

К ВОПРОСУ О ПОСЛЕДСТВИЯХ НАЛИЧИЯ И МЕТОДАХ УСТРАНЕНИЯ ГЕТЕРОСКЕДАСТИЧНОСТИ И АВТОКОРРЕЛЯЦИИ В РЕГРЕССИОННЫХ МОДЕЛЯХ*

УДК 338.27

Нина Леонидовна Тимофеева,
к.т.н., заведующий научно-исследовательской лабораторией Московского государственного университета экономики, статистики и информатики (МЭСИ)
Тел.: 8 (906) 086-75-61;
Эл. почта: NLTimofeeva@mes.i.ru

Артем Игоревич Федосеев,
к.э.н., начальник отдела стратегии развития и координации научной и инновационной деятельности Московского государственного университета экономики, статистики и информатики (МЭСИ)
Тел.: 8 (916) 510-48-33;
Эл. почта: AIFedoseev@mes.i.ru

В статье излагаются методологические и методические основы построения регрессионных моделей, адекватно отражающих действительность. Основное внимание сосредоточено на методах устранения автокорреляции остатков моделей. Приводятся алгоритмы устранения гетероскедастичности и автокорреляции остатков регрессионной модели: взвешенный метод наименьших квадратов, метод Кохрана-Оркутта. Построены модели «чистой» регрессии, а также, для сравнения влияния на зависимую переменную различных объясняющих переменных, когда последние выражаются разными единицами измерения, уравнение регрессии в стандартизированной форме. Разработана схема применения методов устранения гетероскедастичности и автокорреляции для построения регрессионных моделей характерных для социально-культурной сферы.

Ключевые слова: социально-культурная сфера, регрессионные модели, эффективные управленческие решения, гетероскедастичность модели, автокорреляция остатков модели, взвешенный метод наименьших квадратов.

Nina L. Timofeeva,
PhD, Head of Research laboratory, Moscow State University of Economics, Statistics and Informatics (MESI)
Тел.: 7 (906) 086-75-61;
E-mail: NLTimofeeva@mes.i.ru

Artem I. Fedosov
PhD, Head of the Strategy Development and Coordination of research and innovation Department, Moscow State University of Economics, Statistics and Informatics (MESI)
Тел.: 8 (916) 510-48-33;
E-mail: AIFedoseev@mes.i.ru

ON THE EFFECTS OF THE PRESENCE AND METHODS OF THE ELIMINATION HETEROSCEDASTICITY AND AUTOCORRELATION IN THE REGRESSION MODEL

The article presents the methodological and technical bases for the creation of regression models that adequately reflect reality. The focus is on methods of removing residual autocorrelation in models. Algorithms eliminating heteroscedasticity and autocorrelation of the regression model residuals: reweighted least squares method, the method of Cochran-Orkutta are given. A model of "pure" regression is build, as well as to compare the effect on the dependent variable of the different explanatory variables when the latter are expressed in different units, a standardized form of the regression equation. The scheme of abatement techniques of heteroskedasticity and autocorrelation for the creation of regression models specific to the social and cultural sphere is developed.

Keywords: social and cultural sphere, regression models, effective management decisions, heteroscedasticity model, autocorrelation model, the reweighted least squares method.

* Исследование выполнено при финансовой поддержке РГНФ в рамках проектов проведения научных исследований («Управление инвестиционным потенциалом коммерческого банка» и «Методология формирования кадрового потенциала социально-культурной сферы»), проекты № 12-02-00398 и № 13-02-00110.

1. Введение

Социально-культурная сфера является одной из ключевых сфер жизнедеятельности человека. Процессы, происходящие в этой сфере, напрямую связаны с процессами реформирования экономики. Одним из значительных факторов развития социально-культурной сферы является банковская деятельность. Изучение банковской деятельности с помощью экономико-математических методов и моделей, позволяет давать максимально адекватную оценку экономическим тенденциям, которые в свою очередь оказывают значительное влияние на всю социально-культурную сферу. Искусство принятия эффективных управленческих и инвестиционно-финансовых решений, прогнозирование последствий принятия тех или иных экономических решений во многом зависит от качественно построенных регрессионных моделей по различным видам деятельности, в том числе банковской.

