Министерство образования и науки Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра Теории функций и приближений

Оценки матрицы ковариаций и тест на гетероскедастичность Уайта

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента/ки 4 курса 412 группы

направления (специальности) 01.03.02 Прикладная математика и информатика

Механико-математического факультета

Гаязова Рустама Фатыховича

Научный руководитель
Профессор, д.ф.-м.н. ____________ А.Л. Лукашов
подпись, дата

Зав. кафедрой
Доцент, д.ф.-м.н. ____________ С.П. Сидоров
подпись, дата

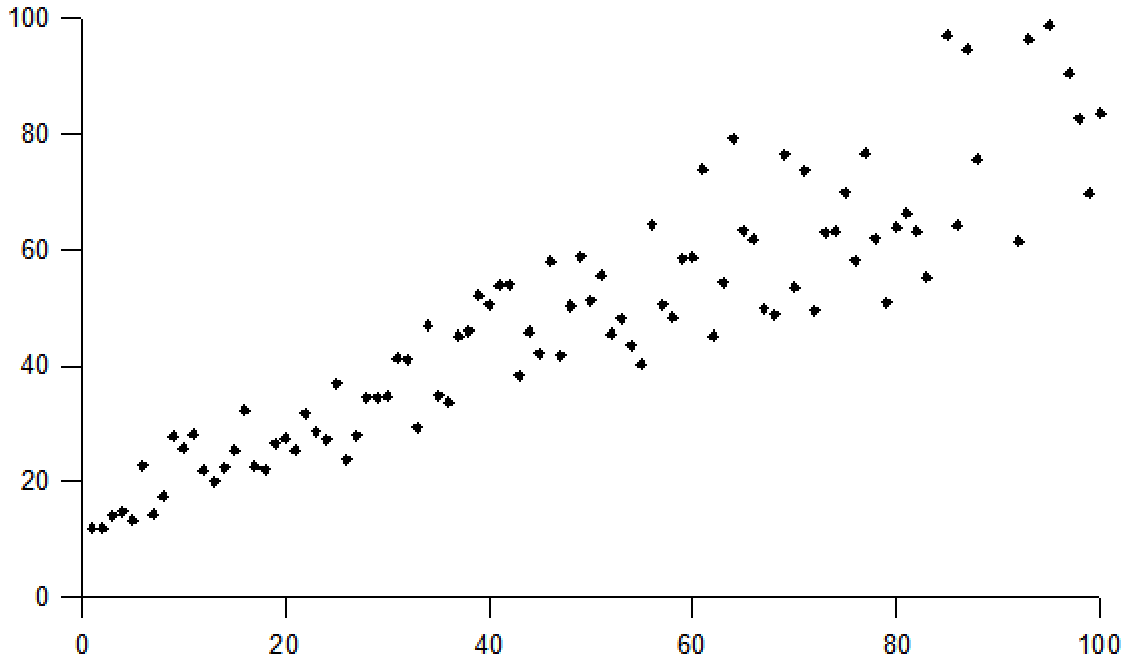
Саратов 2016

ВВЕДЕНИЕ

Экономисты используют количественные данные для наблюдения за ходом развития экономики, ее анализа и прогнозов. Набор статистических методов, используемых для этих целей, называется в совокупности эконометрикой. Для успешного применения этих методов требуется точное (или хотя бы приблизительно верное) моделирование поведения экономических агентов, необходимо также понимание процессов, породивших имеющиеся данные, и насколько эти данные отражают те явления, которые мы пытаемся исследовать. Поскольку наши модели неполны, а данные несовершенны, значительная часть эконометрики посвящена методам, которые могли бы работать с такими моделями и данными. В конце концов, эконометрика является не более чем набором инструментов, хотя и очень полезных. Качество ингредиентов (моделей и данных) и то, как мы их используем, определяют результаты нашего анализа. Но и хорошие инструменты анализа так же необходимы. Эконометрика является одновременно нашим телескопом и нашим микроскопом для изучения окружающего экономического мира.

Тема моей дипломной работы «оценки матрицы ковариаций и тест на гетероскедастичность Уайта»

Гетероскедастичность — понятие, используемое в эконометрике и регрессионном анализе, означающее неоднородность наблюдений, выражающуюся в неодинаковой (непостоянной) дисперсии случайной ошибки регрессионной (эконометрической) модели. Гетероскедастичность противоположна гомоскедастичности, означающей однородность наблюдений, то есть постоянство дисперсии случайных ошибок модели.



