

Рис. 2. График остатков

Список литературы

1. Новиков А.И. Эконометрика: учебное пособие. – М.: 2014. – С. 121.
2. Орлова И.В. Экономико-математическое моделирование: практическое пособие по решению задач. – 2-е издание, испр. и доп. – М.: ИНФРА-М, 2012.
3. Орлова И.В., Половников В.А. Экономико-математические методы и модели: компьютерное моделирование: учебное пособие для студентов высших учебных заведений, обучающихся по специальности «Статистика» и другим экономическим специальностям. – 3-е издание, перераб. и доп. – Серия «Вузовский учебник». – М., 2011.
4. Орлова И.В., Филонова Е.С., Агеев А.В. Эконометрика Компьютерный практикум для студентов третьего курса, обучающихся по специальностям 080105.65 «Финансы и кредит», 080109.65 «Бухгалтерский учет, анализ и аудит». – М., 2011.
5. Турундаевский В.Б. Компьютерное моделирование экономико-математических методов // Международный журнал прикладных и фундаментальных исследований. – 2014. – № 1-2. – С. 229-230.
6. Гармаш А.Н., Орлова И.В. Математические методы в управлении: учебное пособие. – М.: ИНФРА-М, 2012. – 272 с.
7. World Economic Outlook Database [Электронный ресурс]. – Режим доступа: <http://www.imf.org/external/pubs/ft/weo/2014/02/weodata/index.aspx>

ПОСТРОЕНИЕ ЭКОНОМЕТРИЧЕСКОЙ МОДЕЛИ И ИССЛЕДОВАНИЕ ПРОБЛЕМЫ ВЫЯВЛЕНИЯ И КОРРЕКЦИИ ГЕТЕРОСКЕДАСТИЧНОСТИ НА ПРИМЕРЕ ОЦЕНКИ ВАЛОВОГО РЕГИОНАЛЬНОГО ПРОДУКТА

Попова А.В., Родичева Ю.С.

Финансовый Университет при Правительстве РФ,
Москва, e-mail: anna.popova@inbox.ru

Проблема занятости – одна из наиболее обсуждаемых тем в рамках экономической дискуссии любой страны. Количество занятых в экономике во многом определяет темпы и характер её развития, ведь в конечном счёте её движущая сила – люди.

Целью нашего исследования является оценка взаимосвязи количества занятых и уровня развития экономики по субъектам Российской Федерации, проверка на однородность и рассмотрение различных способов приведения ряда к однородному в случае гетероскедастичности. В работе представлена оценка влияния количества занятых в регионах (фактор x) на валовый региональный продукт (фактор y). Основным источником данных для исследования послужила Федеральная служба государственной статистики.

Информация о валовом региональном продукте, а также данные о численности занятых представлены по 83 субъектам на 2012 год.

Перейдём к оценке гетероскедастичности. С помощью регрессии построим остатки модели. Сумма остатков получается равна нулю. Далее построим диаграмму.



Диаграмма 1. Остатки модели регрессии

С помощью визуального анализа по Диаграмме 1 можно сделать вывод, что в данной модели присутствует гетероскедастичность остатков, так как наблюдаются систематические изменения в соотношениях между \hat{y}_i и квадратами e_i^2 . Проверим наличие гетероскедастичности с помощью тестов.

I. Тест Голдфелда-Квандта. Рассчитываем F наблюдаемое по формуле:

$$F_{\text{набл}} = SS_{\text{max}} / SS_{\text{min}} \quad [4].$$

Сравниваем полученное значение с рассчитанным по формуле табличным значением

$$F_{\text{табл}}(k_1 = n_1 - m; k_2 = n - n_1 - m)$$

где n_1 – количество данных в первой выборке, m – количество рассматриваемых факторов, n – количество данных по всей выборке. По результатам проведённого теста получили следующие значения:

F набл	13,49408
F табл	1,694236

Так как наблюдаемое значение больше, чем табличное, то присутствует гетероскедастичность.

II. Тест ранговой корреляции Спирмена. Значимость коэффициента ранговой корреляции Спирмена вычисляется по формуле:

$$T_n = \rho_{xy} * \frac{\sqrt{n-2}}{\sqrt{1-\rho_{xy}^2}},$$

где $\rho_{xy} = 1 - \frac{6 * \sum d^2}{n * (n^2 - 1)}$

Одновременно с этим считаем и табличное значение:

$$t_{табл} \left(n - m - 1; \frac{\alpha}{2} \right)$$

По результатам проведённого теста получили следующие значения:

T_{набл}	2,677917
T_{табл}	1,989686

Так как наблюдаемое значение больше табличного, то присутствует гетероскедастичность.

III. Тест Глейзера. В качестве фактора возьмём значения объясняющего фактора x (численность занятых). Строим вспомогательное уравнение регрессии изменяя значения γ (...-1; -0,5; 0,5; 1...), получаем несколько четыре модели. Если хотя бы в одной модели коэффициент β окажется значимым, то в модели присутствует гетероскедастичность.