Для того чтобы модели регрессионного анализа давали наилучшие из всех возможных результатов, на исходные данные модели накладывается ряд условий, называемых предпосылками регрессионного анализа. Если зависимая переменная, ошибки и объясняющие переменные регрессионной модели удовлетворяют предпосылкам регрессионного анализа, то выборочные оценки параметров модели соответствуют ряду требований, а именно, несмещенности, эффективности и состоятельности. При нарушении предпосылок регрессионного анализа в модели могут возникнуть такие явления, как мультиколлинеарность, гетероскедастичность или автокорреляция. Что может привести к тому, что параметры регрессионной модели потеряют экономический смысл, оценки параметров уравнения регрессии, полученные методом наименьших квадратов, станут не эффективными, случайные ошибки в регрессионной модели окажутся зависимыми.

Таким образом, для того, чтобы регрессионную модель можно было бы считать адекватно отражающую действительность, и на основе этой модели строить различные аналитические прогнозы, одним из наиболее важных аспектов при построении регрессионных моделей является проверка модели на наличие мультиколлинеарности, гетероскедастичности и автокорреляции, а также устранения последних.

2. Построение регрессионных моделей по статистически значимым факторным признакам

При работе над проектами РГНФ «Управление инвестиционным потенциалом коммерческого банка» и «Методология формирования кадрового потенциала социально-культурной сферы» (Номера проектов: №12-02-00398; №13-02-00110) были построены регрессионные модели инвестиционных ресурсов и модели кредитной активности четырех кредитных организаций (ОАО КБ «Зенит», ОАО КБ «Восточный экспресс», ООО КБ «Роспромбанк», ЗАО КБ «ЭкспрессВолга»), представляющих различные группы кредитных институтов, как по величине активов, так и по территориальному размещению.

В качестве факторов, влияющих на инвестиционные ресурсы банка Уинв, были выбраны: X1 – величина активов банка (тыс. руб.), X2 – ставка резервирования (%), X3 – ставка отчислений в агентство по страхованию вкладов АСВ (%), X4 – инфляция в месяц (%), X5 – ставка процента по депозиту сроком от 1 года до 3 лет (%), X6 – ставка процента по депозиту сроком свыше 3 лет (%), X7 – межбанковская ставка привлечения МІАСР (%), X8 – среднедушевые доходы населения (млрд. руб.),

Таблица 1

Корреляционная матрица кредитной активности ЗАО КБ «ЭкспрессВолга»

	Укред	X1	X2	X3	X4	X5	X6	X7	X8
Укред	1,00	0,30	0,77	0,61	-0,65	0,10	0,82	-0,95	-0,44
X1	0,30	1,00	0,29	0,21	-0,44	0,25	0,30	-0,35	-0,21
X2	0,77	0,29	1,00	0,60	-0,48	0,19	0,94	-0,65	-0,31
X3	0,61	0,21	0,60	1,00	-0,64	-0,38	0,75	-0,68	-0,87
X4	-0,65	-0,44	-0,48	-0,64	1,00	0,12	-0,59	0,75	0,64
X5	0,10	0,25	0,19	-0,38	0,12	1,00	0,06	0,00	0,51
X6	0,82	0,30	0,94	0,75	-0,59	0,06	1,00	-0,75	-0,48
X7	-0,95	-0,35	-0,65	-0,68	0,75	0,00	-0,75	1,00	0,63
X8	-0,44	-0,21	-0,31	-0,87	0,64	0,51	-0,48	0,63	1,00

X9 – количество филиалов банка (шт.), X10 – ставка рефинансирования (%). А в качестве факторов, влияющих на кредитную активность Укред: X1 – инфляция в месяц (%), X2 – ставка процента по кредиту сроком от 1 года до 3 лет (%), X3 – ставка процента по кредиту сроком свыше 3 лет (%), X4 – среднедушевые доходы населения (млрд. руб.), X5 – количество филиалов банка (шт.), X6 – ставка рефинансирования (%), X7 – просроченная задолженность (тыс. руб.), X8 – кредитный портфель (тыс. руб.), X9 – рентабельность активов нетто (%).

Проверим наличие корреляции по всем факторам, первоначально включенным в модели. В таблице 1 представлены данные корреляционной матрицы кредитной активности ЗАО КБ «ЭкспрессВолга».