Термин гетероскедастичности применяется в ситуации, когда матрица ковариаций вектора ошибок является диагональной, но элементы главной диагонали, вообще говоря, различны. Иными словами, ошибки в разных наблюдениях некоррелированы, но их дисперсии — разные.

Наличие гетероскедастичности случайных ошибок приводит к неэффективности оценок, полученных с помощью метода наименьших квадратов. Кроме того, в этом случае оказывается смещённой и несостоятельной классическая оценка ковариационной матрицы МНК-оценок параметров. Следовательно, статистические выводы о качестве полученных оценок могут быть неадекватными. В связи с этим тестирование моделей на гетероскедастичность является одной из необходимых процедур при построении регрессионных моделей.

Есть много различных тестов на гетероскедастичность. Одним из самых известных является тест на гетероскедастичность Уайта.

Содержательный смысл этого теста состоит в следующем. Если в модели присутствует гетероскедастичность, то очень часто это связано с тем, что дисперсии ошибок некоторым образом (возможно, довольно сложно) зависят от регрессоров, а гетероскедастичность должна как-то отражаться в остатках обычной регрессии исходной модели. Реализуя эти идеи, Уайт предложил метод тестирования гипотезы наличия гетероскедастичности (H_0) без каких-либо предположений относительно структуры гетероскедастичности. Сначала к исходной модели применяется обычный метод наименьших квадратов и находятся остатки регрессии ε_i. Затем осуществляется регрессия квадратов этих остатков ${\varepsilon_i}^2$ на все регрессоры X, их квадраты, попарные произведения и константу, если ее не было в составе исходных регрессоров. Тогда при гипотезе H_0 величина nR^2 асимптотически имеет распределение χ^2_{N-1}, где R^2 — коэффициент детерминации, n-число наблюдений, а N — число регрессоров второй регрессии.

Привлекательной чертой теста Уайта является его универсальность. Однако если гипотеза H_0отвергается, этот тест не дает указаний на функциональную форму гетероскедастичности, и единственным способом коррекции на гетероскедастичность является применение стандартных ошибок в форме Уайта.

В 1-м разделе рассматриваются некоторые основные понятия регрессионного анализа.

Во 2-м разделе рассматривается тест на гетероскедастичность Уайта.

В 3-м разделе идет проверка теста Уайта к более сложным формам матрицы ковариации.

Целью практической части является проведение вычислительных экспериментов на предмет распознавания с помощью теста Уайта более сложных форм нарушения гомоскедастичности.

Раздел 1. Некоторые основные понятия регрессионного анализа

Регрессия определяет математическую зависимость между зависимой переменной (отклик) и одной или более независимыми переменными (предикторами).

Регрессионный анализ с помощью коэффициента регрессии позволяет количественно прогнозировать изменения одной переменной при изменении другой.

В классической регрессионной модели, использованной в эконометрике, предполагается, что $y = X\beta + \varepsilon$, модель имеет детерминированную $n \times k$ матрицу X, имеющую максимальный ранг k $(k < n)$, а вектор ошибок ε таков, что $E(\varepsilon) = 0$, а матрица ковариации пропорциональна единичной матрице. В этом случае справедлива теорема Гаусса-Маркова.

Теорема Гаусса-Маркова: Предположим, что:

1. $y = X\beta + \varepsilon$;
2. $X-$ детерминированная $n \times k$ матрица, имеющая максимальный ранг k;
3. $E(\varepsilon) = 0; V(\varepsilon) = E(\varepsilon\varepsilon') = \sigma^2 I_n$.

Тогда оценка метода наименьших квадратов $\hat{\beta}_{OLS} = (X'X)^{-1}X'y$ (1) является наиболее эффективной (в смысле наименьшей дисперсии) оценкой в классе линейных(по y) несмещенных оценок.

Большинство статистических выводов, которые получаются при использовании различных эконометрических пакетов (программ), справедливы при выполнении тех же условий Гаусса-Маркова, и, как правило, предполагается, что вектор ошибок ε имеет нормальное распределение. Примером нарушения условия теоремы Гаусса-Маркова является теорема Айткена.

В обобщенной регрессионной модели условие (3) заменяется на $E(\varepsilon) = 0, V(\varepsilon) = \Omega$, где Ω - положительно определенная матрица.

<u>Теорема Айткена:</u> В классе линейных несмещенных оценок вектора β для обобщенной регрессионной модели оценка

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

имеет наименьшую матрицу ковариаций.

