**43. Последствия, симптомы и методика устранения ошибки спецификации эконометрической модели, состоящей в неверном выборе функции регрессии.**

Нередко на этапе проверки адекватности модели выясняется, что оцененная модель оказывается неадекватной. Одной из причин этого может быть допущенная ошибка спецификации, например, пропуск значащей объясняющей переменной. Данная ошибка является частным случаем ошибки неверного выбора типа функции регрессии со всеми последствиями и симптомами.

Пусть экономист выбрал линейную функцию регрессии, в то время как адекватной является
нелинейная.

**Последствием** ошибочного выбора типа функции в уравнении регрессии оказывается нарушение основной предпосылки эконометрической модели: $E(u|x) \neq 0$. Это приводит к нарушению 1-ой предпосылки теоремы Гаусса-Маркова и смещённости оценок коэффициентов в модели, вычисленных по правилу *А* теоремы Гаусса-Маркова. Конечным следствием смещённости оценок оказывается необъективность точечных и интервальных прогнозов значений эндогенных переменных.

**Симптомы данной ошибки:**

*Первый симптом* состоит в несоответствие диаграммы рассеивания, построенной по выборке $(X,\ \vec{y})$ (обучающая выборка) графику функции

$$y = f_F(x; \vec{a}) \tag{15.2}$$

Например, ошибочно выбранная функция (15.2) являяется линейной по $x$, в то время как диаграмма рассеивания свидетельствует, что функция регрессии $E(y|x) = f_T(x)$ суть нелинейнная функция аргумента $x$.

*Второй симптом* - это длительное постоянство знака оценок случайных остатков

$$\widetilde{u}_1,\ \dots,\widetilde{u}_i,\widetilde{u}_{i+1},\ \dots,\widetilde{u}_n$$

в упорядоченных (по возрастанию значений объясняющей переменной) уравнений наблюдений. Именно этот симптом улавливается статистикой $DW$ теста Дарбина-Уотсона.

Для обнаружения *третьего симптома* необходимо разделить выборку $(X,\ \vec{y})$ на две примерно равные по количеству наблюдений части $(X_1,\ \vec{y}_1)$ и $(X_2,\ \vec{y}_2)$. Так чтобы различие в элементах $x_i$ матриц $X_1$ и $X_2$ было по возможности существенным. Затем по каждой из выборок надлежит оценить

$$\begin{cases} y \ = f_F(x; \vec{a}) + u; \\ E(u|x) = 0, E\left(u^2|x\right) \ = \sigma_u^2; \end{cases}$$

Сильное отличие одноимённых коэффициентов в двух оценённых вариантах модели и есть третий симптом неверного выбора функции регрессии.

**Методика устранения**

Допустим ошибка в выборе функции регрессии подтвердилась, в такой ситуации следует, используя диаграмму рассеивания подобрать другую подходящую фукнцию регрессии и повторить процедуру построения регрессионной модели.

**44. Последствия и симптомы ошибки спецификации линейной эконометрической модели, состоящей во включении незначимой объясняющей переменной.**

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

$$H_0 : a_1 = 0 \tag{1}$$

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

$$H_1 : a_1 \neq 0$$

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

***Последствия:*** Наличие в модели незначащих переменных увеличивает дисперсию оценок коэффициентов модели (т.е. ухудшает точность оценок).

***Симптомы:*** Основной симптом наличия в модели незначащей переменной: стандартная ошибка оценки коэффициента при этой переменной $\left(S\tilde{a}_i\right)$ находится на уровне или превышает абсолютное значение оценки коэффициента при этой переменной $\left(\tilde{a}_i\right)$.

Для определения незаначимости переменной используется $t-$тест, который позируется на следующей **теореме:**

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

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

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

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

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

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

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

Именно с таких объясняющих переменных нужно приступать к $t-$ тесту.