В качестве факторов, вошедших в регрессионные уравнения, выбраны не только значимые по *t*-критерию

Стьюдента, но и те, коэффициент корреляции между которыми не превышает 0,3, т.е. не коррелируемые. Из этого следует, что в регрессионных моделях, представленных в проекте, мультиколлинеарность устранена.

Полученные на этом этапе регрессионные модели, представлены в таблицах 2, 3.

Так как мультиколлинеарность удалось устранить наиболее простым методом – методом исключения, то построение «ридж-регрессии» или гребневой регрессии не потребовалось.

Все модели были проверены на наличие гетероскедастичности. В тех моделях, где гетероскедастичность была выявлена, для ее устранения был применен взвешенный метод наименьших квадратов. При применении взвешенного метода наименьших квадратов каждое наблюдение как бы «взвешивается» с помощью коэффициента $1/\sigma_i$, где σ_i – дисперсии ошибок,

но если значения σ_i неизвестны, как в нашем случае, то, сначала находим оценку вектора параметров обычным методом наименьших квадратов. Затем находим регрессию квадратов остатков на квадратичные функции объясняющих переменных, т.е. уравнение

$$e_i^2 = f(x_i) + u_i, \quad i = 1, \dots, n,$$

где $f(x_i)$ – квадратичная функция. Далее по полученному уравнению рассчитываем теоретические значения \hat{e}_i^2 и определяем набор весов $\sigma_i = \sqrt{\hat{e}_i^2}$. Затем вводим новые переменные $Y_i^* = Y/\sigma_i$, $X_j^* = X_j/\sigma_i, j = 1, \dots, m$ и находим уравнение $\hat{Y}^* = X^*b^*$. Полученная оценка и есть оценка взвешенного метода наименьших квадратов. Для новой модели дисперсия остатков равна единице, т.е. она постоянна и поэтому, в силу построения, модель гомоскедастична.

3. Исследование регрессионной модели на примере ЗАО КБ «ЭкспрессВолга»

Было построено регрессионное уравнение Укред от всех рассматриваемых факторов. Из полученных результатов можно заключить, что модель адекватна на 97,8%, т.к. коэффициент детерминации $R^2 = 0,978$, модель значима ($F_{расч} = 220 > F_{табл} = 2,1$). Приведём полученное уравнение модели:

$$\begin{aligned} \text{Укред} = & -0,45 - 0,001 * X1 + 0,01 * X2 + \\ & \quad \quad \quad (-0,15) \quad \quad (1,72) \\ & + 0,02 * X3 + 0,000001 * X4 - 0,01 * X5 + \\ & \quad \quad \quad (3,05) \quad \quad (1,09) \quad \quad (-1,16) \\ & + 0,0001 * X6 - 0,0000003 * X7 + \\ & \quad \quad \quad (0,02) \quad \quad (-22,33) \\ & + 0,00000003 * X8 - 0,002 * X9 + e \\ & \quad \quad \quad (6,73) \quad \quad (-1,24) \end{aligned}$$

Таблица 2

Модели для исследования инвестиционных ресурсов банков

№ п/п	Банк	Регрессионные уравнения, представляющие модели инвестиционных ресурсов банков	Адекватность
1	ОАО КБ «Зенит»	$Y_{инв} = 155224569 + 4636704 * X2 - 698916511 * X3 + 3529868 * X4 + 610 * X8 - 3830321 * X9 + e$	91,8%
2	ОАО КБ «Восточный экспресс»	$Y_{инв} = -1763365 + 0,38 * X1 + 108323694 * X3 - 6374083 * X5 + 5701568 * X6 + 5244751 * X7 - 4457973 * X10 + e$	98,2%
3	ООО КБ «Роспромбанк»	$Y_{инв} = 1211115 + 0,09 * X1 + 816241 * X6 - 659305 * X9 + e$	56,4%
4	ЗАО КБ «ЭкспрессВолга»	$Y_{инв} = 5271019 + 0,68 * X1 + 1801023 * X5 - 1364758 * X6 + 1316353 * X9 - 1163129 * X10 + e$	99,5%

Таблица 3.

