассматриваются две модели парной регрессии. Двадцать наблюдений для города дали следующие результаты:

$$X'X = \begin{pmatrix} 20 & 20 \\ 20 & 25 \end{pmatrix}, X'y = \begin{pmatrix} 10 \\ 20 \end{pmatrix}, y'y = 30$$

Десять для деревни – следующие:

$$X'X = \begin{pmatrix} 10 & 10 \\ 10 & 20 \end{pmatrix}, X'y = \begin{pmatrix} 8 \\ 20 \end{pmatrix}, y'y = 24$$

Надо проверить на уровне значимость 5% гипотезу, что эти модели не совпадают.

Для этого воспользуемся тестом Чоу. Как мы помним, его статистика имеет вид:

$$F_{\text{расч}} = \frac{(ESS_{\sum} - ESS_A - ESS_B)/k}{(ESS_A + ESS_B)/(n-2k)} \sim F(k, n-2k)$$

При этом $ESS_{\sum}$ – ESS, вычисленная по общей модели без разделения на подвыборки, а ESS_A, ESS_B – на основе оцененных по подвыборкам моделям. Подставляем:

$$F_{\text{расч}} = \frac{(ESS_{\sum} - ESS_A - ESS_B)/2}{(ESS_A + ESS_B)/(10+20-2*2)} \sim F(2, 26)$$

$$ESS = e'e = (y-\hat{y})'(y-\hat{y}) = y'y - \hat{y}'\hat{y} = y'y - y'X(X'X)^{-1}X'y$$

Найдем все ESS для трех моделей:

$\boxed{1}$ Для города:

$$(X'X)^{-1} = \frac{1}{100}\begin{pmatrix} 25 & -20 \\ -20 & 20 \end{pmatrix}$$

$$y'X(X'X)^{-1}X'y = 25$$

$$ESS = 5$$

$\boxed{2}$ Для деревни:

$$(X'X)^{-1} = \frac{1}{100}\begin{pmatrix} 20 & -10 \\ -10 & 10 \end{pmatrix}$$

$$y'X(X'X)^{-1}X'y = 20.8$$

$$ESS = 3.2$$

$\boxed{3}$ Для общей модели:

$$X'X = \begin{pmatrix} 30 & 30 \\ 30 & 45 \end{pmatrix}, X'y = \begin{pmatrix} 18 \\ 40 \end{pmatrix}, y'y = 54$$

Тогда:

$$(X'X)^{-1} = \frac{1}{450} \begin{pmatrix} 45 & -30 \\ -30 & 30 \end{pmatrix}$$

$$y'X(X'X)^{-1}X'y = 39.067$$

$$ESS = 14.933$$

Значит, $F_{\text{расч}} = 10.675$. $F_{\text{кр}}(0.05, 2, 26) = 3.37$. Значит, отвергаем гипотезу о совпадении моделей.

10.3.3 МКП. Задача 3.25

Уважаемый исследовать провел две регрессии по 46 наблюдениям с одной зависимой переменной и пересекающимися регрессорами. Он получил следующее:

$$\hat{y}_i = 40 + 0.3x_i^{(2)} + 0.8x_i^{(3)} - 1.8x_i^{(4)}, R^2 = 0.82$$

$$\hat{y}_i = 60 + 0.5x_i^{(2)} + 0.6x_i^{(3)}, R^2 = 0.75$$

Надо проверить для первой регрессии гипотезу о равенстве β_4 минус единице.

Легко осознаем, что $t^2_{\text{расч}}$ равна $F_{\text{расч}}$ теста на короткую и длиинную регрессию с отрубанием последнего регрессора:

$$\frac{(R^2_{UR} - R^2_R)/1}{(1 - R^2_{UR})/(46-4)} = 16.33$$

$$t_{\text{расч}} = \frac{\hat{\beta}_4}{se(\hat{\beta}_4)}$$

Подставляем и находим стандартную ошибку. Как оказывается, $se(\hat{\beta}_4) = 0.445$.

Теперь уже составляем обычную статистику:

$$t_{\text{расч}} = \frac{\hat{\beta}_4 - (-1)}{se(\hat{\beta}_4)} \sim t_{n-k}$$

$t_{\text{расч}} = -1.80$, $t_{\text{кр}}(0.05, 42) = 2.021$. Таким образом, принимаем нулевую гипотезу.

10.3.4 МКП. Задача 3.28

Оценивание производственной функции по МНК дало следующий результат:

$$\ln Q = 1.37 + \underset{0.257}{0.632} \ln K + \underset{0.219}{0.452} \ln L$$

При этом $R^2 = 0.98$, $\widehat{\text{Cov}}(\hat{\beta}_K, \hat{\beta}_L) = 0.055$.

