ECONOMY

ГЕТЕРОСКЕДАСТИЧНОСТЬ: ЧТО ПРОИЗОЙДЕТ, ЕСЛИ ДИСПЕРСИЯ ОШИБКИ ЯВЛЯЕТСЯ НЕПОСТОЯННОЙ

Нуруллаева Ш. Т.,
 Рузметова Н. Ш.,
 Муминова М. А.,
 Сайдуллаева С. А.,

Узбекистан, Ташкент, Ташкентский государственный экономический университет

DOI: https://doi.org/10.31435/rsglobal_wos/31012019/6313

ARTICLE INFO

Received: 06 November 2018
 Accepted: 17 January 2019
 Published: 31 January 2019

KEYWORDS

heteroscedasticity,
 homoscedasticity, classical linear
 regression model, disturbance term,
 equal variance, best linear unbiased
 estimator, multiple regression
 model, discretionary income.

ABSTRACT

This article is about heteroscedasticity, it opens the nature of heteroscedasticity and what are the consequences of heteroscedasticity. Heteroscedasticity means the heterogeneity of observations, expressed in an unequal (non-constant) variance of the random error of the regression (econometric) model. Heteroscedasticity is the opposite of homoscedasticity, meaning the uniformity of observations, that is, the constancy of the variance of random model errors. The presence of heteroscedasticity of random errors leads to the ineffectiveness of estimates obtained using the method of least squares. Therefore, statistical conclusions about the quality of the obtained estimates may be inadequate. In this regard, testing models for heteroscedasticity are one of the necessary procedures when building regression models.

Citation: Нуруллаева Ш. Т., Рузметова Н. Ш., Муминова М. А., Сайдуллаева С. А. (2019) Geteroskedastichnost': chto Proizojdet, esli Dispersiya Oshibki Yavlyaetsya Nepostoyannoj. *International Academy Journal Web of Scholar*. 1(31), Vol.2. doi: 10.31435/rsglobal_wos/31012019/6313

Copyright: © 2019 Нуруллаева Ш. Т., Рузметова Н. Ш., Муминова М. А., Сайдуллаева С. А. This is an open-access article distributed under the terms of the **Creative Commons Attribution License (CC BY)**. The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Важным предположением классической модели линейной регрессии является то, что случайные ошибки u_i , возникающие в функции регрессии являются гомоскедастичными. То есть, все они имеют один и ту же дисперсию. Рассмотрим справедливость этого предположения и выясним, что произойдет, если это предположение не будет выполнено. Нас интересуют ответы на следующие вопросы:

1. Какова природа гетероскедастичности?
2. Каковы последствия гетероскедастичности?
3. Как можно её обнаружить?
4. Каковы меры по исправлению положения?

Как отмечалось, одним из важных предположений классической модели линейной регрессии является то, что дисперсия каждой случайной ошибки u_i в зависимости от выбранных значений объясняющих переменных, есть некоторое постоянное число, равное σ^2

Это предположение о гомоскедастичности, или равенстве (homo) размаха (scedasticity), то есть равенстве дисперсии.

Символично,

$$E(u_i^2) = \sigma^2 \quad i = 1, 2, \dots, n \quad (1)$$

Схематически, для двух факторной регрессионной модели гомоскедастичность можно представить, как на рисунке 1. Как видно из рисунка, условная дисперсия Y_i (которая равна u_i), относительно данной X_i остается неизменной, независимо от значений, принимаемых переменной X .

В противоположность этому, рассмотрим рисунок 2, который показывает, что условная дисперсия Y_i возрастает с увеличением X . Здесь, дисперсии Y_i не те же самые. Следовательно, существует гетероскедастичность. Символично,

$$E(u_i^2) = \sigma_i^2 \quad (2)$$

Обратите внимание на индекс σ^2 , который напоминает нам, что условные дисперсии u_i (= условные дисперсии Y_i) больше не постоянные.

Рис.1. Гомоскедастичность остатков¹

Рис.2. Гетероскедастичность остатков²

¹ Gujarati D.N. Basic Econometrics. McGraw-Hill, 4th edition, 2003 (Gu), p. 366

² Gujarati D.N. Basic Econometrics. McGraw-Hill, 4th edition, 2003 (Gu), p. 366

Для того, чтобы сделать разницу между гомоскедастичностью и гетероскедастичностью более ясной, предположим, что в двух факторной модели $Y_i = \beta_1 + \beta_2 X_i + u_i$, Y представляет собой сбережения и X представляет собой доход.

