

ЗАДАЧНИК

СБОРНИК ЗАДАЧ по эконометрике временных рядов и панельных данных

Ф. С. Картаев, О. А. Клачкова,
В. М. Ромашова, О. В. Сучкова



Экономический
факультет
МГУ
имени
М.В. Ломоносова

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М. В. Ломоносова
Экономический факультет



Ф. С. Картаев, О. А. Клачкова,
В. М. Ромашова, О. В. Сучкова

СБОРНИК ЗАДАЧ
по эконометрике временных рядов
и панельных данных

Москва
2016

УДК 330.43
ББК 65в631
К27

Каргаев Ф. С., Клячкова О. А., Ромашова В. М., Сучкова О. В.
К27 Сборник задач по эконометрике временных рядов и панельных данных. — М.: Экономический факультет МГУ имени М. В. Ломоносова, 2016. — 64 с.

ISBN 978-5-906783-48-6

Сборник задач подготовлен сотрудниками кафедры математических методов анализа экономики экономического факультета МГУ имени М. В. Ломоносова и предназначен для студентов учебных курсов бакалавриата и магистратуры. Материалы сборника позволят студентам приобрести навык решения задач продвинутого уровня по курсу эконометрики, развить понимание следующих тем дисциплины эконометрика: панельные данные, временные ряды.

Сборник содержит не только условия задач и ответы к ним, но и подробные решения значительной их части, что позволит осуществлять эффективную самостоятельную подготовку.

ISBN 978-5-906783-48-6

© Экономический факультет
МГУ имени М. В. Ломоносова, 2016

СОДЕРЖАНИЕ

Предисловие	4
Задача 1. ARIMA	5
Задача 2. Динамика реальных доходов	15
Задача 3. Реклама и продажи	20
Задача 4. Коинтеграция	25
Задача 5. Лисы и кролики	28
Задача 6. Кривая Филлипа	32
Задача 7. Волатильность инфляции и рост экономики	35
Задача 8. Стоимость авиабилетов	38
Задача 9. Списывание	46
Задача 10. Заработная плата в Мексике	56
Задача 11. Государственный долг и экономический рост	60
Литература	65

ПРЕДИСЛОВИЕ

Данное пособие содержит задания по эконометрике временных рядов и панельных данных. Задания охватывают следующие темы: модель ARIMA, обобщенная модель авторегрессионной условной гетероскедастичности (GARCH), авторегрессионная модель распределенных лагов (ADL), модель векторной авторегрессии (VAR), модели с фиксированными и случайными эффектами, а также метод «разность разностей».

Овладение методами прикладной эконометрики невозможно без практической работы с данными. Поэтому задачник будет полезен студентам, изучающим соответствующие темы, а также преподавателям эконометрики. Все задания сопровождаются подробным решением.

Все расчеты для решения заданий осуществлялись в эконометрическом пакете Gretl. Поэтому все таблицы и графики, приведенные в работе, получены в результате использования этого пакета. Впрочем, это ни к чему не обязывает читателя, который волен решать задачи, используя любое удобное ему программное обеспечение.

В конце задачника приведен список литературы, в которой описаны все нужные для решения представленных задач методы и модели.

Архив с данными, необходимыми для расчетов, доступен в сети по адресу: <http://www.econ.msu.ru/sys/raw.php?o=34427&p=attachment>. Также вы можете получить его, написав авторам (kartaev@gmail.com). Если в процессе решения заданий вы обнаружите в сборнике неточности и опечатки, то авторы будут вам признательны за сообщение о них.

ЗАДАЧА 1

ARIMA

Для каждого из представленных в файле **ARIMA_example** временных рядов подберите модель ARIMA, наилучшим образом описывающую его динамику.

Решение

Для подбора модели авторегрессии — скользящего среднего (ARIMA-модели) для каждого из трех временных рядов воспользуемся методологией Бокса—Дженкинса¹.

Ряд Y1

Для начала рассмотрим график временного ряда (см. рис. 1.1)², на котором видно, что ряд «колеблется» около значения 20, при этом явный тренд отсутствует, ряд «похож» на стационарный, но необходимо провести формальные тесты.

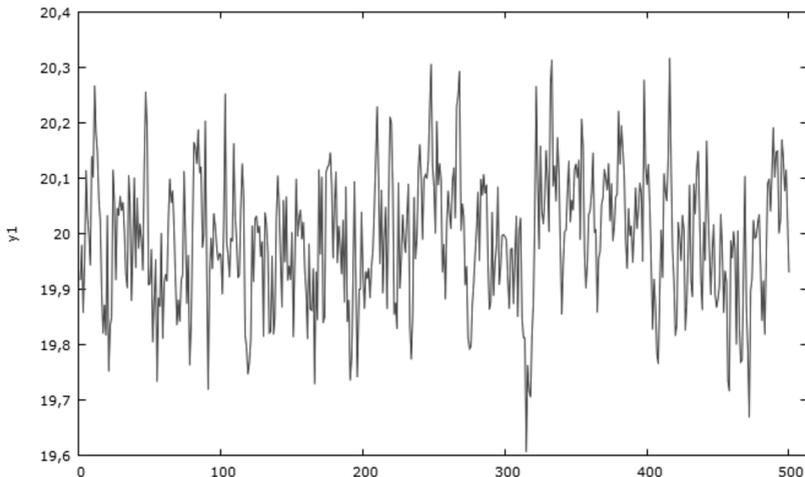


Рис. 1.1. График временного ряда Y1, состоящего из 500 наблюдений в последовательные моменты времени

¹ Магнус Я. Р., Катыйшев П. К., Пересецкий А. А. Эконометрика. Начальный курс: Учебник. — 6-е изд., перераб. и доп. — М.: Дело, 2004, глава 11.4

² Здесь и далее рисунки вставлены из эконометрического пакета Gretl.

Проверим ряд на стационарность, используя расширенный тест Дики—Фуллера с константой без тренда.

В лекциях Вашингтонского университета¹ процедура выбора числа лагов для ADF-теста описывается следующим образом. На первом шаге выбирается максимально возможный лаг P_{\max} , по **правилу Шварта**² для определения максимального числа лагов в ADF-тесте в зависимости от длины ряда T :

$$p_{\max} = \left\lceil A^* \left(\frac{T}{100} \right)^{1/4} \right\rceil, \quad (1.1)$$

где $[x]$ — целая часть x , $c = 12$ или 4 в зависимости от длины ряда.

На втором шаге на основе критерия Шварца выбирается оптимальное число лагов от одного до P_{\max} , определенного на первом шаге.

$$p^* = \arg \min(l_k) \quad (1.2)$$

$$l_k = \ln(\hat{\sigma}_k^2) + \frac{k * C_T}{T},$$

где $\hat{\sigma}_k^2 = \frac{\sum_{t=k+1}^T e_t^2}{T-k}$, а $C_T = \ln(T)$ в случае критерия Шварца и $C_T = 2$ для критерия Акаике.

По автоматическим настройкам в программе Gretl для данного ряда Y1 длиной в 500 наблюдений ставится максимальное количество лагов 17, исходя из «правила-12» Шварта, и выбирается оптимальное количество лагов с помощью критерия Шварца³.