**Шаг 2.** Расчёт статистики $t$:

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

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

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

Если оно справедли, то гипоза $H_0 : a_j = 0$ может быть принята, как не противоречащая реальным данным и переменная $x_j$ удалена из модели, в противном случае приниматся гипотеза $H_1$ переменная $x_j$ интерпритируется, как значащая и сохраняется в моделе.
*Замечание.* Из курса математической статистики известно, что процедура проверк статистической гипотезы, может приводить к ошибкам I или II рода. Ошибка I рода - отвергнуть гипотезу $H_0$, когда она верна. Ошибка II рода - когда мы приняли гипотезу, когда она не верна. В ситуации $t-$ теста гораздо опаснее принять гипотезу $H_0$, когда она не верна и следовательно исключить значащую переменную из модели (лучше сохранить незначащую чем сохранить значащую). После удаления из модели незначащих переменных необходимо повторить все тесты в моделе.
**45. Последствия и симптомы ошибки спецификации линейной эконометрической модели, состоящей в пропуске значимой объясняющей переменной.**
Эта ошибка по последствиям и симптомам эквивалентна неверному выбору типа функции регрессии.
Пусть на первом этапе экономист составил спецификацию модели:

$$\begin{cases} y \ = a_0 + a_1 x_1 + u; \\ E(u|x_1) = 0, E\left(u^2|x_1\right) \ = \sigma_u^2; \end{cases}$$

в ситуации, когда истинной является спецификация:

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

при условии, что гипотеза $H_0 : a_0 = 0$ неверна. Это означает, что экономист сделал ошибочный выбор типа функции регрессии, а именно: вместо функции двух аргументов выбрал функцию одного аргумента. Влияние данной ошибки на случайный остаток в модели (1): $E(u|\vec{x}_1) = (a_0 + a_1 x_1 + a_2 x_2) - (a_0 + a_1 x_1) = a_2 x_2 = \phi(x_2) \neq 0$.
Действительно пропуск значащей объясняющей переменной эквивалентен неверному выбору типа функции регрессии. Если пропущенная переменная (скажем, регрессор $x_2$) недоступна для наблюдений, то экономист может включить в модель ее заместителя - такую переменную $x_2^{pr}$, которая доступна для наблюдений и

коррелирует с переменной $x_2$.

Последствия и симптомы смотри в пункте 44.

**46. Последствия и симптомы ошибки спецификации линейной эконометрической модели, состоящей в игнорировании гетероскедастичности случайного возмущения.**

Гетероскедастичность случайных возмущений – возмущения обладают различными дисперсиями, но не коррелированны друг с другом.

**Симпотомы:** при гетероскедастичности распределение $u$ для каждого наблюдения имеет нормальное распределение и нулевое ожидание, но дисперсия распределений различна.

**Последствия** гетерескедастичности случайных возмущений**:**

1. Потеря эффективности оценок коэффициентов регрессии, т.е. можно найти другие, отличные от Метода Наименьших Квадратов и более эффективные оценки;
2. Смещенность стандартных ошибок коэффициентов в связи с некорректностью процедур их оценки;

Для определения гетероскедастичности случайных возмущений используют тест Голдфилда-Кванта.

**Тест Голдфелда-Кванта гипотезы о гомоскедастичности случайного возмущения**

Шаги теста предпосылки №2 Голдфелда-Кванта

**Шаг 1.** Составляется система уравнений наблюдений объекта

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

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

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

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

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

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

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

**Шаг 4.** Вычисляется по следующему правилу дробь:

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

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

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

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

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

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

**47. Оценивание линейной модели с автокоррелированным остатком AR(1) алгоритмом Хилдрета–Лу.**

Спецификация AR(1) выглядит следующим образом:

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

Если бы нам был известен параметр авторегрессии $\rho$, то никаких проблем с оцениванием этой модели у нас бы не возникло. Однако в практике реальных исследований этот параметр

практически никогда нам не бывает известен. Следовательно, встает задача: оценить $\rho$.

Один из методов оценки является **алгоритм Хилдрета–Лу**:

Все значения параметра авторегрессии лежат в пределах $\rho \in (-1;1)$. Следовательно, идея состоит в том, чтобы из этого интервала с небольшим шагом выбирать различные значения и оценивать для каждого из них регрессию:

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

При этом следим за суммой квадратов остатков $RSS$. В качестве оценки $\rho$ берем то его значение, для которого $RSS$ минимальна.