$\boxed{\text{а}}$ Проверим гипотезу о равенстве эластичностей по труду и капиталу в зависимости от числа наблюдений.

$$H_0 : \beta_K - \beta_L = 0$$

Выпишем t-статистику, раз одно линейное ограничение:

$$t_{\text{расч}} = \frac{(\beta_K - \beta_L) - 0}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_K - \hat{\beta}_L)}}$$

Раскрываем дисперсию под корнем, находим выражение, подставляем числа:

$$t_{\text{расч}} = \frac{0.632 - 0.452}{\sqrt{0.004}} = 2.846 \sim t_{n-3}$$

Будем изучать соотношение $t_{\text{кр}}(0.05)$ и $t_{\text{расч}}$. Критический переход между 3 и 4 степенями свободы – критическое значение переходит от 3.182 к 2.776. Значит, если у нас не меньше семи наблюдений, то мы отвергаем нулевую гипотезу. Если меньше семи, то мы не можем ее отвергнуть.

10.3.5 МКП. Задача 4.5

Рассмотрим регрессию в рамках КЛММР:

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

x – фиктивная переменная. Пусть $\overline{y}_0$ – среднее значение переменной y для наблюдений с $x = 0$, а $\overline{y}_1$ – для тех, у которых $x = 1$. Пусть первых n_0 штук, а вторых n_1 штука.

$\boxed{\text{а}}$ Найдем МНК-оценки для этих коэффициентов. Упорядочим наблюдения так, что сначала идут все с нулевым значением x, а потом с единичным.

Тогда матрица X:

$$X = \begin{pmatrix} 1 & 0 \\ \dots & \dots \\ 1 & 0 \\ 1 & 1 \\ \dots & \dots \\ 1 & 1 \end{pmatrix}$$

Тогда элементарно получаем нужные матрицы:

$$X'X = \begin{pmatrix} n & n_1 \\ n_1 & n_1 \end{pmatrix}$$

$$(X'X)^{-1} = \frac{1}{n_0 n_1} \begin{pmatrix} n_1 & -n_1 \\ -n_1 & n \end{pmatrix}$$

$$X'y = \begin{pmatrix} \overline{y}n \\ \overline{y}_1 n_1 \end{pmatrix} = \begin{pmatrix} \overline{y}_i n_0 + \overline{y}_1 n_1 \\ \overline{y}_1 n_1 \end{pmatrix}$$

Наконец:

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \overline{y}_0 \\ \overline{y}_1 - \overline{y}_0 \end{pmatrix}$$

$\boxed{\text{б}}$ Найдем дисперсию коэффициентов в таком случае:

$$\operatorname{Var}(\hat{\beta}_1) = \operatorname{Var}(\overline{y}_0) = \frac{\sum\limits_{i=1}^{n_o} \operatorname{Var}(y_i)}{n_0^2} = \frac{\sigma^2}{n_0}$$

$$\operatorname{Var}(\hat{\beta}_2) = \operatorname{Var}(\overline{y}_1 - \overline{y}_0) = \operatorname{Var}(\overline{y}_1) + \operatorname{Var}(\overline{y}_0) = \frac{n\sigma^2}{n_0 n_1}$$

10.3.6 МКП. Задача 4.23

Уважаемый исследователь получил оценку регрессии удельного выпуска на уровень заработной платы:

$$\ln\frac{Q}{L} = 0.374 + \underset{0.049}{0.805}W + \varepsilon$$

При этом $R^2 = 0.929$

Проверим, значима ли эта связь.

Как известно, этот вопрос сводится к тестированию гипотезы $H_0 : \beta_2 = 0$. Мы не знаем число наблюдений, но знаем, что $t^2_{\text{расч}} = F_{\text{расч}}$, где:

$$F_{\text{расч}} = \frac{R^2}{1-R^2}(n-2)$$

Выражаем число наблюдений:

$$n = 2 + \frac{t^2(1-R^2)}{R^2}$$

Подставляем сюда $t^2_{\text{расч}} = \left(\frac{0.805}{0.049}\right)^2 = 269.9$. $R^2 = 0.929$. Значит, $n = 23$. Сравниваем с критическим значением и отвергаем нулевую гипотезу.

10.3.7 МКП. Задача 5.6

Рассмотрим уравнение регрессии:

$$y_i = \beta + \varepsilon_i$$

Пусть ошибки регрессии удовлетворяют следующим условиям: $E(\varepsilon_i) = 0$, $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ для $i \neq j$, $\text{Var}(\varepsilon_i) = \sigma^2 x_i$, $x_i > 0$.