Таблица 1.1

<p>Расширенный тест Дики-Фуллера для y1 включая 0 лага(-ов) для (1-L)y1 (max was 17, criterion Крит. Шварца) объем выборки 499 нулевая гипотеза единичного корня: a = 1</p> <p>тест с константой модель: (1-L)y = b0 + (a-1)*y(-1) + e оценка для (a - 1): -0,455737 тестовая статистика: tau_c(1) = -12,1132 Р-значение 1,333e-023 коэф. автокорреляции 1-го порядка для e: -0,023</p>

¹ Электронный ресурс, URL: <http://faculty.washington.edu/ezivot/econ584/notes/unitrootLecture2.pdf> (дата обращения 15.09.2016).

² Schwert. Test for Unit Roots: A Monte Carlo Investigation, JBES, 1989.

³ Здесь и далее для результатов тестов и оценки моделей вставки из программы Gretl.

По результатам теста (табл. 1.1) нулевая гипотеза о наличии единичного корня отвергается, поскольку полученное p -значение меньше любого разумного уровня значимости. Поэтому делается вывод, что ряд Y_1 стационарен, следовательно, для него определена автокорреляционная функция.

Рассмотрим график автокорреляционной и частной автокорреляционной функции (см. рис. 1.2).

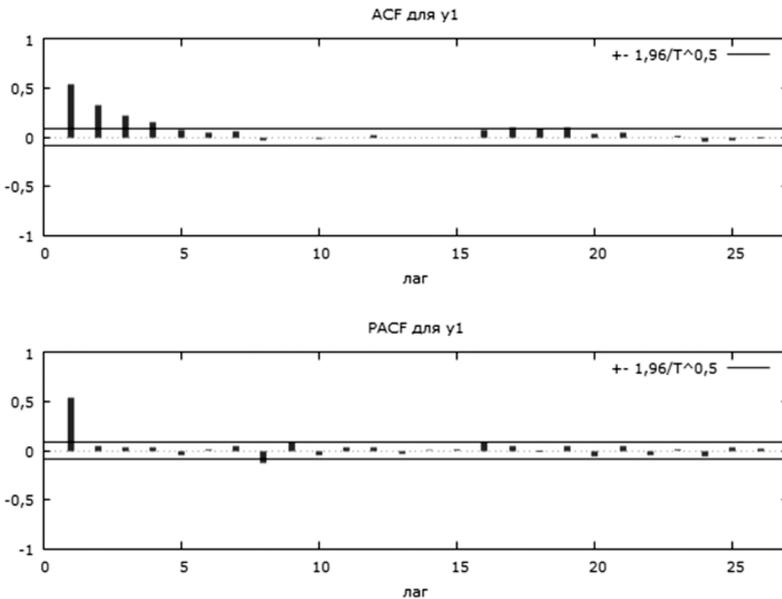


Рис. 1.2. График автокорреляционной и частной автокорреляционной функций ряда Y_1 . Столбцами обозначены коэффициенты автокорреляционной и частной автокорреляционной функций, горизонтальными линиями – границы 95%-ного доверительного интервала

Как видно из рис. 1.2, коэффициенты авторегрессионной функции постепенно сходятся к нулю, а для частной авторегрессионной функции значимым является только первый коэффициент, т.е. коррелограмма характерна для AR(1)-процесса.

Оценим AR(1)-модель $Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varepsilon_t$. Результаты оценки представлены в табл. 1.2.

Как видно из результатов оценки модели для временного ряда Y_1 , константа и коэффициент при первом авторегрессионном лаге значимы на 1%-ном уровне. Чтобы убедиться в качестве построенной модели, проверим, являются ли остатки модели белым шумом. Для этого рассмотрим коррелограмму остатков модели № 1 (см. рис. 1.3).

Таблица 1.2

Модель 1: ARMA, использованы наблюдения 1-500					
Зависимая переменная: y1					
Стандартные ошибки рассчитаны на основе Гессiana					
	Коэффициент	Ст. ошибка	Z	P-значение	
const	19,991	0,00973907	2052,6567	<0,0001	***
phi_1	0,543611	0,0374714	14,5074	<0,0001	***
Среднее зав. перемен	19,99129	Ст. откл. зав. перемен		0,118879	
Среднее инноваций	0,000081	Ст. откл. инноваций		0,099624	
Лог. правдоподобие	443,5315	Крит. Акаике		-881,0630	
Крит. Шварца	-868,4192	Крит. Хеннана-Куинна		-876,1016	
	Действительная часть	Мнимая часть	Модуль	Частота	
AR					
	Корень 1	1,8395	0,0000	1,8395	0,0000

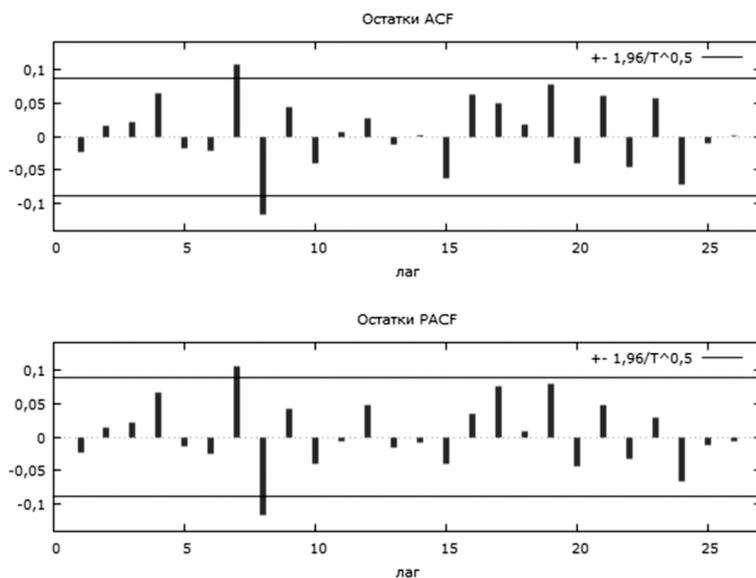


Рис. 1.3. Коррелограмма остатков модели № 1 для ряда Y1

Как видно из рис. 1.3, среди первых шести коэффициентов автокорреляционной функции нет значимых (построение модели с большим количеством лагов не представляется целесообразным), поэтому можно сделать вывод, что остатки модели № 1 являются белым шумом, т.е. качество модели можно считать удовлетворительным.

Ответ: ряд Y1 представляет собой AR(1)-процесс: $\hat{Y}_t = 19,99 + 0,54 Y_{t-1}$
(0,01) (0,037)

Ряд Y2

Аналогично с предыдущим случаем рассмотрим график временного ряда Y2. На рис. 1.4 видно, что тренд отсутствует, ряд колеблется около значения 5.