Разберём алгоритм на примере для модели

$$|e_i| = \alpha + \beta * \frac{1}{X_i} + \varepsilon_i$$

(то есть, при $\gamma = -1$):

T_{набл}	-0,225590395
T_{табл}	1,989686323

Так как наблюдаемое значение меньше табличного, то мы признаём статистическую незначимость коэффициента β . Расчёты показали, что коэффициент β статистически значим только для моделей, построенных при $\gamma = 0,5$ и при $\gamma = 1$. Это означает, что присутствует гетероскедастичность.

IV. Тест Уайта. Метод заключается в том, чтобы определить общую значимость уравнения с помощью сравнения критерия χ^2 и тестовой статистики $U = nR^2$, где n – число наблюдений. Применим к данному уравнению инструмент «Анализ данных – Регрессия» и выясним значение $R^2 = 0,834195404$. Легко подсчитать, что:

U	69,23821853
$\chi^2_{0,05,2}$	5,991464547

Так как тестовая статистика U больше критерия χ^2 , то мы отвергаем гипотезу о гомоскедастичности и говорим о наличии гетероскедастичности.

Все четыре выполненных теста подтвердили нашу первоначальную теорию, основанную на присутствии в выборке гетероскедастичности. Причинами такого явления могут являться следующие признаки:

1. Неоднородность исследуемых объектов (что как раз наблюдается в данной выборке);
2. Характер наблюдений (например, данные временного ряда).

Одним из серьёзных последствий гетероскедастичности является неточная оценка параметров мо-

дели. Методы избавления от гетероскедастичности мы выполним по следующим направлениям: кластеризация данных для борьбы с неоднородностью объектов и применение обобщённого метода наименьших квадратов для получения более точной оценки данных.

Был проведён кластерный анализ в программе VSTAT. В результате анализа были выявлены пять групп. Что не удивительно, в пятой группе оказался только один субъект, и им стал город федерального значения – Москва, с самым наибольшим показателем валового регионального продукта, а также с наибольшей численностью занятого населения. В четвёртый кластер тоже входит лишь один показатель, и это – данные по Ханты-Мансийскому автономному округу-Югра. Для дальнейшего избавления от гетероскедастичности мы будем использовать только первый кластер с более однородными данными, который содержит 61 субъект. Так же, как и для всей выборки, построим диаграмму остатков для визуального анализа.

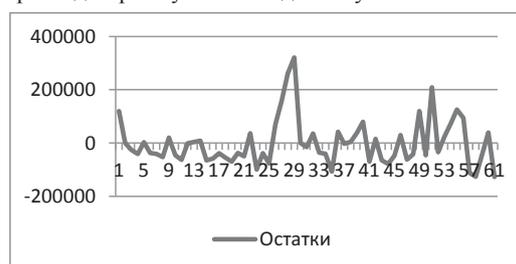


Диаграмма 2. График остатков модели (первый кластер)

Визуальный анализ показывает, что колебания всё равно остались, однако чтобы лучше понять, поможет ли нам такое разделение на группы убрать гетероскедастичность, необходимо выполнить проверку. Тест Голдфелда-Квандта, а также тест ранговой корреляции Спирмена показали присутствие гетероскедастичности. Раз не получилось избавиться от гетероскедастичности с помощью кластеризации, то можно попробовать хотя бы улучшить, сделать более точными оценки параметров регрессии с помощью обобщённого метода наименьших квадратов. Рассчитаем среднеквадратические отклонения S_{e_j} для параметров модели регрессии, которые будут считаться оценками модели. Расчёт производим по формулам:

$$S_e = \sqrt{\frac{1}{n-2} * \sum_{i=1}^n e_j^2}; S_{a0} = S_e * \sqrt{b_{00}};$$

$$S_{a1} = S_e * \sqrt{b_{11}}$$

Для наилучшего восприятия приведём посчитанные оценки в конце данных вычислений. Проведём оценку вектора A с помощью обобщённого метода наименьших квадратов. Приведём оценки модели для первого кластера, а также для модели, рассчитанной с помощью обобщённого метода наименьших квадратов.

	I	II
Se	88171,24	539,7977
Sa0	22040,51	79,83235
Sa1	36,30038	9640,581

Сравним статистические характеристики уравнений регрессий, полученных с помощью МНК и

ОМНК. Можно заметить, что второе уравнение имеет более точную оценку a_0 , однако вторая оценка a_1 менее точна.

Заключение

Вывод: ни кластерный анализ, ни обобщенный метод наименьших квадратов не помогли уйти от гетероскедастичности, а также не помогли улучшить оценки регрессии. Это может быть связано с тем, что данные, отобранные для исследования, первоначально слишком разнородные, так как все субъекты Российской Федерации слишком разные по количеству занятых и по характеру и объёму производства. Оказалось, что не все модели можно избавиться от гетероскедастичности и привести к однородности. Проведём сравнительный анализ регрессионных статистик до проведения кластерного анализа, а также после проведения процедуры в программе VSTAT с помощью метода наименьших квадратов, а также с помощью обобщенного метода наименьших квадратов.