Модели для исследования кредитной активности банков

№ п/п	Банк	Регрессионные уравнения, представляющие модели кредитной активности банков	Адекватность
1	ОАО КБ «Зенит»	$Y_{кред} = -1,18 + 0,01 * X1 + 0,03 * X2 + 0,02 * X3 + 0,000000004 * X8 + e$	86,5%
2	ОАО КБ «Восточный экспресс»	$Y_{кред} = 0,95 - 0,03 * X5 - 0,000000002 * X8 + 0,01 * X9 + e$	67,1%
3	ООО КБ «Роспромбанк»	$Y_{кред} = 0,57 - 0,05 * X1 + 0,04 * X2 - 0,04 * X3 - 0,0000001 * X7 - 0,00000004 * X8 - 0,01 * X9 + e$	59%
4	ЗАО КБ «ЭкспрессВолга»	$Y_{кред} = -0,49 + 0,01 * X2 + 0,02 * X3 - 0,0000003 * X7 + 0,00000003 * X8 + e$	97,4%

В скобках указаны значения *t*-статистика для коэффициентов регрессии.

Так при уровне значимости $\alpha = 0,1$ значимыми оказались коэффициенты при X2, X3, X7 и X8, остальные факторы должны быть исключены из дальнейшего рассмотрения.

Перед построением модели регрессии от всех значимых факторов необходимо построить корреляционную матрицу и оценить их взаимосвязь.

Построим уравнение регрессии от всех статистически значимых факторов. Полученная модель адекватна на 97,4%, и значима, все входящие в модель факторы значимы при $\alpha = 0,05$. В качестве факторов, вошедших в регрессионные уравнения, выбраны не только значимые по *t*-критерию Стьюдента, но и те, коэффициент корреляции между которыми не превышает 0,3, т.е. не коррелируемые, что свидетельствует об отсутствии в модели мультиколлинеарности. Уравнение модели:

$$Y_{\text{кред}} = -0,49 + 0,01 \cdot X_2 + 0,02 \cdot X_3 - 0,0000003 \cdot X_7 + 0,0000003 \cdot X_8 + e.$$

(2,63) (3,88) (-25,08) (6,86)

В скобках указаны значения *t*-статистика для коэффициентов регрессии.

После построения уравнения регрессии и проверки на адекватность и значимость, следует провести проверку остатков модели на наличие автокорреляции. В связи с тем, что наличие автокорреляции между остатками модели может привести к негативным результатам всего процесса оценивания неизвестных параметров модели, автокорреляция остатков должна быть устранена.

Вспользуемся для определения наличия автокорреляции критерием Дарбина-Уотсона, критериальное расчётное значение составило $d_{\text{расч}} = 1,08$. По таблице Дарбина-Уотсона находим пороговые значения d_n и d_v . При длине реализации $n = 53$ и числе факторов $m = 4$, имеем $d_n = 1,41$, $d_v = 1,72$. Так как расчётное значение попадает в интервал $0 < d_{\text{расч}} < d_n$, то в соответствии с критерием Дарбина-Уотсона имеем наличие положительной автокорреляции (рис. 1).

Но прежде чем устранять автокорреляцию, необходимо провести проверку на наличие гетероскедастичности. В соответствии со схемой теста Голдфелда-Кванда, упорядочивание было проведено по фактору X2. Получили $d_{\text{расч}} = 4,59 > d_{\text{табл}} = 2,48$,

что свидетельствует о наличии гетероскедастичности. Для ее устранения был применен взвешенный метод наименьших квадратов.

Следует проверить уравнение регрессии, полученное после устранения гетероскедастичности, на наличие автокорреляции остатков по критерию Дарбина-Уотсона. Получили $d_{\text{расч}} = 0,75$, т.е. $0 < d_{\text{расч}} < d_n$, что свидетельствует о наличии положительной автокорреляции. К сожалению, применение взвешенного метода наименьших квадратов не позволило устранить автокорреляцию.