$\boxed{\text{а}}$ Найдем МНК-оценку и ее дисперсию в таком случае.

Как мы помним, в регрессии без переменной:

$$\hat{\beta}^{OLS} = \overline{y} = \frac{1}{n}\left(\sum y_i\right)$$

Из условия очевидно, что:

$$E(y_i) = \beta$$
$$\text{Var}(y_i) = \sigma^2 x_i$$
$$\text{Cov}(y_i, y_j) = 0,\ i \neq j$$

Поэтому мы сразу получаем, что наша оценка несмещенная и имеет следующую дисперсию:

$$\mathrm{Var}(\hat{\beta}^{OLS}) = \frac{\sigma^2}{n^2}\sum x_i$$

$\boxed{\text{б}}$ Найдем ОМНК-оценку. В данном случае это просто ВМНК-оценка, взвешенная по корням из иксов.

$$\frac{y_i}{\sqrt{x_i}} = \beta\frac{1}{\sqrt{x_i}} + \frac{\varepsilon_i}{\sqrt{x_i}}$$

Здесь уже очевидно гомоскедастичность победила. Применяем теперь МНК:

$$\hat{\beta}^{GLS} = \left(\sum\frac{y_i}{x_i}\right)\left(\sum\frac{1}{x_i}\right)$$

Дисперсия тогда:

$$\mathrm{Var}(\hat{\beta}^{GLS}) = \left(\sum\frac{\sigma^2 x_i}{x_i^2}\right)\left(\sum\frac{1}{x_i}\right)^2 = \frac{\sigma^2}{\sum\frac{1}{x_i}}$$

Очевидно, что МНК-дисперсия больше, чем найденное выражение:

$$\mathrm{Var}(\hat{\beta}^{OLS}) \geq \mathrm{Var}(\hat{\beta}^{GLS})$$

Это эквивалентно:

$$\left(\sum x_i\right)\left(\sum\frac{1}{x_i}\right) \geq n^2$$

Это вытекает из неравенства Коши-Шварца.

10.3.8 МКП. Задача 6.12

Оценить ОМНК следующую модель:

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

В модели только гетероскедастичность: $\mathrm{Var}(\varepsilon_i) = \sigma^2 x_i^2$.

x	4	1	5	8	2
y	6	3	12	15	4

$\boxed{\text{а}}$ Оценка ОМНК в данном случае совпадает с ВМНК. Делим обе части на x.

$$\frac{y_i}{x_i} = \beta\frac{1}{x_i} + \beta_2 + u_i$$

Все очень гомоскедастично.

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'y$$

$$X'X = \begin{pmatrix} \sum \frac{1}{x_i^2} & \sum \frac{1}{x_i} \\ \sum \frac{1}{x_i} & n \end{pmatrix}$$

Помним, что игреки у нас тоже изменились при замене.

$$X'y = \begin{pmatrix} \sum \frac{y_i}{x_i^2} \\ \sum \frac{y_i}{x_i} \end{pmatrix}$$

Значит, $\hat{\boldsymbol{\beta}} = (1.218, 1.649)$.

$\boxed{\text{б}}$ Оценим ковариационную матрицу.
Довольно очевидно, что $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}) = S^2(X'X)^{-1}$.
Таким образом,

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}) = \begin{pmatrix} 0.363 & -0.151 \\ -0.151 & 0.099 \end{pmatrix}$$

10.3.9 МКП. Задача 7.1

Утверждение 10.3.2. В векторном случае КЛММР дисперсия ошибка прогноза имеет следующую формулу:

$$E(\hat{y} - y_{n+1})^2 = \sigma^2(1 + x'_{n+1}(X'X)^{-1}x_{n+1})$$

Доказательство. $\boxed{\text{а}}$ Ошибка прогноза равна:

$$e = \hat{y} - y_{n+1} = x'_{n+1}\hat{\boldsymbol{\beta}} - (x'_{n+1}\boldsymbol{\beta} + \varepsilon_{n+1})$$

Мы помним, что МНК-оценка несмещенная, а случайная ошибка имеет нулевое матожидание. Тогда прогноз тоже несмещенный:

$$E(e) = E(x'_{n+1}\hat{\boldsymbol{\beta}}) - x'_{n+1}\boldsymbol{\beta} - E(\varepsilon_{n+1}) = x'_{n+1}\boldsymbol{\beta} - x'_{n+1}\boldsymbol{\beta} = 0$$