На рисунках 1 и 2 показано, что по мере увеличения доходов, сбережения в среднем также увеличиваются. Однако на рисунке 1 дисперсии сбережений остаются неизменными на всех уровнях дохода, в то время как на рисунке 2 растут с доходом. Кажется, что на рисунке 2 семьи с более высокими доходами в среднем откладывают больше, чем семьи с низким уровнем дохода, но существует к тому же большая вариабельность их сбережений.

Есть несколько причин, почему дисперсии u_i могут изменяться¹.

1. Следуя модели изучения ошибок, так как люди учатся, их ошибки поведения становятся меньше со временем или количество ошибок становится более последовательным. В этом случае, σ^2 как ожидается, снизится. В качестве примера, рассмотрим рисунок 3, связывающий количество опечаток, сделанных в определенный период времени на испытании практики печатания за несколько часов. Как показано на рисунке, количество часов практики печатания увеличивает среднее количество ошибок ввода текста, а также уменьшает их дисперсию.

2. По мере роста доходов, люди имеют больше возможностей для выбора размещения своих доходов. Следовательно, σ_i^2 , вероятно, увеличится с доходом. Таким образом, в регрессии сбережений на доходы возможно найти σ_i^2 увеличивающийся с доходом (как на рисунке 2), потому что люди имеют более широкий выбор относительно поведения их сбережений. Кроме того, в компании с большей прибылью, как правило, ожидается большая изменчивость в их дивидендной политике, чем в компании с низкой прибылью. Кроме того, ориентированные на рост компании, скорее всего, показывают большую изменчивость в отношении выплаты дивидендов, чем стабильные компании.

Рис.3. Иллюстрация гетероскедастичности

3. С улучшением техники сбора данных, σ_i^2 , вероятно, уменьшится. Таким образом, банки, имеющие сложное оборудование для обработки данных, скорее всего, совершают меньше ошибок в ежемесячных или ежеквартальных отчетах своих клиентов, чем банки без таких приспособлений.

4. Гетероскедастичность может также возникнуть в результате присутствия выбросов. Внешнее наблюдение, или постороннее, это наблюдение, которое сильно отличается (либо очень мало или очень велико) относительно наблюдений в выборке. Включение или исключение такого наблюдения, особенно если размер выборки мал, может существенно изменить результаты регрессионного анализа.

¹ Stefan Valavanis, Econometrics, McGraw-Hill, New York, 1959, p. 48

5. Другим источником гетероскедастичности является нарушение девятой предпосылки классической модели линейной регрессии (CLRM), а именно, что модель регрессии правильно указана. Очень часто выглядит так, что гетероскедастичность может быть связана с тем фактом, что некоторые важные переменные исключены из модели. Таким образом, в функции спроса на товар, если мы не будем включать цены товаров, комплементарных или конкурирующих с товаром в этом вопросе (пропущенная переменная), регрессия может дать четкое впечатление, что дисперсия ошибки не может быть постоянной. Но если опущенные переменные включены в модель, то впечатление может исчезнуть.

6. Другим источником гетероскедастичности является перекосяк в распределении одного или нескольких регрессоров включенных в модель. Примерами являются экономические переменные, такие как доход, богатство, и образование. Хорошо известно, что распределение доходов и богатства в большинстве обществ неравномерно, причем основная часть доходов и богатства принадлежит нескольким на самом верху.