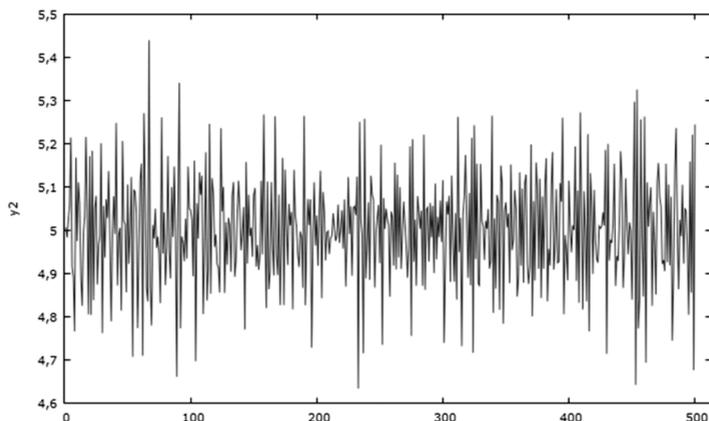


Рис. 1.4. График временного ряда Y2

Проведем расширенный тест Дики—Фулера на наличие единичного корня.

Таблица 1.3

Расширенный тест Дики-Фулера для y2
включая 6 лага(-ов) для $(1-L)y_2$
(max was 17, criterion Крит. Шварца)
объем выборки 493
нулевая гипотеза единичного корня: $a = 1$

тест с константой
модель: $(1-L)y = b_0 + (a-1)y(-1) + \dots + e$
оценка для $(a - 1)$: -4,27818
тестовая статистика: $\tau_{c(1)} = -14,6794$
асимпт. p-значение 3,354e-034
коэф. автокорреляции 1-го порядка для e: -0,006
лаг для разностей: $F(6, 485) = 24,012 [0,0000]$

По результатам теста нулевая гипотеза о наличии единичного корня отвергается, поскольку p -значение меньше любого уровня значимости. Это позволяет сделать вывод о стационарности ряда.

По коррелограмме временного ряда Y2 (см. рис. 1.5) видно, что первый коэффициент автокорреляционной функции значим, а коэффици-

енты частной автокорреляционной функции стремятся к нулю по мере роста лага.

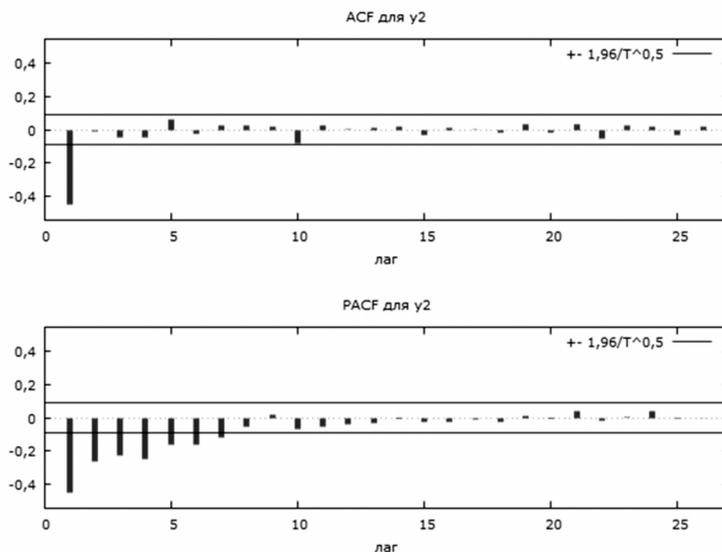


Рис. 1.5. График автокорреляционной и частной автокорреляционной функций временного ряда Y_2

Эта ситуация характерна для процесса скользящего среднего первого порядка (MA(1)-процесса). Построим MA(1)-модель: $Y_t = \theta_0 + \theta_1 \varepsilon_{t-1} + \varepsilon_t$. Оценки представлены в табл. 1.4.

Таблица 1.4

Модель 2: ARMA, использованы наблюдения 1-500				
Зависимая переменная: y_2				
Стандартные ошибки рассчитаны на основе Гессииана				
	Коэффициент	Ст. ошибка	Z	P-значение
const	5,00087	0,000782403	6391,6767	<0,0001 ***
theta_1	-0,829799	0,0240793	-34,4611	<0,0001 ***
Среднее зав. перемен	5,001028	Ст. откл. зав. перемен	0,128549	
Среднее инноваций	0,000258	Ст. откл. инноваций	0,101780	
Лог. правдоподобие	432,4189	Крит. Акаике	-858,8379	
Крит. Шварца	-846,1941	Крит. Хеннана-Куинна	-853,8765	
	<i>Действительная часть</i>	<i>Мнимая часть</i>	<i>Модуль</i>	<i>Частота</i>
МА				
Корень 1	1,2051	0,0000	1,2051	0,0000

Константа и коэффициент пери первом МА-лаге значимы. Проверим, являются ли остатки модели № 2 белым шумом.

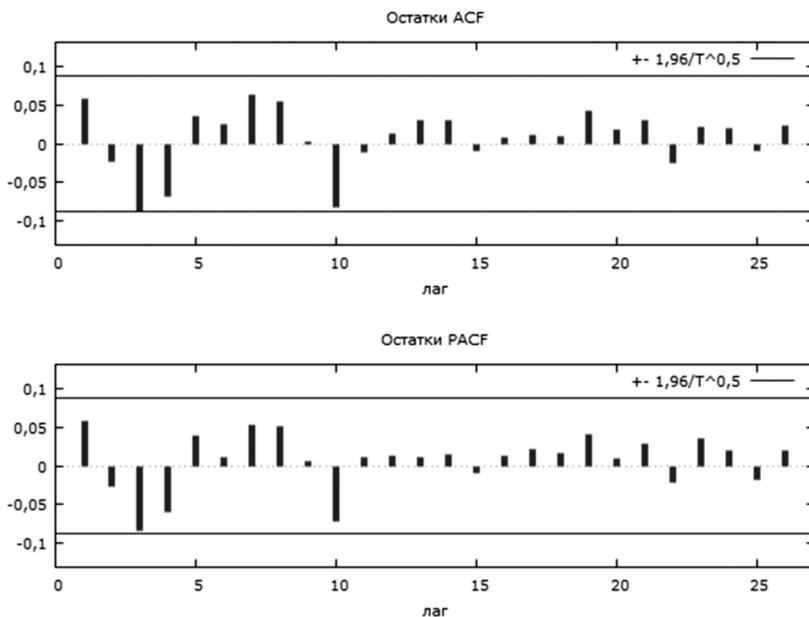


Рис. 1.6. Коррелограмма остатков модели № 2

Из рис. 1.6 видно, что коэффициенты автокорреляционной функции остатков модели № 2 не являются значимыми (на 5%-ном уровне), что позволяет сделать вывод об удовлетворительном качестве модели № 2.

Ответ: временной ряд Y2 представляет собой MA(1)-процесс

$$\hat{Y}_t = 5,00 - 0,83\varepsilon_{t-1}$$

(0,0008) (0,024)

Ряд Y3

График временного ряда Y3 (см. рис. 1.7) характерен для нестационарного процесса: для разных периодов отличаются средние значения и амплитуда колебаний при отсутствии явного тренда.