$\boxed{\text{б}}$ Делаем из ошибки квадрат ошибки:

$$E(e^2) = \operatorname{Var}(e) = \operatorname{Var}(x'_{n+1}\hat{\boldsymbol{\beta}} - x'_{n+1}\boldsymbol{\beta} - \varepsilon_{n+1}) =$$
$$= \operatorname{Var}(x'_{n+1}\hat{\boldsymbol{\beta}} - \varepsilon_{n+1}) = \operatorname{Var}(x'_{n+1}\hat{\boldsymbol{\beta}}) + \operatorname{Var}(\varepsilon_{n+1}) - 2\operatorname{Cov}(x'_{n+1}\hat{\boldsymbol{\beta}}, \varepsilon_{n+1})$$

$\boxed{\text{в}}$ Используем некоррелированность оценок и ошибок. Получаем:

$$E(e^2) = \operatorname{Var}(x'_{n+1}\hat{\boldsymbol{\beta}}) + \operatorname{Var}(\varepsilon_{n+1}) = x'_{n+1}\operatorname{Var}(\hat{\boldsymbol{\beta}})x_{n+1} + \sigma^2 =$$
$$= \sigma^2 x'_{n+1}(X'X)^{-1}x_{n+1} + \sigma^2 = \sigma^2(1 + x'_{n+1}(X'X)^{-1}x_{n+1})$$

□

10.4 Не вошедшее в семинары 15-21

10.4.1 МКП. Задача 10.10

Пусть у нас есть модель:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

В ней есть гетероскедастичность. У первых n_1 наблюдений $\mathrm{Var}(\varepsilon_i) = \sigma_1^2$, у остальных $\mathrm{Var}(\varepsilon_t) = \sigma_2^2$. Ошибки не коррелированы друг с другом.

Вектор ошибок нормален. Построим тест отношения правдоподобия для проверки гипотезы:

$$H_0 : \sigma_1^2 = \sigma_2^2$$

Статистику мы помним:

$$LR = -2(\ln(\tilde{\boldsymbol{\beta}},\tilde{\sigma}^2) - \ln L(\hat{\boldsymbol{\beta}},\hat{\sigma}_1^2,\hat{\sigma}_2^2))$$

Параметры с тильдой получаются в результате оценки с ограничением на равенство дисперсий, с крышками – без ограничений.

Начнем решать задачу:

$$\ln L(\boldsymbol{\beta},\sigma^2) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}\sum (y_i - x_i'\boldsymbol{\beta})^2$$

Очевидно, что мы можем получить вектор ошибок:

$$\tilde{\boldsymbol{e}} = \boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}$$

Отсюда возникает очевидная оценка дисперсии:

$$\tilde{\sigma}^2 = \frac{\tilde{\boldsymbol{e}}'\tilde{\boldsymbol{e}}}{n}$$

Тогда оцененная функция правдоподобия:

$$\ln(\tilde{\boldsymbol{\beta}},\tilde{\sigma}^2) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \tilde{\sigma}^2 - \frac{1}{2}\frac{\tilde{\boldsymbol{e}}'\tilde{\boldsymbol{e}}}{\tilde{\sigma}^2} = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \tilde{\sigma}^2 - \frac{n}{2}$$

Если нет ограничения, то все получается симпатично.

$$\ln L(\boldsymbol{\beta}, \sigma_1^2, \sigma_2^2) = -\frac{n}{2}\ln 2\pi - \frac{n_1}{2}\ln \sigma_1^2 - \frac{n_2}{2}\ln \sigma_2^2 -$$

$$-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n_1}(y_i-\boldsymbol{x_i'\beta})^2-\frac{1}{2\sigma_2^2}\sum_{i=n_1+1}^{n_1+n_2}(y_i-\boldsymbol{x_i'\beta})^2$$

Это надо продифференцировать по сигмам. Тогда мы получим:

$$\hat{\sigma_1^2}=\frac{1}{n_1}\sum_{i=1}^{n_1}(y_i-\boldsymbol{x_i'\beta})^2=\frac{1}{n_1}\sum_{i=1}^{n_1}e_i^2$$

$$\hat{\sigma_2^2}=\frac{1}{n_2}\sum_{i=n_1+1}^{n_1+n_2}(y_i-\boldsymbol{x_i'\beta})^2=\frac{1}{n_2}\sum_{i=n_1+1}^{n_1+n_2}e_i^2$$

Отсюда очевидной подстановкой:

$$LR=-n\ln\tilde{\sigma}^2+n_1\ln\hat{\sigma}_1^2+n_2\ln\hat{\sigma}_2^2$$

Это очевидно обобщается очевидным образом на много разных дисперсий:

$$H_0:\sigma_1^2=\cdots=\sigma_r^2$$

Тогда статистика здесь:

$$LR=-n\ln\tilde{\sigma}^2+n_1\ln\hat{\sigma}_1^2+\cdots+n_r\ln\hat{\sigma}_r^2$$

