

## ГЕТЕРОСКЕДАСТИЧНОСТЬ: СУЩНОСТЬ И МЕТОДЫ УСТРАНЕНИЯ

Временной ряд является одной из возможных реализаций случайного процесса. Члены ряда не являются статистически независимыми и одинаково распределенными. При проведении анализа временного ряда важно знать стационарен данный ряд или нет, так как наблюдения в стационарном ряду не зависят от времени, т. е. закон распределения и его характеристики не зависят от времени, а, следовательно, математическое ожидание и среднее квадратическое отклонение могут быть оценены по наблюдениям.

В силу того, что оценки параметров уравнения регрессии являются выборочными характеристиками, в процессе оценивания необходимо проводить статистическую проверку существенности полученных параметров.

Для оценки неизвестных параметров уравнения регрессии чаще всего используют метод наименьших квадратов, который позволяет получить несмещенные оценки.

Обобщенная линейная модель регрессии имеет вид  $Y = X\beta + \varepsilon$ .

При применении метода наименьших квадратов для нахождения оценок параметров простой множественной регрессии предполагается выполнение нескольких предпосылок, касающихся в первую очередь случайной величины  $\varepsilon$ , учитывающей ошибки измерения и ошибки спецификации. Эти предпосылки называют условиями Гаусса-Маркова.

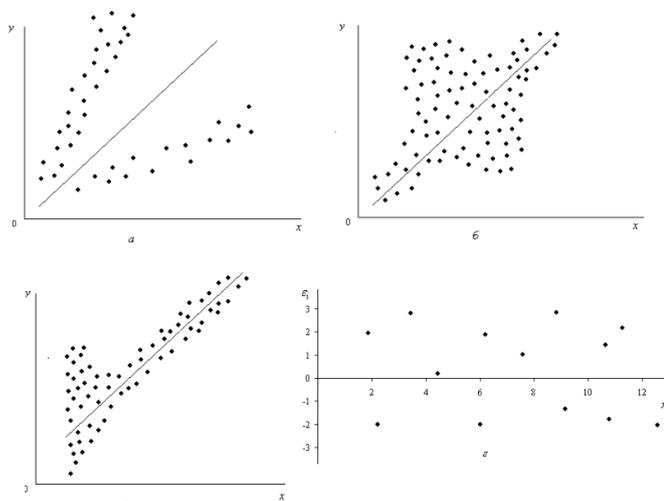
Согласно теореме Гаусса-Маркова остатки должны обладать свойствами, близкими к свойствам случайных ошибок.

Одним из нарушений этой теоремы является нарушение третьего условия Гаусса-Маркова, а именно гомоскедастичность остатков  $M\varepsilon_i^2 = \sigma^2, i = 1, \dots, n$ .

Гетероскедастичными называют остатки, для которых не выполняется условие равенства дисперсий (гомоскедастичности), при этом предполагается, что выполняются условия отсутствия автокорреляции остатков и подчинения их нормальному закону распределению. Гетероскедастичность обозначает, что исследуемые объекты неоднородны, а дисперсия случайных остатков изменяется при изменении значений независимой переменной.

Проблема гетероскедастичности зачастую характерна для пространственных наблюдений и довольно редко встречается при исследовании временных рядов.

Наличие гетероскедастичности можно наглядно продемонстрировать на графике:



При наличии гетероскедастичности последствия применения метода наименьших квадратов будут следующими:

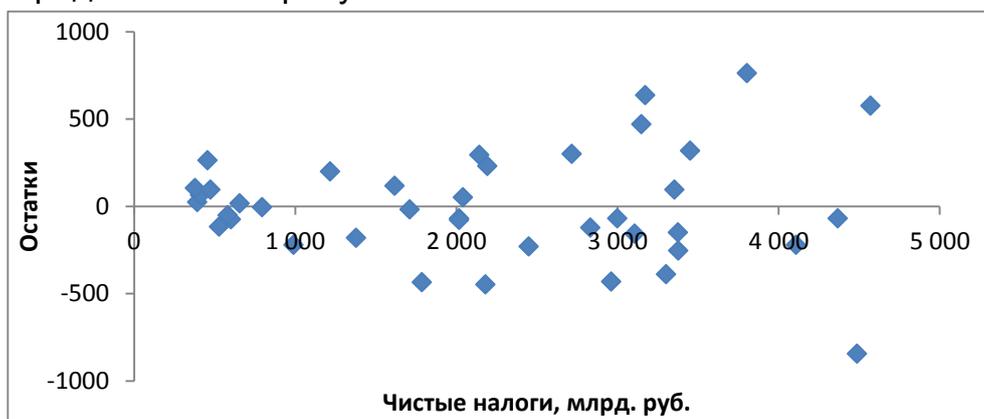
- 1) оценки коэффициентов по-прежнему останутся несмещенными и линейными;
- 2) оценки не будут эффективными (не будут иметь наименьшую дисперсию по сравнению с другими оценками такого же параметра). Вероятность получения максимально точных оценок снижается с увеличением дисперсий оценок;
- 3) дисперсии оценок будут рассчитываться со смещением;
- 4) в результате вышеперечисленного, все выводы, получаемые на основе  $t$ - и  $F$ -статистик, а также интервальные прогнозы будут ненадежными и приводить к неверным выводам по полученной модели. Стандартные ошибки коэффициентов уравнения регрессии будут занижены, а, как следствие,  $t$ -статистики завышены, что приведет к признанию статистически значимыми коэффициентов, которые на самом деле таковыми не являются.