**48. Проблема совершенной мультиколлинеарности и её выявление методом дополнительной регрессии.**

Критерий мультиколениарности:

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

На практике часто встречаются ситуации, когда упомянутый выше определитель не равен $0$, но очень мал. Тогда говорят, что в исходной модели присутствует несовершенная мультиколениарность. Отметим симптомы несовершенной мультиколениарности.

**Симптомы:**

**Симптом №1.** $F$ – тест отвергает гипотезу:

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

но $t$ – тест **не отвергает** гипотезу $H_0 : a_j = 0$ о незначимости многих $x_j$.

**Симптом № 2.** При добавлении в обучающую выборку одного уравнения (или удаления одного из уравнений), оценки коэффициентов модели резко меняют свои значения, то есть оценки неустойчивые $\tilde{a}_j$ и значит они ненадёжные.

При несовершенной мультиколлениарности часто используется процедура пошаговой регрессии отбора в модель объясняющих переменных. Процедура состоит из следующих шагов:

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

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

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

**49. Вложенные модели. Тест Вальда вложенной модели.**

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

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

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

если для модели М2 справедлива гипотеза:

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

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

Модель М2 следует предпочесть, если справедлива гипотеза 8:

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

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

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

**Шаг 2.** Оценивается модель (6) МНК и отмечается $ESS_2$.
И шаг 1 и шаг 2 осуществляется по одной и той же выборке.
**Шаг 3.** По правилу (9) вычисляется статистика критерия гипотезы $H_0$:

$$W = (ESS_1 - ESS_2) / (ESS_2 / (n - k_2)) \tag{9}$$
$$W \sim \chi_r^2 \text{ при } n \to \infty; r - \text{число}$$

Пусть справеливы все предпосылки теоремы Гаусса-Маркова при произовальном законе распределения случайных возмущений. Если справедлива гипотеза $H_0$, то статистика асимптотически (с ростом объёма выборки) имеет распределение хи-квадрат с числом степеней свободы $r$, где $r$ число ограничений (7) в модели (6).
**Шаг 4.** Гипотеза $H_0$ отвергается в пользу гипотезы $H_1$ (т.е принимается модель (6)), если значение статистики превосходит квантиль распределения $\chi_r^2$ уровня $1 - \alpha$:

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

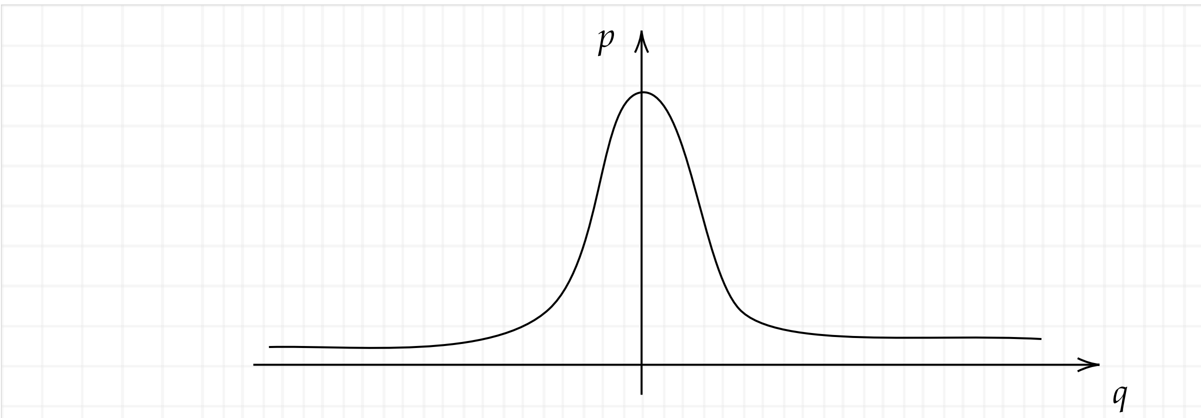
***Следствие.*** Если случайное возмущение в модели М2 имеет нормальный закон распределения, то:

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

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

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

Квантиль:




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