7. Другие источники гетероскедастичности: Как отмечает Дэвид Хендри гетероскедастичность может также возникать из-за (1) неверного преобразования данных и (2) неправильной функциональной формы (например, линейная по сравнению с линейно-логарифмической моделью).¹

Отметим, что проблема гетероскедастичности, вероятно, будет более распространенной для перекрестных данных чем для данных временных рядов. В перекрестных данных, как правило, имеем дело с единицами совокупности в данный момент времени, такими как индивидуальные потребители или члены их семей, фирмы, отрасли или географические подразделения, такие как государства, страны, города и т.д. Кроме того, эти единицы могут быть разных размеров, такие как малые, средние или крупные фирмы или с низким, средним или высоким уровнем дохода. В данных временных рядов, с другой стороны, переменные, как правило, имеют тенденцию быть подобными порядков, так как, как правило собираются данные для одного и того же субъекта в течение периода времени. Примерами являются валовой национальный продукт (ВНП), потребительские расходы, сбережения и т.д.,

Что происходит с обычными МНК-оценками и их дисперсиями, если предположить гетероскедастичность, допустив $E(u_i^2) = \sigma_i^2$, но сохранив все остальные предположения классической модели? Чтобы ответить на этот вопрос, вернемся к модели с двумя переменными:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

Применяя обычную формулу, МНК-оценка β_2 равна

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2} \quad (3)$$

Но его дисперсия теперь определяется следующим выражением:

$$\text{var}(\hat{\beta}_2) = \frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2} \quad (4)$$

Которая, очевидно, отличается от обычной формулы дисперсии, полученной в предположении гомоскедастичности, а именно,

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \quad (5)$$

Конечно, если $\sigma_i^2 = \sigma^2$ для каждого i , обе формулы будут идентичны.

Напомним, что $\hat{\beta}_2$ является наилучшей линейной несмещенной оценкой (BLUE), если выполнены предположения классической модели, в том числе и гомоскедастичности. Является ли она все ещё BLUE, когда мы отбросим только предположение о гомоскедастичности и заменим его предположением о гетероскедастичности? Нетрудно доказать, что $\hat{\beta}_2$ все еще

¹ David F. Hendry, *Dynamic Econometrics*, Oxford University Press, 1995, p. 45.

линейна и несмещена. На самом деле, для установления несмещенности β_2 нет необходимости, чтобы возмущения u_i были гомоскедастичными. На самом деле, дисперсия u_i , гомоскедастична она или гетероскедастична, не играет роли в определении свойства несмещенности. Мы показали, что β_2 является последовательной оценкой в предположениях классической модели линейной регрессии. Хотя мы не будем доказывать это, можно показать, что β_2 является последовательной оценкой, несмотря на гетероскедастичность. То есть, когда размер выборки неограниченно возрастает, оценка β_2 сходится к его истинному значению. Разумеется, то, что мы говорили о β_2 , справедливо и для других параметров модели множественной регрессии.

Допустим, что β_2 по-прежнему является линейной несмещенной и последовательной, является ли она «эффективной» или «лучшей»? То есть, имеет ли она минимальную дисперсию в классе несмещенных оценок? И является ли минимальной дисперсия, определяемая уравнением 4? Ответом является *нет* на оба вопроса: β_2 больше не является лучшей, и минимальная дисперсия не задается уравнением. 4.

ЛИТЕРАТУРА

1. Gujarati D.N. Basic Econometrics. McGraw-Hill, 4th edition, 2003 (Gu); 5th edition (2009, Gujarati D.N., and D.C. Porter).
2. Stefan Valavanis, Econometrics, McGraw-Hill, New York, 1959
3. David F. Hendry, Dynamic Econometrics, Oxford University Press, 1995
4. Магнус Я. Р., Катышев П. К., Пересецкий А. А. Эконометрика. — М.: Дело, 2004. — 576 с.
5. William H. Greene. Econometric analysis. — New York: Pearson Education, Inc., 2003. — 1026 с.
6. Елисеева И.И., Курышева С.В. и др. Эконометрика: Учебник. —М.: Финансы и статистика, 2007. — 260 с.
7. Кремер Н.Ш. Эконометрика: Учебник. —М.: ЮНИТИ-ДАНА, 2008. — 562 с.
8. Бабешко Л.О. Основы эконометрического моделирования: Учебное пособие. —М.: КомКнига, 2010. — 520 с.
9. Гладили А.В. Эконометрика: Учебное пособие —М.: КНОРУС, 2006. — 250 с.
10. Кундышева Е.С. Математическое моделирование в экономике: Учебное пособие. /под науч. ред. проф. Б.А. Сулакова. — М.: изд. «Дашков и К°», 2006. —410 с.
11. Фандеева Л.Н., Лебедев А.В. Теория вероятностей. Учебное пособие. —М.: Эксмо, 2010. — 382 с.