10.4.2 МКП. Задача 10.12

Пусть p – вероятность выпадения орла при бросании монеты. В ста испытаниях $x=42$ раза выпал орел, а 58 – решка. Проверим гипотезу:

$$H_0:p=0.5$$

Выпишем функцию правдоподобия и найдем информацию Фишера:

$$\ln L=x\ln p+(n-x)\ln(1-p)$$

$$\frac{\partial\ln L}{\partial p}=\frac{x}{p}-\frac{n-x}{1-p}=\frac{x-np}{p(1-p)}=0$$

$$I(p) = \frac{n}{p(1-p)}$$

Раз ограничение одно, то сравниваем все статистики с $\chi^2(1)$. По жизни пользуемся асимптотической нормальностью, чтобы строить статистики.

$\boxed{\text{а}}$ Проверим тестом Вальда это. Используя эффективность оценки ММП и теоремой Рао-Фреше, понимаем, что асимптотическая дисперсия $\hat{p}$ по одному наблюдению:

$$\widehat{\mathrm{Var}}(\hat{p}) = \hat{p}(1-\hat{p})$$

Подставляем в статистику Вальда:

$$W = (\hat{p} - p_0)' \frac{n}{\hat{p}(1-\hat{p})} (\hat{p} - p_0) = \frac{n(\hat{p} - p_0)^2}{\hat{p}(1-\hat{p})}$$

Высчитываем конкретные числа:

$$W = \frac{100 \cdot (0.42 - 0.5)^2}{0.42(1-0.42)} = 2.627$$

Даже $\chi^2(1{,}0.1) = 2.706$. Таким образом, мы не можем отклонить нулевую гипотезу.

$\boxed{\text{б}}$ Проверим тестом множителей Лагранжа это. Стоит помнить, что в этом тесте используются только оценки, сделанные с ограничением.

Получаем замечательную статистику, не забывая, что $x = n\hat{p}$:

$$LM = \left(\frac{x - np_0}{p_0(1-p_0)} \right)' \frac{n}{p_0(1-p_0)} \left(\frac{x - np_0}{p_0(1-p_0)} \right) = \frac{n(\hat{p} - p_0)^2}{p_0(1-p_0)}$$

Делаем подстановку:

$$LM = \frac{100 \cdot (0.42 - 0.5)^2}{0.5(1-0.5)} = 2.56$$

Здесь тем более отклонить не можем.

$\boxed{\text{в}}$ Проверим тестом отношения правдоподобия это. Здесь достаточно подставить в логарифмы и посмотреть разность.

$$\ln L = x \ln p + (n - x) \ln(1-p)$$

Сначала без ограничения:

$$\ln L_{UR} = 42\ln 0.42 + 58\ln 0.58 = -68.03$$

Теперь с ограничением:

$$\ln L_R = 42\ln 0.5 + 58\ln 0.5 = -69.32$$

Вычисляем статистику:

$$LR = 2.58$$

Здесь снова не можем отклонить. Таким образом, во всех трех ситуациях мы не отклонили нулевую гипотезу.

10.4.3 МКП. Задача 10.13

Пусть теперь есть 80 наблюдений величины из распределения Пуассона[9] . При этом знаем, что $\overline{x} = 1.7$. Нам надо проверить гипотезу:

$$H_0 : \lambda = 2$$

Снова честно придется найти оценки ММП и информацию Фишера о нашей оценке.

$$L = \prod \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}$$

Из этого сразу получается логарифм функции правдоподобия:

$$\ln L = -\sum \ln(x_i!) + \sum x_i \ln\lambda - n\lambda = -\sum \ln x_i! + n\overline{x}\ln\lambda - n\lambda$$

$$\frac{\partial \ln L}{\partial\lambda} = \frac{n\overline{x}}{\lambda} - n == 0$$

[9]Плотность распределения:

$$p(x = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$\hat{\lambda} = \overline{x}$$

$$I_n(\lambda) = \frac{n}{\lambda}$$

Здесь снова $\chi^2(1)$ мы используем для сравнения. Снова опираемся на теорему Рао, чтобы получать статистики.

$\boxed{\text{а}}$ Проверим тестом Вальда.

Выписываем статистику:

$$W = (\hat{\lambda} - \lambda_0)' \frac{n}{\hat{\lambda}} (\hat{\lambda} - \lambda_0)$$

Подставляем:

$$W = \frac{80 \cdot (1.7 - 2)^2}{1.7} = 4.235$$

$\chi^2(0.05, 1) = 3.841$. Значит, случилось отклонение нулевой гипотезы на уровне 5%.