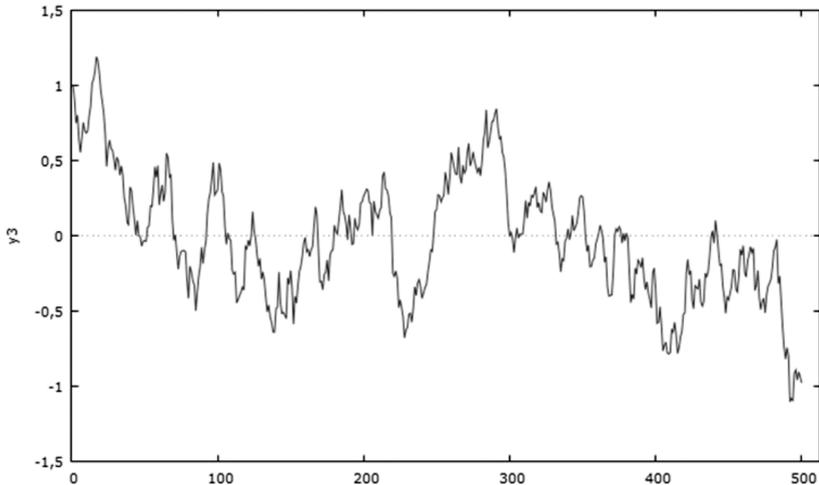


Рис. 1.7. График временного ряда y_3

Проведем расширенный тест Дики—Фулера с константой на наличие единичного корня.

Таблица 1.5

Расширенный тест Дики-Фулера для y_3
 включая 0 лага(-ов) для $(1-L)y_3$
 (max was 17, criterion Крит. Шварца)
 объем выборки 499
 нулевая гипотеза единичного корня: $a = 1$

тест с константой
 модель: $(1-L)y = b_0 + (a-1)*y(-1) + e$
 оценка для $(a - 1)$: -0,0331852
 тестовая статистика: $\tau_c(1) = -2,93261$
 Р-значение 0,04238
 коэф. автокорреляции 1-го порядка для e : 0,053

По результатам теста нулевая гипотеза о наличии единичного корня не отвергается на 1%-ном уровне значимости, поскольку p -значение составило 0,042. Ряд нестационарен, необходимо перейти к его первой разности $\Delta Y_t = Y_t - Y_{t-1}$, график которой представлен на рис. 1.8.

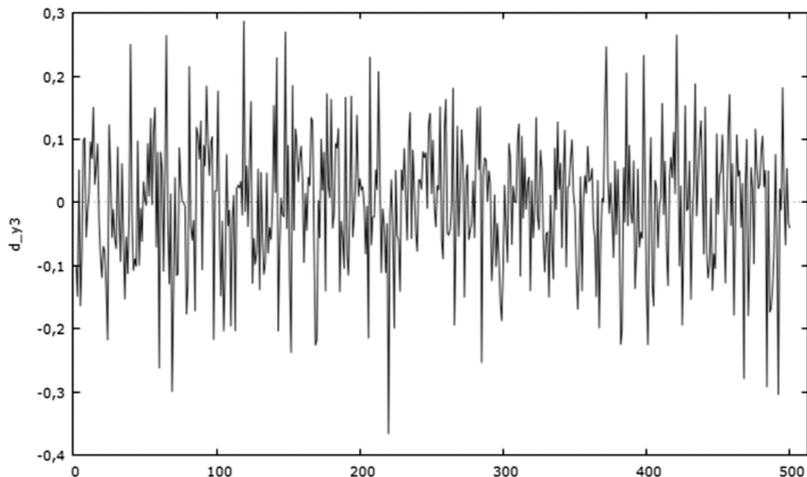


Рис. 1.8. График временного ряда ΔY_3

Проверим первую разность ряда Y_3 на стационарность, так же проведя расширенный тест Дики—Фулера (см. табл. 1.6).

Таблица 1.6

Расширенный тест Дики-Фулера для d_y3
включая 0 лага(-ов) для $(1-L)d_y3$
(max was 17, criterion Крит. Шварца)
объем выборки 498
нулевая гипотеза единичного корня: $a = 1$

тест с константой
модель: $(1-L)y = b0 + (a-1)*y(-1) + e$
оценка для $(a - 1)$: -0,962918
тестовая статистика: $\tau_c(1) = -21,4761$
Р-значение 1,669e-037
коэф. автокорреляции 1-го порядка для e : -0,002

По результатам теста ряд стационарный. Его коррелограмма представлена на рис. 1.9.

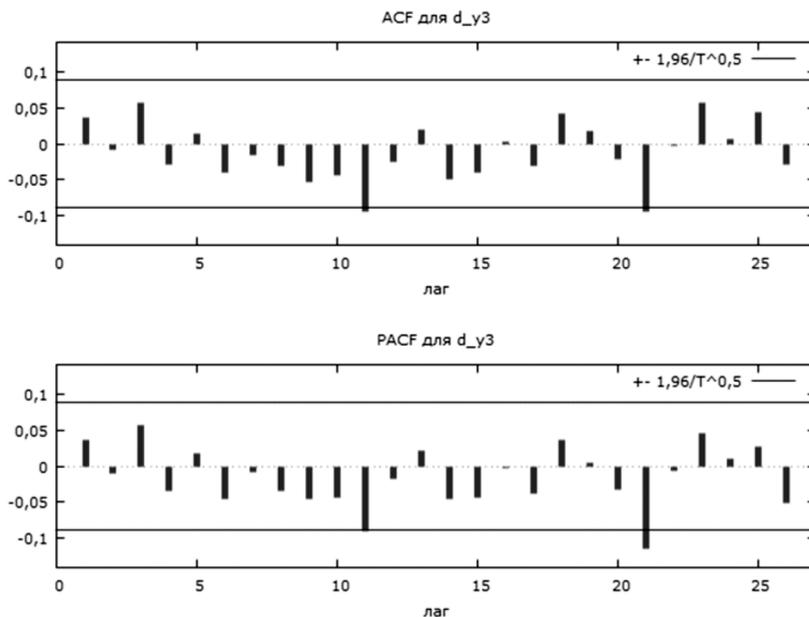


Рис. 1.9. График автокорреляционной и частной автокорреляционной функций временного ряда ΔY_3

Все коэффициенты автокорреляционной и частной автокорреляционной функций (до 11-го лага) являются незначимыми на 5%-ном уровне. Это ситуация говорит о том, что первая разность временного ряда Y_3 представляет собой белый шум: $\Delta Y_t = \varepsilon_t$. То есть сам ряд Y_3 является случайным блужданием.

Ответ: ряд Y_3 является случайным блужданием $Y_t = Y_{t-1} + \varepsilon_t$.

ЗАДАЧА 2

ДИНАМИКА РЕАЛЬНЫХ ДОХОДОВ

На основе представленных в файле **rus_income** данных о динамике индекса реальных денежных доходов населения России:

- (а) подберите ARIMA-модель, удачно описывающую динамику индекса реального дохода;
- (б) на основе полученной модели постройте прогноз индекса реального дохода на полгода (2 квартала) вперед (по сравнению с имеющимися в файле данными);
- (в) сравните прогнозные значения со значениями, наблюдавшимися на практике в рассматриваемые месяцы. (Примечание: фактическое значение индекса реальных денежных доходов населения в первом квартале 2008 г., согласно данным Росстата, составило 165,6.)