В некоторых случаях, когда характер данных известен, появление проблемы гетероскедастичности можно предвидеть и попробовать устранить еще на этапе спецификации модели. Однако гораздо чаще эту проблему приходится решать после того, как уравнение регрессии уже построено. Для диагностики гетероскедастичности разработано довольно много тестов и критериев для них. Рассмотрим на примере более подробно некоторые из них.

В качестве исходных данных были выбраны квартальные данные о расходах на конечное потребление государственных организаций и о налоговых поступлениях в бюджет Республики Беларусь с 2009 по 2018 года.

Для множественной регрессии вид графиков остатков является наиболее наглядным способом изучения гомо- и гетероскедастичности. О наличии гетероскедастичности будет свидетельствовать величина разброса остатков относительно оси абсцисс. Если же разброс примерно одинаковый, то можно говорить о выполнении условия гомоскедастичности.

Для исследуемых данных построим график остатков. Результат этого построения представлен на рисунке 1.



**Рисунок 1 – График зависимости остатков от чистых налоговых поступлений в бюджет Республики Беларусь**

На графике остатков видно, что интервал значений разбивается примерно на три части, на каждой из которых дисперсия разброса остатков одинакова. Вывод о величине дисперсии остатков, полученный визуально на графике, следует уточнить с помощью других тестов.

Другим наиболее распространенным методом обнаружения гетероскедастичности является тест Спирмена. Рассмотрим более подробно принцип его использования. Он предполагает, что дисперсия отклонений будет увеличи-

ваться либо уменьшаться при увеличении независимой переменной  $x$ . Поэтому для построенной по методу наименьших квадратов регрессии абсолютные значения отклонений  $\varepsilon_i$  и зависимой переменной  $x_i$  будут коррелированы. Значения  $x_i$  и  $\varepsilon_i$  ранжируются, и затем определяется коэффициент ранговой корреляции:

$$r_{x,|\varepsilon|} = 1 - \frac{\sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

Далее по таблице Стьюдента находим критическое значение  $t_{\alpha, n-1}$ , где  $\alpha = \frac{1-\gamma}{2}$  – уровень значимости. Расчетное значение t-статистики вычисляется по формуле:

$$t = \frac{r_{x,|\varepsilon|} \sqrt{n-2}}{\sqrt{1-r_{x,|\varepsilon|}^2}}$$

Затем сравниваются полученные значения: если  $t < t_{\text{крит}}$ , то на уровне значимости  $\alpha$  принимается гипотеза об отсутствии гетероскедастичности, в противном случае она принимается. Если модель содержит больше одной объясняющей переменной, то проверка гипотез проводится при помощи t-статистик для каждой из них отдельно.

Вернемся к нашему примеру и проведем тест Спирмена для исходных данных. Подставляя расчеты в формулу коэффициента ранговой корреляции, получим следующее значение:

$$r_{x,|\varepsilon|} = 1 - \frac{6 \cdot 16366}{40 \cdot (40^2 - 1)} = -0,53527.$$

Зная это, найдем расчетное значение t-статистики:

$$t = \frac{\sqrt{0,53527^2 (40-2)}}{\sqrt{1-0,53527^2}} = 3,9064.$$

По таблице находим граничную точку  $t_{\alpha, n-2} = t_{0,05,40-2} = 2,0244$ . Поскольку рассчитанное значение больше критического, то с вероятностью в 95% можно утверждать о гетероскедастичности остатков модели регрессии.

Обобщая, стоит еще раз отметить, что проблему гетероскедастичности можно предвидеть и попытаться устранить еще на этапе спецификации. Однако значительно чаще эту проблему приходится решать после построения уравнения регрессии. В данной статье был рассмотрен графический метод выявления гетероскедастичности и тест Спирмена. Оба метода показали наличие гетероскедастичности в исходных исследуемых данных.

#### Список цитированных источников

1. Харин, Ю.С. Экономическое моделирование / Ю.С. Харин, В.И. Малюгин, А.Ю. Харин. – Минск: БГУ, 2003.
2. Эконометрика и экономико-математические методы и модели: учеб. пособие / Г.О. Читая [и др.] ; под ред. Г.О Читая, С.Ф. Миксюк. – Минск : БГЭУ, 2018. – 511 с.
3. Национальный статистический комитет Республики Беларусь [Электронный ресурс]. – Режим доступа: <http://belstat.gov.by/>. – Дата доступа: 06.04.2020.