$\boxed{\text{б}}$ Проверим тестом множителей Лагранжа.

$$LM = \left(n \frac{\overline{x} - \lambda_0}{\lambda_0}\right)' \frac{n}{\lambda_0} \left(n \frac{\overline{x} - \lambda_0}{\lambda_0}\right) = \frac{n(\hat{\lambda} - \lambda_0)^2}{\lambda_0}$$

Подставляем это:

$$LM = \frac{80 \cdot (1.7 - 2)^2}{2} = 3.6$$

Здесь вот уже не можем отклонить на уровне значимости в 5%.

$\boxed{\text{в}}$ Проверим тестом отношения правдоподобия.

Посчитать пару логарифмов будет затруднительно, придется делать преобразования:

$$\ln L = -\sum \ln x_i! + n\overline{x} \ln \lambda - n\lambda$$

$$LR = -2 \left(-\sum \ln x_i! + n\overline{x} \ln \lambda_0 - n\lambda_0 - \left(-\sum \ln x_i! + n\overline{x} \ln \hat{\lambda} - n\hat{\lambda}\right)\right) =$$

$$= 2n\left(\hat{\lambda}\ln\frac{\hat{\lambda}}{\lambda_0} + (\lambda_0 - \hat{\lambda})\right)$$

В это уже можно подставить:

$$LR = 2\cdot 80\cdot\left(1.7\ln\frac{1.7}{2} + 0.3\right) = 3.795$$

Здесь тоже не вышло отклонить нулевую гипотезу на уровне значимости 5%.

10.4.4 МКП. Задача 12.1

Утверждение 10.4.1. Если среди регрессоров логит-модели есть константа, то среднее значение прогнозных вероятностей равно доле единиц во всей выборке зависимой переменной.

Доказательство. Доказательство довольно ясно:

$$p(y_i = 1) = \Lambda(\boldsymbol{x_i'}\boldsymbol{\beta})$$

Наш логит: $\Lambda(u) = \dfrac{1}{1-e^{-u}}$. Выписываем сразу логарифм функции правдоподобия.

$$\ln L = \sum(y_i\ln\Lambda + (1-y_i)\ln(1-\Lambda))$$

Дифференцируем по $\boldsymbol{\beta}$:

$$\frac{\partial\ln L}{\partial\boldsymbol{\beta}} = \sum\left(y_i\frac{\Lambda'}{\Lambda} - (1-y_i)\frac{\Lambda'}{1-\Lambda}\right)\boldsymbol{x} = 0$$

В это надо подставить, что $\Lambda' = \Lambda(1-\Lambda)$.

Получаем:

$$\frac{\partial\ln L}{\partial\boldsymbol{\beta}} = \sum(y_i(1-\Lambda) - (1-y_i)\Lambda)\boldsymbol{x} = \sum(y_i - \Lambda)\boldsymbol{x} = \sum(y_i - \hat{p})\boldsymbol{x} = 0$$

$$(\boldsymbol{y} - \hat{\boldsymbol{p}})'\boldsymbol{X} = 0$$

Первый стобец матрицы $\boldsymbol{X}$ – единичный. Умножаем частично:

$$(\boldsymbol{y}-\hat{\boldsymbol{p}})'\boldsymbol{i} = n\left(\frac{1}{n}\sum y_i - \frac{1}{n}\sum p_i\right) = 0$$

$$\sum y_i = \sum p_i$$

$$\overline{y} = \overline{p}$$

Так как сумма y есть сумма единиц, мы как раз получаем утверждение.

□

10.4.5 МКП. Задача 12.8

Пусть у нас есть 100 наблюдений трех бинарных переменных:

Число наблюдений	y	x	z
12	0	0	0
10	0	0	1
20	0	1	0
0	0	1	1
8	1	0	0
28	1	0	1
22	1	1	0
0	1	1	1

$\boxed{\text{а}}$ Проверим в рамках логита значимость влияния z на y.
Мы оцениваем:

$$P(y_i = 1) = \Lambda(\alpha + \beta x_i + \gamma z_i)$$

Выпишем функцию правдоподобия:

$$L = \Lambda(\alpha)^8 \Lambda(\alpha+\gamma)^{28} \Lambda(\alpha+\beta)^{22} (1-\Lambda\alpha)^{12} (1-\Lambda(\alpha+\gamma))^{10} (1-\Lambda(\alpha+\beta))^{20}$$