Решение

(а) В качестве первого шага эконометрического моделирования полезно посмотреть на график анализируемого ряда (см. рис. 2.1).

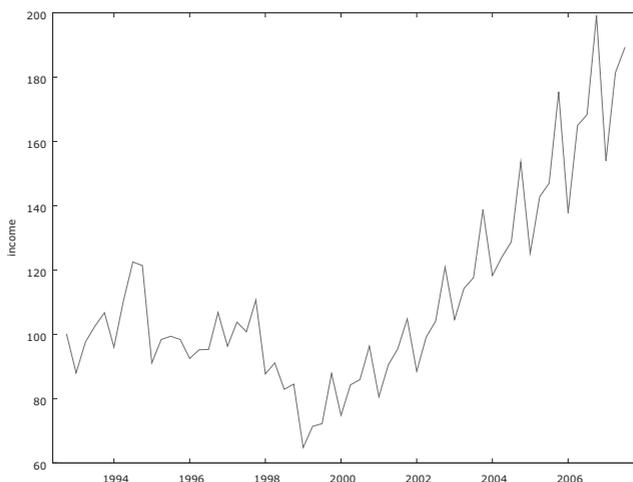


Рис. 2.1. Динамика зависимой переменной (полная выборка)

Легко видеть, что в 1998 г. произошел структурный сдвиг, вызванный дефолтом и кризисом, случившимися в середине этого года. Для целей прогнозирования целесообразно использовать однородные данные, по-

этому ограничим выборку. Будем анализировать данные, начиная с первого квартала 1999 г. (см. рис. 2.2).

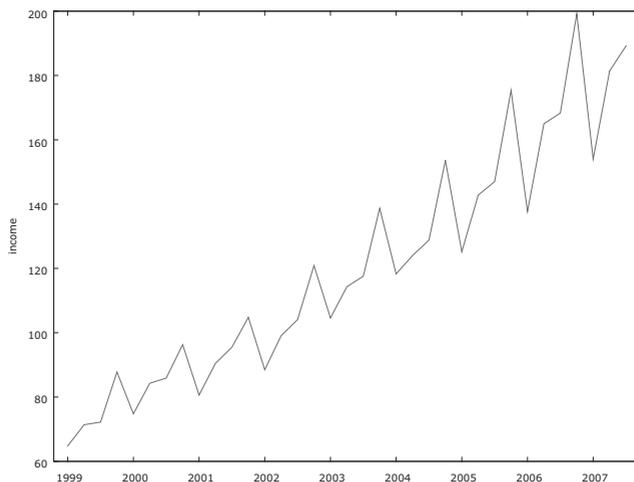


Рис. 2.2. Динамика зависимой переменной (начиная с 1999 г.)

Осуществив расширенный тест Дики—Фуллера, приходим к выводу, что ряд является интегрированным первого порядка. Результаты теста представлены в табл. 2.1 и 2.2. В обоих случаях используется автоматический подбор порядка лага на основе минимизации критерия Шварца.

Таблица 2.1

Результаты теста Дики—Фуллера для уровней ряда Income

Расширенный тест Дики-Фуллера для income
включая 6 лага(-ов) для $(1-L)income$
(max was 10, criterion Крит. Акаике)
объем выборки 35
нулевая гипотеза единичного корня: $a = 1$
тест с константой
модель: $(1-L)y = b_0 + (a-1)*y(-1) + \dots + e$
оценка для $(a - 1)$: 0,0488965
тестовая статистика: $\tau_c(1) = 1,50685$
асимпт. p-значение 0,9994
коэф. автокорреляции 1-го порядка для e : -0,128
лаг для разностей: $F(6, 27) = 109,488 [0,0000]$
с константой и трендом
модель: $(1-L)y = b_0 + b_1*t + (a-1)*y(-1) + \dots + e$
оценка для $(a - 1)$: -0,365652
тестовая статистика: $\tau_{ct}(1) = -2,8844$
асимпт. p-значение 0,1675
коэф. автокорреляции 1-го порядка для e : -0,020
лаг для разностей: $F(8, 24) = 38,868 [0,0000]$

Таблица 2.2

Результаты теста Дики—Фуллера для первых разностей ряда

Расширенный тест Дики-Фуллера для d_income
включая 3 лага(-ов) для (1-L)d_income
(max was 10, cregion Крит. Акаике)
объем выборки 35
нулевая гипотеза единичного корня: $a = 1$

тест с константой
модель: $(1-L)y = b_0 + (a-1)y(-1) + \dots + e$
оценка для (a - 1): -0,856137
тестовая статистика: $\tau_c(1) = -3,53183$
асимпт. p-значение 0,007234
коэф. автокорреляции 1-го порядка для e: 0,032
лаг для разностей: $F(3, 30) = 126,359 [0,0000]$

с константой и трендом
модель: $(1-L)y = b_0 + b_1*t + (a-1)y(-1) + \dots + e$
оценка для (a - 1): -1,61068
тестовая статистика: $\tau_{ct}(1) = -4,11919$
асимпт. p-значение 0,005836
коэф. автокорреляции 1-го порядка для e: -0,144
лаг для разностей: $F(5, 27) = 85,244 [0,0000]$

На рис. 2.3 представлена коррелограмма для первых разностей ряда зависимой переменной. Эта коррелограмма может быть косвенным сви-

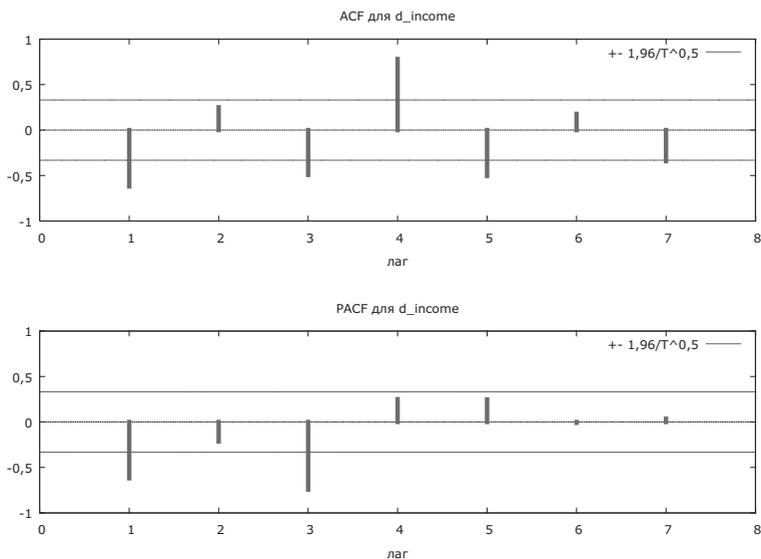


Рис. 2.3. Коррелограмма первых разностей моделируемой переменной

детельством в пользу наличия в модели авторегрессии четвертого порядка (соответствующий лаг значим). На это же указывает и характер поведения ряда на рис. 2.2, где ярко выражена его квартальная сезонность, что также соответствует корреляции между текущим значением ряда и значением ряда 1 год (т.е. 4 квартала) назад.