Раздел 2. Тест на гетероскедастичность Уайта

Предположение 1: Имеет место модель

$Y_i = X_i\beta_0 + \varepsilon_i$ i=(1,…,n)

где $(X_i\varepsilon_i)$ представляет собой последовательность независимых (не обязательно) одинаково распределенных случайных векторов, $(X_i\varepsilon_i)$ таких, что X_i (1 x K вектор) и ε_i(скаляр) удовлетворяют $E(X_i^{'}\varepsilon_i) = 0$. ε_i ненаблюдаема тогда Y_i и X_iявляются наблюдаемыми. β_0 является конечным неизвестным вектором К x 1 параметр должен быть оценен.

Предположение 2: (a) Существуют такие положительные конечные константы δ и Δ такие

что для всех i,$E\left(\left|\varepsilon_i^2\right|^{1+\delta}\right) < \Delta$ и $E(\left|X_{ij}X_{ik}\right|^{1+\delta}) < \Delta$ (j, k = 1,…,K)

(b)$M_n = n^{-1}\sum_{i=1}^{n}(X'_iX_i)$ невырожденная для (всех) n достаточно больших, таких, что $\det M_n > \delta > 0$.

Лемма 1: Указанные в предположениях 1 и 2 $\hat{\beta}_n$ существует почти наверное для достаточно больших n и $\hat{\beta}_n \xrightarrow{a.s.} \beta_0$.

Предположение 3: (a) Существуют такие положительные конечные константы δ и Δ такие, что для всех i $E(\left|\varepsilon_i^2 X_{ij}X_{ik}\right|^{1+\delta}) < \Delta$ (j,k=1,..,K);

(b) Усредненная ковариационная матрица $\bar{V}_n = n^{-1}\sum_{i=1}^{n}E(\varepsilon_i^2 X'_iX_i)$ невырожденная для достаточно больших n, таких, что $\det V_n > \delta > 0$.

Лемма 2: Согласно предположениям 1-3,

$$\sqrt{n}V_n^{-1/2}\bar{M}_n(\hat{\beta}_n - \beta_0) \sim^A N(0, I_k).$$

Предположении 4: Существуют такие положительные константы δ и Δ такие, что для всех i

$$E\left(\left|X_{ij}^2 X_{ik} X_{il}\right|^{1+\delta}\right) < \Delta \quad \text{(j, k, l=1,..,K).}$$

<u>Теорема 1:</u> (I) $\left|\widehat{V_n} - \overline{V_n}\right| \xrightarrow{a.s.} 0$ при предположениях 1, 2, 3 (a) и 4; (II)

$$\left|(X'X/n)^{-1}\widehat{V_n}(X'X/n) - \overline{M_n^{-1}\, V_n\, M_n^{-1}}\right| \xrightarrow{a.s.} 0$$

при предположениях 1, 2, 3 (a), и 4; (III)

$$n\left(R\widehat{\beta_n} - r\right)'\left[R(X'X/n)^{-1}\widehat{V_n}(X'X/n)^{-1}R'\right]^{-1}\left(R\widehat{\beta_n} - r\right) \overset{A}{\sim} \chi_q^2,$$

при справедливости Ho и предположений 1-4.

<u>Предположение 5:</u> Существуют положительные постоянные δ и Δ такие, что для всех i

$$E\left(\left|\varepsilon_i^4\right|^{1+\delta}\right) < \Delta, \qquad и \;\; E\left(\left|X_{ij} X_{ik} X_{il} X_{im}\right|^{1+\delta}\right) < \Delta \quad (\mathrm{j,k,l,m} = 1, \dots, \mathrm{K})$$

Заметим, что предположение 5 является достаточным для предположение 4 и предположения 2(a).

Далее дадим определение следующей величине

$$\Psi_{is} \equiv X_{ik} X_{il} \qquad \text{(s=1,…,K(K+1)/2; k=1,…,K; l=1,…,k)}$$

и пусть Ψ_i будет 1 x K (K + 1) / 2 вектор с элементами Ψ_{is}. Таким образом, Ψ_i есть вектор, содержащий элементы ниже диагонали матрицы $X_i X_i$. Кроме того, определение

$$\overline{\Psi}_{ns} \equiv n^{-1} \sum_{i=1}^{n} E(\Psi_{is}) \qquad \text{(s=1,…,K(K+1)/2)}$$

<u>Предположение 6:</u> Усредненная ковариационная матрица

$$\bar{B}_n = n^{-1} \sum_{i=1}^{n} E([\varepsilon_i^2 - \sigma_i^2]^2 (\Psi_i - \overline{\Psi}_n)'(\Psi_i - \overline{\Psi}_n))$$

невырожденная для достаточно больших n, так что $\det \bar{B}_n > \delta > 0$; кроме того для достаточно больших n $n^{-1}\sum_{i=1}^{n} E([\varepsilon_i^2 - \sigma_i^2]^2) > \delta > 0$.