Это выглядит слегка досадно для нас. Придется сплющить параметры $\alpha=\alpha$, $\delta=\alpha+\beta$, $\phi=\alpha+\gamma$. Тогда:

$$\ln L = 8\ln\Lambda(\alpha) + 28\ln\Lambda(\phi) + 22\ln\Lambda(\delta) + 12\ln(1-\Lambda(\alpha)) + \\ +10\ln(1-\Lambda(\phi)) + 20\ln(1-\Lambda(\delta))$$

Выписываем частные производные:

$$\frac{\partial \ln L}{\partial \alpha} = 8\frac{\Lambda'(\alpha)}{\Lambda(\alpha)} - 12\frac{\Lambda'(\alpha)}{1-\Lambda(\alpha)} = 0$$
$$\frac{\partial \ln L}{\partial \delta} = 22\frac{\Lambda'(\delta)}{\Lambda(\delta)} - 20\frac{\Lambda'(\delta)}{1-\Lambda(\delta)} = 0$$
$$\frac{\partial \ln L}{\partial \phi} = 28\frac{\Lambda'(\phi)}{\Lambda(\phi)} - 10\frac{\Lambda'(\phi)}{1-\Lambda(\phi)} = 0$$

Каждое уравнение снова элементарно разрешается относительно Λ:

$$\Lambda(\alpha) = \frac{2}{5}, \quad \Lambda(\delta) = \frac{11}{21}, \quad \Lambda(\phi) = \frac{14}{29}$$

Подставляем это в логарифм функции правдоподобия, получаем -64.426.

Теперь мы должны оценить модель с ограничением. Когда мы посчитаем ее, сможем провести тест отношения правдоподобия для гипотезы $\gamma = 0$. Здесь выходит, что $\phi = \alpha$:

$$\ln L = 36\ln\Lambda(\alpha) + 22\ln\Lambda(\delta) + 22\ln(1-\Lambda(\alpha)) + 20\ln(1-\Lambda(\delta))$$

$$\frac{\partial \ln L}{\partial \alpha} = 36\frac{\Lambda'(\alpha)}{\Lambda(\alpha)} - 22\frac{\Lambda'(\alpha)}{1-\Lambda(\alpha)} = 0$$
$$\frac{\partial \ln L}{\partial \delta} = 22\frac{\Lambda'(\delta)}{\Lambda(\delta)} - 20\frac{\Lambda'(\delta)}{1-\Lambda(\delta)} = 0$$

Отсюда сразу же получаем, что:

$$\Lambda(\alpha) = \frac{18}{19}, \quad \Lambda(\delta) = \frac{11}{21}$$

Здесь $\ln \tilde{L} = -67.56$.

Проводим теперь тест. Наша стаститика для сравнения $\chi^2(1)$.

$$LR = -2(\ln\tilde{L} - \ln\hat{L}) = 6.27$$

Этого нам хватает, чтобы отклонить нулевую гипотезу на уровне значимости 5%. Может возникнуть вопрос, зачем нам вообще было условие, что у нас логит-модель. Разумеется, это условие мы не использовали, а наши результаты справедливы для произвольной функции распределения.

$\boxed{\text{б}}$ Теперь сделаем то же для ЛВМ. Стоит понимать, что оценивать эту модель с помощью МНК без использования робастных ошибок не стоит.

Поступим иначе. Модель имеет вид:

$$y_i = \alpha + \beta x_i + \gamma z_i = \boldsymbol{x'\beta} + \varepsilon_i$$

В мире случаются две штуки:

$$-\boldsymbol{x'\beta} = \varepsilon_i$$

$$1 - \boldsymbol{x'\beta} = \varepsilon_i$$

Довольно ясно, что у нас всегда в жизни есть условие $E(\varepsilon_i) = 0$. Ошибки у нас принимают только эти два значения. Значит, выходит, что вероятности этих условий как раз и равны левым частям.

Значит, у нас получается условие:

$$0 \leqslant \boldsymbol{x'\beta} \leqslant 1$$

Мы обнаружили, что у нас есть симпатичная функция распределения $F(u) = \boldsymbol{x'\beta} = u$. А раз это функция распределения, то в первой части мы уже доказали все.

10.4.6 МКП. Задача 12.11

Пусть у нас есть следующий факт о данных:

n	y	x
n_1	1	1
n_2	0	1
n_3	0	0

Покажем, что тогда уравнение правдоподобие не будет иметь решения для произвольной обычной функции распределения:

$$L = F(\alpha+\beta)^{n_1}(1-F(\alpha+\beta))^{n_2}(1-F(\alpha))^{n_3}$$

Сплющим: $\gamma = \alpha + \beta$:

$$\ln L = n_1 \ln F(\gamma) + n_2 \ln(1-F(\gamma)) + n_3 \ln(1-F(\alpha))$$

Берем производные:

$$\frac{\partial \ln L}{\partial \alpha} = -n_3 \frac{f(\alpha)}{1-F(\alpha)} = 0$$

$$\frac{\partial \ln L}{\partial \gamma} = n_1 \frac{f(\gamma)}{F(\gamma)} - n_2 \frac{f(\gamma)}{1-F(\gamma)} = 0$$

И вот у любой произаичной функции распределения происходит катастрофа в первом уравнении.