Поэтому оценим параметры модели $ARIMA(4,1,0)$ для ряда Income. Отметим, что, если оценить модели с чуть большим или чуть меньшим порядком авторегрессии, то значение информационного критерия Шварца в них окажется выше, что также является аргументом в пользу выбора модели именно с четырьмя лагами. Результаты оценивания модели содержатся в табл. 2.3.

Таблица 2.3

Модель ARIMA для зависимой переменной

Модель 1: ARIMA, использованы наблюдения 1999:1-2007:3 (T = 35)					
Зависимая переменная: (1-L) income					
	Коэффициент	Ст. ошибка	z	P-значение	
Const	3,46835	0,913216	3,7980	0,0001	***
phi_1	-0,249829	0,0687012	-3,6365	0,0003	***
phi_2	-0,253904	0,0723654	-3,5086	0,0005	***
phi_3	-0,248891	0,0736955	-3,3773	0,0007	***
phi_4	0,896487	0,0725088	12,3638	<0,0001	***
Среднее зав. перемен	2,991429	Ст. откл. зав. перемен	18,55338		
Среднее инноваций	0,000000	Ст. откл. инноваций	4,209520		
Лог. правдоподобие	-97,27242	Крит. Акаике	204,5448		
Крит. Шварца	212,3216	Крит. Хеннана-Куинна	207,2294		

(б) Прогноз, который можно сделать на основе этой модели, представлен в табл. 2.4 и на рис. 2.4.

Таблица 2.4

Прогноз реальных денежных доходов населения

Для 95% доверительных интервалов, $z(0,025) = 1,96$			
Набл.	Предсказание	Ст. ошибка	95% доверительный интервал
2007:4	222,714	4,20952	(214,464, 230,965)
2008:1	168,488	5,26233	(158,174, 178,802)

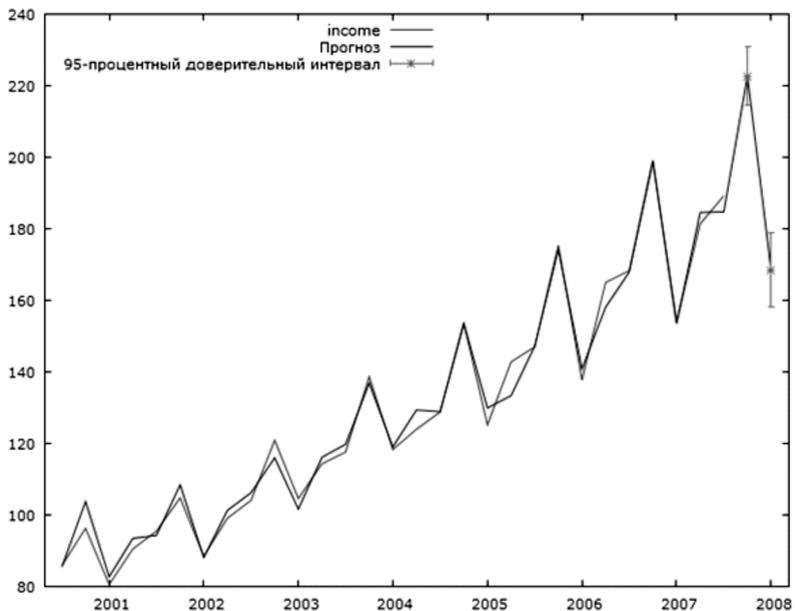


Рис. 2.4. Прогноз реальных денежных доходов населения

(в) Сравнивая прогноз на первый квартал 2008 г. с фактическим значением индекса реальных доходов населения, которое наблюдалось в этот период, можно отметить, что фактическое значение оказалось внутри 95%-ного доверительного интервала для прогноза (и довольно близко к его середине).

ЗАДАЧА 3

РЕКЛАМА И ПРОДАЖИ

В вашем распоряжении имеются ежемесячные данные (файл **advert.xlsx**) за период с января 2001 по декабрь 2010 г. о расходах на рекламу и объемах продаж регионального подразделения транснациональной корпорации: x_t — расходы фирмы на рекламу, y_t — объем продаж фирмы.

- (а) Осуществите для рассматриваемых переменных тест Грейнджера на причинно-следственную связь (используйте первый, второй и третий лаги для каждой из переменных). Интерпретируйте результаты теста.
- (б) Оцените регрессию y_t на константу, x_t , x_{t-1} , x_{t-2} . Дайте содержательную интерпретацию полученных результатов. Осуществите тест Дарбина—Уотсона и интерпретируйте его результаты. Можно ли использовать в представленной модели обычные стандартные ошибки или необходимо использовать состоятельные в условиях автокорреляции стандартные ошибки?
- (в) Если в исходной модели были незначимые переменные, то оцените модель заново, исключив их. Осуществите тест Дарбина—Уотсона и интерпретируйте его результаты.
- (г) Оцените регрессию y_t на константу, x_t , x_{t-1} , y_{t-1} . Дайте содержательную интерпретацию полученных результатов. Почему в этой модели нельзя использовать тест Дарбина—Уотсона? Каким образом можно проверить, есть ли автокорреляция в новой модели?
- (д) Сравните полученные в предыдущих пунктах модели и выберите наиболее удачную из них.

Решение

(а) Результаты проведения теста Грейнджера, полученные в программном пакете Gretl, представлены в табл. 3.1.

Таблица 3.1

Метод оценки - МНК, наблюдения 2001:04-2010:12 (T = 117)					
Уравнение 1: зависимая переменная - x					
	Коэффициент	Ст. ошибка	t-статистика	P-значение	
const	7,91218	4,35237	1,8179	0,0718	*
x_{t-1}	-0,11157	0,0944463	-1,1813	0,2400	
x_{t-2}	0,131526	0,220934	0,5953	0,5529	
x_{t-3}	0,161901	0,220049	0,7358	0,4634	

Окончание табл. 3.1

y_{t-1}	-0,0326293	0,243404	-0,1341	0,8936	
y_{t-2}	-0,0987355	0,271033	-0,3643	0,7163	
y_{t-3}	-0,0261092	0,152815	-0,1709	0,8647	
Среднее зав. перемен	4,444444		Ст. откл. зав. перемен	2,966609	
Сумма кв. остатков	984,3176		Ст. ошибка модели	2,991378	
R-квадрат	0,035823		Испр. R-квадрат	-0,016768	
F(6, 110)	0,681156		P-значение (F)	0,665154	
Параметр rho	-0,005988		Стат. Дарбина-Вотсона	2,003786	
F-тесты для нулевых ограничений:					
Все лаги для x F(3, 110) = 0,74644 [0,5267]					
Все лаги для y F(3, 110) = 0,35307 [0,7870]					
Уравнение 2: зависимая переменная - y					
	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>t-статистика</i>	<i>P-значение</i>	
<i>const</i>	11,6475	1,6867	6,9055	<0,0001	***
x_{t-1}	0,816099	0,0366014	22,2969	<0,0001	***
x_{t-2}	0,0226147	0,0856201	0,2641	0,7922	
x_{t-3}	0,105255	0,085277	1,2343	0,2197	
y_{t-1}	0,488741	0,0943279	5,1813	<0,0001	***
y_{t-2}	-0,103673	0,105035	-0,9870	0,3258	
y_{t-3}	0,0330453	0,0592213	0,5580	0,5780	
Среднее зав. перемен	27,10342		Ст. откл. зав. перемен	3,017790	
Сумма кв. остатков	147,8292		Ст. ошибка модели	1,159268	
R-квадрат	0,860066		Испр. R-квадрат	0,852433	
F(6, 110)	112,6805		P-значение (F)	1,26e-44	
Параметр rho	-0,017481		Стат. Дарбина-Вотсона	2,027780	
F-тесты для нулевых ограничений:					
Все лаги для x F(3, 110) = 166,97 [0,0000]					
Все лаги для y F(3, 110) = 12,122 [0,0000]					