<u>Предположение 7:</u> Существуют положительные постоянные δ и Δ такие, что для всех i

$$E\left(\left|X_{ij}^4 \Psi_{is} \Psi_{it}\right|^{1+\delta}\right) < \Delta \quad (j = 1, \dots, K; \;\; s, t = 1, \dots, K(K+1)/2)$$

<u>Теорема 2:</u> При выполнении предположений 1, 2 (б), 3 (б), и 5-7, если ε_i не зависит от X_i и $E(\varepsilon_i^2) = \sigma_0^2$ для всех i, то

$$nD_n(\hat{\beta}_n, \hat{\sigma}_n^2)' \hat{B}_n^{-1} D_n(\hat{\beta}_n, \hat{\sigma}_n^2) \overset{A}{\sim} \chi^2_{K(K+1)/2}$$

<u>Следствие из теоремы:</u> При выполнении предположений 1, 2 (б), 3 (б), 4-6, если ε_i не зависит от X_i , и $E(\varepsilon_i^2) = \sigma_0^2$; $E(\varepsilon_i^4) = \mu_4$ для всех *I*, тогда

$$nR^2 \overset{A}{\sim} \chi^2_{K(K+1)/2}$$

где R^2 является коэффициентом детерминации (при необходимости с добавленной константой-регрессором) соответствующей регрессии.

Раздел 3. Чувствительность теста Уайта к более сложным формам матрицы ковариации

Тест Уайта рассчитан на проверку гипотезы гомоскедастичности исходной модели. На практике это означает, что выполнение теста влечет признание наличия гомоскедастичности, а его не выполнение – к наличию гетероскедастичности. Гетероскедастичность – лишь одна из множества возможных форм матрицы ковариации ошибок модели регрессии (диагональная, но не пропорциональная единичной, матрица). Цель данного раздела – проведение вычислительных экспериментов на предмет распознавания с помощью теста Уайта более сложных форм нарушения гомоскедастичности.

Генерируем случайные величины с помощью нормального закона распределения(10 величин). Из сгенерированных случайных величин мы составляем вектор. После этого умножаем этот вектор на трехдиагональную матрицу, на главной диагонали которой расположены все 3-ки, по-соседству - 1-ки, а остальные – 0. В результате получаем вектор ошибок ε.

Далее работаем по уравнению

$$Y = X\beta + \varepsilon.$$

Берем произвольные векторы X и β и находим y.

С полученными X и Y прогоняем метод наименьших квадратов и находим $\hat{\beta}_{OLS}$

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'Y.$$

Теперь прогоняем тест на гетероскедастичность Уайта:

1) Получаем остатки оцененного уравнения регрессии:

$$e_i = Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \ldots - \beta_k X_{ki} .$$

2) Оцениваем регрессии квадратов остатков на все регрессоры, их квадратов и попарных произведений:

$$e_i^2 = a_0 + \sum_{i=1}^{k} a_i X_i + \sum_{i=1}^{k} \beta_i X_i^2 + \sum_{i=1}^{k} \gamma_{ij} X_i X_j + \varepsilon_i \ .$$

3) Рассчитываем коэффициент детерминации регрессии R^2.
4) Составляем наблюдаемую статистику $\chi^2_{obs} = nR^2$. Данное значение сравнивается с табличным значением $\chi^2_{\alpha;m-1}$, где m – количество параметров в регрессии.
5) Если выполняется неравенство

$$\chi^2_{obs} > \chi^2_{\alpha;m-1},$$

то гипотеза H_0 отвергается. В противном случае H_0 принимается.

ЗАКЛЮЧЕНИЕ

В дипломной работе рассмотрен тест Уайта на гетероскедастичность. В теоретической части изложены основные результаты работы Уайта, в которых был представлен этот тест. В практической части на примерах искусственно созданных данных показано, что тест не дает ложного заключения о гомоскедастичности даже в случае матрицы ковариации, являющейся квадратом трехдиагональной матрицы.