10.4.7 Задача 7

Уважаемый исследователь анализирует влияние вступления в брак на уровень доходов. Он собрал данные за четыре года о тысяче работников обоих полов, часть из которых в течение рассматриваемого периода вступили в брак. В качестве объясняющей исследователь использовал фиктивную переменную Одинокий, которая равна единице для тех работников, которые в данном году не женаты (не замужем). В качестве контрольных переменных он использовал переменные Возраст (возраст в годах) и Образование (число лет обучения в годах). Исследователь оценил четыре уравнения (см. таблицу): первые два оценены обычным МНК, третье и четвертое оценены при помощи модели с фиксированными эффектами (внутригрупповое преобразование).

Модель	Модель 1	Модель 2	Модель 3	Модель 4
Метод оценивания	МНК	МНК	FE	FE
Возраст	0,80 (0,10)	0,78 (0,09)	0,67 (0,11)	0,68 (0,11)
Число лет обучения	1,20 (0,24)	1,41 (0,23)	0,91 (0,50)	0,90 (0,52)
Одинокий	−0,05 (0,02)	−0,04 (0,02)	−0,02 (0,01)	−0,03 (0,01)
Индивидуальные эффекты	Нет	Нет	Да	Да
Фиктивные переменные времени	Нет	Да	Нет	Да
Число наблюдений	4000	4000	4000	4000
R-квадрат	0,612	0,723	0,520	0,521
P-значение теста на отсутствие индивидуальных эффектов	—	—	0,001	0,002
P-значение теста на равенство нулю коэффициентов при фиктивных переменных времени	—	0,090	—	0,170

$\boxed{\text{а}}$ Поможем ему выбрать лучшую модель. Довольно очевидно, что по результатам теста на наличие индивидуальных эффектов побеждают обе модели с фиксированными эффектами. p-значения для них ниже любого разумного уровня значимости. При этом тест на значимость фиктивных переменных для разных периодов не позволяет нам отклонить нулевую гипотезу. Значит, очевидным образом лучше третья модель.

$\boxed{\text{б}}$ Возникает вопрос, есть ли все же связь между семейным положением и доходами. Сразу видно, что $t_{\text{расч}} = -2$. Критическое значение на уровне значимости 5% $t_{\text{кр}} = 1.96$. Значит, есть. Легко дать интерпретацию коэффициенту:

$$\Delta \ln W = 0.02$$

$$\frac{W_0}{W_1} = e^{0.02} = 1.02$$

Значит, неодинокие люди в среднем зарабатывают на 2% больше.

$\boxed{\text{в}}$ Какой-то менее уважаемый исследователь предлагает включить в модель фиктивные переменные пола. Сразу ясно, почему он менее уважаемый, ведь за редким исключением переменная пола будет оставаться постоянной во времени, быть частью индивидуального эффекта человека. Поэтому оценить ее влияние не представляется возможным.

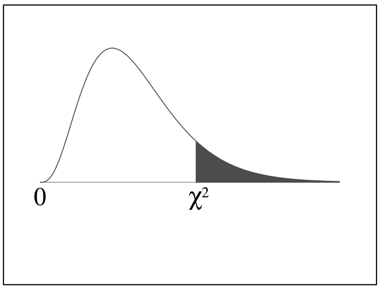
Приложение

Источники

1. **Картаев Ф.С.** : материалы лекций и семинарских занятий по эконометрике;

2. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс: Учеб. — 6 - е изд., перераб. и доп. — М.: Дело, 2004

Chi-Square Distribution Table



The shaded area is equal to α for $\chi^2 = \chi^2_\alpha$.

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Large-Sample Critical Values for the t-statistic
from the Standard Normal Distribution

	Significance Level		
	10%	5%	1%
2-Sided Test	1,64	1,96	2,58

F-распределение: критические значения F с v1 и v2 степенями свободы
уровень значимости 5%

v2 \ v1	1	2	3	4	5	6	7	8	9	10
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Процентные точки распределения Хи-квадрат для уровня значимости 5%

N	1	2	3	4	5	6	7	8	9	10
	3.84	5.99	7.81	9.49	11.07	12.59	14.07	15.51	16.92	18.31

МОСКВА