Интерпретация: исходя из первого уравнения, гипотеза «у не влияет на x» принимается, исходя из второго уравнения, гипотеза «x не влияет на у» отклоняется. Следовательно, x является причиной по Грейнджеру для у.

(6) Результаты оценивания регрессии y_t на константу, x_t , x_{t-1} , x_{t-2} (см. табл. 3.2).

Таблица 3.2

МНК, использованы наблюдения 2001:03-2010:12 (T = 118)					
Зависимая переменная - y					
Стандартные ошибки НАС, ширина окна 3 (Ядро Бартлетта (Bartlett))					
	Коэффициент	Ст. ошибка	t-статистика	P-значение	
const	21,6889	0,356816	60,7846	<0,0001	***
x_t	-0,0109618	0,0440358	-0,2489	0,8039	
x_{t-1}	0,830543	0,043612	19,0439	<0,0001	***
x_{t-2}	0,414882	0,0438739	9,4562	<0,0001	***
Среднее зав. перемен	27,07966		Ст. откл. зав. перемен	3,015928	
Сумма кв. остатков	247,8437		Ст. ошибка модели	1,474472	
R-квадрат	0,767110		Испр. R-квадрат	0,760982	
F(3, 114)	159,8971		P-значение (F)	1,09e-40	
Лог. правдоподобие	-211,2195		Крит. Акаике	430,4389	
Крит. Шварца	441,5217		Крит. Хеннана-Куинна	434,9388	
Параметр rho	0,436848		Стат. Дарбина-Вотсона	1,116712	

Содержательная интерпретация: рост расходов на рекламу на одну единицу ведет к:

- мгновенному сокращению продаж фирмы на 0,011 единицы;
- увеличению продаж фирмы на 0,83 единицы по итогам следующего месяца;
- увеличению продаж фирмы на 0,415 единицы спустя два месяца.

Суммарное увеличение объема продаж в долгосрочном периоде (долгосрочный динамический мультипликатор расходов на рекламу) составит $-0,011 + 0,83 + 0,415 = 1,234$ единицы.

Осуществим тест Дарбина—Уотсона для данной модели. Программный пакет Gretl рассчитывает следующее p -значение для данного теста: $2,43 \cdot 10^{-7}$. Следовательно, нулевая гипотеза отклоняется, в модели есть проблема автокорреляции ошибок первого порядка, т.е. в регрессии $u_t = \rho u_{t-1} + \varepsilon_t$, где u_t — ошибки исходной модели, отклоняется гипотеза $H_0: \rho = 0$.

Таким образом, для получения состоятельных оценок стандартных ошибок необходимо использовать состоятельные в условиях автокорреляции стандартные ошибки.

(в) В исходной модели была незначимой переменная x_t . Оценим регрессию без нее (см. табл. 3.3).

Таблица 3.3

МНК, использованы наблюдения 2001:03-2010:12 (T = 118)					
Зависимая переменная - y					
Стандартные ошибки НАС, ширина окна 3 (Ядро Бартлетта (Bartlett))					
	Коэффициент	Ст. ошибка	t-статистика	P-значение	
const	21,6406	0,271646	79,6648	<0,0001	***
x_{t-1}	0,831631	0,0420047	19,7985	<0,0001	***
x_{t-2}	0,413805	0,0427633	9,6766	<0,0001	***
Среднее зав. перемен	27,07966		Ст. откл. зав. перемен	3,015928	
Сумма кв. остатков	247,9662		Ст. ошибка модели	1,468410	
R-квадрат	0,766995		Испр. R-квадрат	0,762943	
F(2, 115)	232,4056		P-значение (F)	3,99e-41	
Лог. правдоподобие	-211,2486		Крит. Акаике	428,4972	
Крит. Шварца	436,8093		Крит. Хеннана-Куинна	431,8722	
Параметр rho	0,439229		Стат. Дарбина-Вотсона	1,112549	

Проведем тест Дарбина—Уотсона для этой модели. Программный пакет Gretl рассчитывает следующее p -значение для данного теста: $1,9 \cdot 10^{-7}$. Аналогично предыдущим рассуждениям в данном случае также надо использовать состоятельные в условиях автокорреляции стандартные ошибки.

(г) Оценим регрессию y_t на константу, x_t , x_{t-1} , y_{t-1} (табл. 3.4).

Таблица 3.4

МНК, использованы наблюдения 2001:02-2010:12 (T = 119)					
Зависимая переменная - y					
Стандартные ошибки НАС, ширина окна 3 (Ядро Бартлетта (Bartlett))					
	Коэффициент	Ст. ошибка	t-статистика	P-значение	
const	9,92939	1,13322	8,7621	<0,0001	***
x	-0,00771794	0,0336524	-0,2293	0,8190	
x_{t-1}	0,817293	0,0301132	27,1407	<0,0001	***
y_{t-1}	0,502631	0,0431649	11,6445	<0,0001	***
Среднее зав. перемен	27,06807		Ст. откл. зав. перемен	3,005783	
Сумма кв. остатков	152,5433		Ст. ошибка модели	1,151722	
R-квадрат	0,856914		Испр. R-квадрат	0,853182	
F(3, 115)	372,4367		P-значение (F)	4,87e-59	
Лог. правдоподобие	-183,6290		Крит. Акаике	375,2581	
Крит. Шварца	386,3746		Крит. Хеннана-Куинна	379,7721	
Параметр rho	-0,003065		h-статистика Дарбина	-0,037905	

Содержательная интерпретация: рост расходов на рекламу на одну единицу ведет к:

- мгновенному сокращению продаж фирмы на 0,0077 единицы;
- увеличению продаж фирмы на $0,817 + (-0,0077) \cdot 0,5 = 0,813$ единицы по итогам следующего месяца;

Долгосрочный динамический мультипликатор составит

$$\frac{-0,0077 + 0,817}{1 - 0,5} = 1,62 \text{ единицы.}$$

Для проверки проблемы автокорреляции в данной модели нельзя использовать тест Дарбина—Уотсона, так как его нельзя применять, если в правой части уравнения есть лагированное значение зависимой переменной (y_{t-1}).