Множественный R	0,819
R-квадрат	0,671
Нормированный R-квадрат	0,666
Стандартная ошибка	88171,239
Наблюдения	61

Множественный R	0,828
R-квадрат	0,685
Нормированный R-квадрат	0,680
Стандартная ошибка	539,798
Наблюдения	61

Рис. 1. Регрессионная статистика после кластеризации

Множественный R	0,875
R-квадрат	0,765
Нормированный R-квадрат	0,762
Стандартная ошибка	597379,392
Наблюдения	83

Рис. 2. Регрессионная статистика до кластеризации

Стоит отметить, что показатели, приведённые в таблицах не дают хорошего представления о единственной подходящей модели. Поэтому воспользуемся тестом Чоу для выяснения, можно ли нам использовать одну модель для всех данных, либо же надо применять различные модели для разных подвыборок. Первоначально выдвинута гипотеза об однородности двух подвыборок и об использовании общей модели. Для регрессионного анализа используем подвыборку, состоящую из наблюдений первого кластера, а также вторую подвыборку из оставшихся данных. Также находим остатки. Полученные результаты приведены в таблице:

	df	
Остатки общие	81	2,89058E+13
Остатки 1	59	4,58676E+11
Остатки 2	20	2,38791E+13

На основе полученных данных рассчитываем $F_{\text{наблюдаемое}}$ с помощью формулы:

$$F_{(k,n-2k)} = \frac{ESS_{\text{общ}} - (ESS_1 + ESS_2)}{ESS_1 + ESS_2} * \frac{n - 2k - 2}{k + 1}$$

Также рассчитываем $F_{\text{критическое}}$ по формуле:

$$F_{\text{крит}} = F_{\text{распоб}}(0.05; k + 1; n - 2k - 2).$$

Результаты проведённых вычислений представлены в таблице:

$F_{\text{набл}}$	7,413876
$F_{\text{крит}}$	3,11226

Наблюдаемое значение превышает критическое, это значит, что гипотеза об однородности подвыборок отвергается. В данном случае следует рассматривать отдельно первый кластер данных и остальные показатели. Итак, в данной работе не получилось привести данные к однородной выборке. Вероятнее всего, это произошло из-за того, что на валовый региональный продукт влияет множество факторов помимо экономических, которые невозможно охватить. Конечно, влияние численности занятых тоже очень важно. Ведь среди факторов ясно прослеживается тенденция: чем больше работающих, тем выше региональный продукт. Безусловно, данная проблема требует дальнейшего изучения и рассмотрения с помощью различных эконометрических методов.

Список литературы

1. Бабешко Л.О. Основы эконометрического моделирования: учебное пособие. – М.: КомКнига, 2006.
2. Бобков В., Васильев В., Гулюгина А., Одинцова Е., Смирнов М. Качество и уровень жизни населения: территориальный разрез // Экономист. – 2009. – №1. – С. 27-37.
3. Кара-оол Ш.В. Бюджет субъекта Российской Федерации в социально-экономическом развитии региона // Финансы. – 2009. – №2. – С. 9-13.
4. Орлова И.В., Половников В.А. Экономико-математические методы и модели: компьютерное моделирование: учебное пособие для студентов высших учебных заведений, обучающихся по специальности «Статистика» и другим экономическим специальностям. – 3-е издание, перераб. и доп. Серия «Вузовский учебник». – М., 2011.
5. Орлова И.В. Экономико-математическое моделирование: практическое пособие по решению задач. – 2-е издание, испр. и доп. – М.: ИНФРА-М, 2012.
6. Турундаевский В.Б. Компьютерное моделирование экономико-математических методов // Международный журнал прикладных и фундаментальных исследований. – 2014. – № 1-2. – С. 229-230.

АНАЛИЗ ФАКТОРОВ, ВЛИЯЮЩИХ НА КУРС РУБЛЯ; СВЯЗЬ МЕЖДУ ТОРГОВЫМ БАЛАНСОМ РФ И ЦЕНОЙ НА НЕФТЬ; ПРОГНОЗИРОВАНИЕ МЕЖБАНКОВСКОЙ СТАВКИ НЕДЕЛЬНОГО ДЕПОЗИТА

Рассказов В.Е.

Финансовый Университет при Правительстве РФ,
Москва, e-mail: rasskazov.vladislav@gmail.com

Какие факторы влияют на курс рубля, как они соотносятся между собой, и возможно ли, вообще, работать с макроэкономическими показателями или же их так много, и они так взаимосвязаны между собой, что они не поддаются никакому эконометрическому моделированию? Именно на эти вопросы я попробую ответить в своей работе.

Исследование по данным вопросам в процессе работы сложилось в несколько отдельных этапов:

I) определение необходимых факторов и сбор данных

II) анализ исходного массива собранных данных

III) создание модели парной регрессии по торговому балансу РФ и ценой на нефть марки Urals

IV) создание модели множественной регрессии по факторам, влияющим на межбанковскую ставку недельного депозита

I. Первоначальной задачей исследования было выстроить модель для прогнозирования изменения курса рубля, и, как бы наивно это не было, было инте-