Т.к. в процессе исследования была выявлена автокорреляция между соседними уровнями, то для ее устранения была построена авторегрессионная схема первого порядка. По рекуррентной формуле получаем преобразованную модель регрессии, учитывающую процесс автокорреляции первого порядка:

$$y_i - \rho y_{i-1} = \beta_0(1 - \rho) + \sum_{i=1}^m \beta_i(x_{it} - \rho x_{it-1}) + v_i,$$

где выборочный коэффициент остатков первого порядка рассчитывается по формуле:

$$\rho = \frac{\sum_{i=2}^n e_i e_{i-1}}{\sum_{i=1}^n e_i^2}$$

а первые наблюдения вычисляем с помощью поправки Прайса-Уинстена:

$$y_1^* = \sqrt{1 - \rho^2} y_1; x_1^* = \sqrt{1 - \rho^2} x_1,$$

Используя вышеприведённые преобразования получены новые исходные данные по методу Кохрана-

Оркутта, позволяющему устранять автокорреляцию первого порядка.

Проверим на наличие автокорреляции используя тест Дарбина-Уотсона, получаем $d_{\text{расч}} = 1,45$. Попадаем в интервал $d_n < d_{\text{расч}} < d_v$, в соответствии с критерием Дарбина-Уотсона – это зона неопределенности. Для ответа на вопрос о наличии автокорреляции надо применять другой критерий, например, критерий Бреуша-Годфри. В соответствии со схемой теста Бреуша-Годфри построено лаговое уравнение регрессии остатков от четырех предыдущих уровней:

$$e_i = -0,0006 + 0,2e_{i-1} + 0,2e_{i-2} - 0,06e_{i-3} - 0,04e_{i-4} + u.$$

(1,37) (1,15) (-0,38) (-0,26)

В скобках указаны значения *t*-статистика для коэффициентов регрессии.

Из представленных результатов видно, что модель статистически не значима ($R^2 = 0,08$; $F_{\text{расч}} = 1,02 < F_{\text{табл}} = 1,9$), и не имеет ни одного значимого коэффициента, что в свою очередь означает, что в построенной авторегрессионной модели первого порядка, автокорреляция отсутствует (рис. 2) и коэффициенты модели, отвечают требованиям несмещенности, эффективности и состоятельности.

Уравнение регрессионной модели кредитной активности ЗАО КБ «ЭкспрессВолга», получившиеся после устранения гетероскедастичности и автокорреляции³:

$$Y_{\text{кред}} = -0,22 + 0,01 \cdot X_2 + 0,008 \cdot X_3 - 0,0000003 \cdot X_7 + 0,0000002 \cdot X_8 + e.$$

(2,89) (2,42) (-23,36) (6,76)

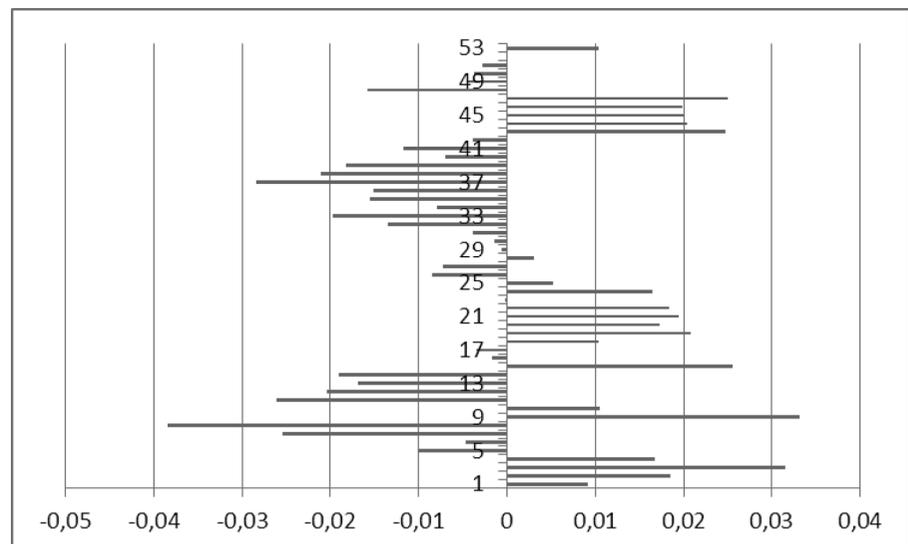


Рис. 1. Распределение остатков модели, с характерным наличием положительной автокорреляции

