ТЕМА 4

Предпосылки МНК для парной линейной регрессии

Дисциплина: Эконометрика

2020-21 уч.г., 1 семестр

Группы 3ЭББ1-3ЭБФ1-3ЭБФ3

Преподаватель – Н.В. Никаноркина

После построения модели регрессии необходимо проверить выполнение предпосылок МНК, т.к. нарушение этих условий снижает качество модели.

Предпосылки МНК

- Коэффициенты регрессии, найденные исходя из системы нормальных уравнений, представляют собой выборочные оценки характеристики силы связи.
- Для практических целей важно, чтобы они были несмещенными, эффективными и состоятельными.

- МНК строит оценки регрессии на основе минимизации суммы квадратов остатков. Поэтому очень важно исследовать поведение остаточных величин регрессии ε_i.
- Исследование остатков предполагает проверку наличия пяти предпосылок, которые называются ***условиями Гаусса-Маркова.***

Теорема Гаусса-Маркова:

Оценки параметров линейной регрессии, полученные МНК, будут несмещенными и эффективными в классе линейных несмещенных оценок при выполнении ряда условий (***условий Гаусса-Маркова или предпосылок МНК***).

Модель считается адекватной, если ряд ее остатков удовлетворяет следующим требованиям:

1. Уровни ряда остатков имеют случайный характер

2. Математическое ожидание уровней ряда остатков равно нулю

3. Дисперсия каждого отклонения одинакова для всех значений

4. Значения уровней ряда остатков независимы друг от друга (отсутствует автокорреляция)

5. Уровни ряда остатков распределены по нормальному закону

Требование	Методы проверки требований
Первое	Для *проверки свойства случайности ряда ост*атков можно использовать *критерий поворотных точек (пиков)*. Точка считается поворотной, если выполняются следующие условия: $\varepsilon_{i-1} < \varepsilon_i > \varepsilon_{i+1}$ или $\varepsilon_{i-1} > \varepsilon_i < \varepsilon_{i+1}$. Далее подсчитывается число поворотных точек p. Критерием случайности с 5%-ным уровнем значимости, т.е. с доверительной вероятностью 95%, является выполнение неравенства: $p > \left[\frac{2}{3}(n-2) - 1{,}96\sqrt{\frac{16n-29}{90}}\right].$ Квадратные скобки означают, что берется целая часть числа, заключенного в скобки. Если неравенство выполняется, то модель считается адекватной

Второе	Для *проверки равенства математического ожидания остаточной последовательности нулю* вычисляется среднее значение ряда остатков $\bar{\varepsilon} = \sum(\varepsilon_i)/n$. Если $\bar{\varepsilon} \approx 0$, то считается, что модель не содержит постоянной систематической ошибки и адекватна по критерию нулевого среднего. Если $\bar{\varepsilon} \neq 0$, то проверяется нулевая гипотеза о равенстве нулю математического ожидания. Для этого вычисляют *t*-критерий Стьюдента по формуле: $$t = \frac{\lvert\bar{\varepsilon}\rvert - 0}{S_\varepsilon} \cdot \sqrt{n},$$ где S_e - стандартное отклонение остатков модели (стандартная ошибка). Значение *t*-критерий сравнивают с табличным $t_{\alpha,\gamma}$. В случае, если выполняется неравенство $t > t_{\alpha,\gamma}$, то модель неадекватна по данному критерию

Третье	Дисперсия уровней ряда остатков должна быть одинаковой для всех значений x_i (*свойство гомоскедастичности*). Если это условие не соблюдается, то имеет место *гетероскедастичность*. Для оценки гетероскедастичности, при малом объеме выборки, можно использовать *метод Гольдфельда-Квандта*, суть которого заключается в следующем: - необходимо расположить значения переменной x_i в порядке возрастания; - разделить совокупность упорядоченных наблюдений на две группы; - по каждой группе наблюдений построить уравнения регрессии; - определить остаточные суммы квадратов для первой и второй групп по формулам: $$S_1 = \sum_{i=1}^{n} \varepsilon_i^2; \qquad S_2 = \sum_{i=n_1+1}^{n_2} \varepsilon_i^2$$ - рассчитать критерий $F_{расч} = S_1 : S_2$ или $F_{расч} = S_2 : S_1$ (в числителе должна быть большая сумма квадратов). При выполнении нулевой гипотезы о гомоскедастичности критерий $F_{расч}$ будет удовлетворять F-критерию со степенями свободы $\gamma_1 = n_1 - m$ и $\gamma_2 = n - n_1 - m$ для каждой остаточной суммы квадратов (где m – число оцениваемых параметров в уравнении регрессии). Чем больше величина $F_{расч}$ превышает табличное значение F-критерия, тем более нарушена предпосылка о равенстве дисперсий остаточных величин

$d'_{1кр} < d < d_2$

Четвертое

Проверку независимости последовательности остатков (отсутствие автокорреляции) осуществляют с помощью *d*-критерия Дарбина-Уотсона. Он определяется по формуле:

$$d = \frac{\sum_{i=1}^{n} (\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^{n} \varepsilon^2{}_i}.$$

Расчетное значение критерия сравнивается с нижним d_1 и верхним d_2 критическими значениями статистики Дарбина-Уотсона.

Возможны следующие случаи:

а) если $d < d_1$, то гипотеза о независимости остатков отвергается и модель признается неадекватной по критерию независимости остатков;

б) если $d_1 < d < d_2$ (включая сами эти значения), то считается, что нет достаточных оснований сделать тот или иной вывод и следует использовать дополнительный критерий, например, *первый коэффициент автокорреляции*:

$$r_1 = \frac{\sum_{i=1}^{n} \varepsilon_i \cdot \varepsilon_{i-1}}{\sum_{i=1}^{n} \varepsilon^2{}_i}$$

Если расчетное значение коэффициента по модулю меньше табличного значения $r_{1кр}$, то гипотеза об отсутствии автокорреляции принимается; в противном случае эта гипотеза отвергается.

в) если $d_2 < d < 2$, то гипотеза о независимости остатков принимается и модель признается адекватной по данному критерию;

г) если d>2, то это свидетельствует от отрицательной автокорреляции остатков. В этом случае расчетное значение критерия необходимо преобразовать по формуле: $d' = 4 - d$ и сравнивать с критическим значением d', а не d

Пятое	*Проверку соответствия распределения остаточной последовательности нормальному закону распределения* можно осуществить с помощью *R/S*-критерия, который определяется по формуле: $$R/S=(\varepsilon_{\max}-\varepsilon_{\min})/S_{\varepsilon},$$ где S_{ε} - стандартное отклонение остатков модели (стандартная ошибка). Расчетное значение *R/S*-критерия сравнивается с табличными значениями (нижней и верхней границами данного отношения) и если значение не попадает в интервал между критическими границами, то с заданным уровнем значимости гипотеза о нормальности распределения отвергается; в противном случае гипотеза принимается

Нарушение хотя бы одного из условий Гаусса-Маркова приводит к нарушению эффективности оценок, т. е. в классе несмещённых оценок можно найти такие, которые имеют меньшую дисперсию.

Если условия нарушаются, следует модернизировать модель соответствующим образом.