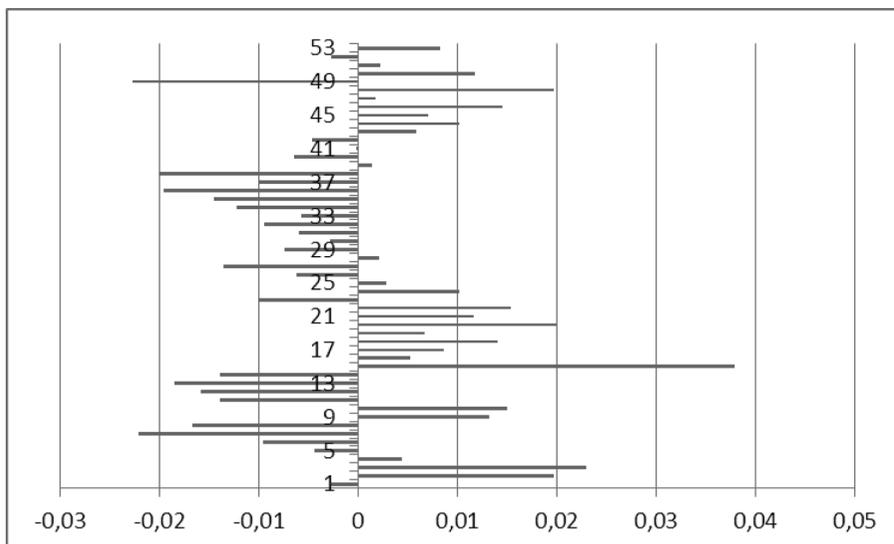


Рис. 2. Распределение остатков модели, автокорреляция отсутствует.

В скобках указаны значения t -статистика для коэффициентов регрессии.

4. Построение уравнение регрессии кредитной активности ЗАО КБ «ЭкспрессВолга» в стандартизованной форме

Так как факторы, которые входят в систему показателей имеют различные единицы измерения, то для того, чтобы сравнивать их влияние на зависимую переменную, необходимо использовать стандартизованные коэффициенты регрессии β_i ($i = 1, 2, \dots, n$).

Уравнение регрессии в стандартизованной форме:

$$t_y = \beta_1 \cdot t_{x1} + \beta_2 \cdot t_{x2} + \beta_n \cdot t_{xn} + \varepsilon,$$

где $t_y = \frac{y - \bar{y}}{\sigma_y}$, $t_{x_i} = \frac{x_i - \bar{x}_i}{\sigma_{x_i}}$ – стандартизованные переменные.

Сравнивая стандартизованные коэффициенты друг с другом, можно ранжировать факторы по силе их воздействия на результат. Наибольшее влияние оказывает фактор с наибольшим коэффициентом, взятым по модулю. Стандартизованные коэффициенты показывают, на сколько сигм изменится в среднем результат, если соответствующий фактор t_{x_i} изменится на одну сигму при неизменном среднем уровне других факторов.

Уравнение регрессионной модели кредитной активности ЗАО КБ «Экс-

прессВолга», полученное в стандартизованной форме:

$$tY_{\text{кред}} = 0,11 \cdot tX_2 + 0,27 \cdot tX_3 - 0,97 \cdot tX_7 + 0,43 \cdot tX_8 + e.$$

(2,98) (2,52)
(-26,03) (7,13)

В скобках указаны значения t -статистика для коэффициентов регрессии.

5. Заключение

Приведённая методика позволяет с большой степенью достоверности устранять гетероскедастичность и автокорреляцию в регрессионных моделях характерных для социально-культурной сферы, в таких её аспектах как: торговля и туризм, банковская деятельность и финансы, наука и образование, культура и спорт, и т.д. При этом полученные модели будут способны адекватно отражать действительность.

Литература

1. Кремер Н.Ш., Путко Б.А. Эконометрика. Учебник. М.: ЮНИТИ, 2008.
2. Шелобаев С.И. Экономико-математические методы и модели. Учебное пособие для вузов. Издание 2. М: ЮНИТИ-ДАНА, 2005.

References

1. Kremer N.S., Putko B.A. Econometrics. Textbook. Moscow: UNITY 2008.
2. Shelobaev S.I. Economic and mathematical methods and models. Textbook for high schools. Edition 2. M: UNITY-DANA, 2005.