Наличие автокорреляции в данной модели можно проверить с помощью теста Бреуша—Годфри. Встроенная в программный пакет Gretl процедура применения теста Бреуша—Годфри дает следующее p -значение для данного теста: 0,971. Тем самым, в данном случае для регрессии $u_t = \text{const} + \delta_0 x_t + \delta_1 x_{t-1} + \beta_1 y_{t-1} + \rho u_{t-1} + \varepsilon_t$, где u_t — ошибки исходной модели, принимается гипотеза $H_0: \rho = 0$, автокорреляции в модели нет.

(д) Для сравнения построенных моделей будем пользоваться информационными критериями Акаике и Шварца. Они принимают наименьшие значения для последней построенной модели, следовательно, она и является наиболее удачной с точки зрения этих критериев.

ЗАДАЧА 4

КОИНТЕГРАЦИЯ

Моделируется зависимость курса ценной бумаги (переменная P) от фондового индекса (переменная $INDEX$). Исходный файл с данными **cointegration.wfl**.

- (а) Коинтегрированы ли рассматриваемые переменные? Осуществите соответствующий тест.
- (б) Оцените параметры модели коррекции ошибок:

$$\Delta P_t = \gamma_0 + \gamma_1 \cdot \Delta INDEX_t + \gamma_2 \cdot e_{t-1} + \varepsilon_t.$$

Быстро ли в рассматриваемом случае происходит возврат модели к долгосрочному равновесию?

Решение

(а) Для ответа на вопрос осуществим тест Энгла—Грейнджера. Реализация теста состоит из четырех шагов, результаты каждого из которых последовательно представлены в табл. 4.1. На первом и втором шагах тестируется нестационарность каждой из двух переменных. Из определения коинтеграции следует, что коинтегрированы могут быть только нестационарные временные ряды.

Как видно из таблицы, соответствующие P -значения в обоих случаях больше чем 0,01 и даже 0,05. Таким образом, при любом разумном уровне значимости гипотеза о нестационарности не отвергается. Это означает, что имеет смысл третий шаг, на котором оценивается коинтеграционное соотношение (также представленное в таблице).

На последнем, четвертом шаге тестируется нестационарность ряда остатков, полученных в ходе оценивания коинтеграционного соотношения. Соответствующее P -значение меньше чем 0,01. Следовательно, мы делаем вывод, что ряд остатков стационарен. Таким образом, можно сделать вывод о том, что временные ряды коинтегрированы.

(б) Результаты оценки модели коррекции ошибок представлены в табл. 4.2, где переменная e_{-1} представляет собой остатки, полученные в ходе оценивания коинтеграционного соотношения и взятые с лагом в один период.

Следует отметить, что коэффициент при этой переменной является отрицательным и статистически значимым (расчетное значение тестовой статистики по модулю равно 29,6, что больше соответствующего критического значения из таблиц распределения Дэвидсона и Маккин-

нона¹, что еще раз подтверждает наличие коинтеграции между динамикой курса ценной бумаги и динамикой фондового индекса.

Этот коэффициент по абсолютной величине близок к единице, что свидетельствует о том, что возврат модели к долгосрочному равновесию происходит быстро.

Таблица 4.1

Результаты теста Энга—Грейджера

Шаг 1: тестирование единичного корня для P				
Расширенный тест Дики-Фуллера для P				
включая один лаг для (1-L)P				
(max was 1, criterion Крит. Акаике)				
объем выборки 998				
нулевая гипотеза единичного корня: $a = 1$				
тест с константой				
модель: $(1-L)y = b_0 + (a-1)y(-1) + \dots + e$				
оценка для $(a - 1)$: -0,015931				
тестовая статистика: $\tau_c(1) = -2,52884$				
асимпт. p-значение 0,1086				
коэф. автокорреляции 1-го порядка для e : -0,106				
Шаг 2: тестирование единичного корня для INDEX				
Расширенный тест Дики-Фуллера для INDEX				
включая один лаг для (1-L)INDEX				
(max was 1, criterion Крит. Акаике)				
объем выборки 998				
нулевая гипотеза единичного корня: $a = 1$				
тест с константой				
модель: $(1-L)y = b_0 + (a-1)y(-1) + \dots + e$				
оценка для $(a - 1)$: -0,00105014				
тестовая статистика: $\tau_c(1) = -0,627105$				
асимпт. p-значение 0,8623				
коэф. автокорреляции 1-го порядка для e : -0,002				
Шаг 3: коинтеграционная регрессия				
Коинтеграционная регрессия -				
МНК, использованы наблюдения 1-1000				
Зависимая переменная: P				
	Коэффициент	Ст. ошибка	t-статистика	P-значение

const	-0,145641	0,373916	-0,3895	0,6970
INDEX	0,627453	0,00396381	158,3	0,0000 ***

¹ Таблица критических значений доступна, например, в параграфе 11.3 учебника: *Магнус Я. Р., Катышев П. К., Пересецкий А. А. Эконометрика. Начальный курс: Учеб. — 6-е изд., перераб. и доп. — М.: Дело, 2004.*

Окончание табл. 4.1

Среднее зав. перемен	57,78679	Ст. откл. зав. перемен	12,37770
Сумма кв. остатков	5862,444	Ст. ошибка модели	2,423673
R-квадрат	0,961697	Испр. R-квадрат	0,961659
Лог. правдоподобие	-2303,222	Крит. Акаике	4610,444
Крит. Шварца	4620,259	Крит. Хеннана-Куинна	4614,174
Параметр rho	0,219333	Стат. Дарбина-Вотсона	1,322683

Шаг 4: тестирование единичного корня для uhat

Расширенный тест Дики-Фуллера для uhat
включая 0 лага(-ов) для (1-L)uhat
(max was 1, criterion Крит. Акаике)
объем выборки 999
нулевая гипотеза единичного корня: $a = 1$

модель: $(1-L)y = (a-1)y(-1) + e$
оценка для $(a - 1)$: -0,780667
тестовая статистика: $\tau_c(2) = -29,1996$
Р-значение 2,178e-045
коэф. автокорреляции 1-го порядка для e: -0,004

Таблица 4.2

Модель коррекции ошибок

Модель 2: МНК, использованы наблюдения 2-1000 (T = 999)				
Зависимая переменная: d_P				
	Коэффициент	Ст. ошибка	t-статистика	P-значение
const	-0,047131	0,0639195	-0,7373	0,4611
d_INDEX	0,291462	0,0624449	4,6675	<0,0001 ***
e_1	-0,780407	0,0263773	-29,5863	<0,0001 ***
Среднее зав. перемен	-0,055968	Ст. откл. зав. перемен	2,781111	
Сумма кв. остатков	4062,170	Ст. ошибка модели	2,019526	
R-квадрат	0,473751	Испр. R-квадрат	0,472695	
F(2, 996)	448,3208	Р-значение (F)	1,4e-139	
Лог. правдоподобие	-2118,177	Крит. Акаике	4242,354	
Крит. Шварца	4257,075	Крит. Хеннана-Куинна	4247,949	
Параметр rho	-0,015438	Стат. Дарбина-Вотсона	2,030691	

ЗАДАЧА 5

ЛИСЫ И КРОЛИКИ

Вы располагаете данными о динамике популяций лис и кроликов в некотором регионе (файл **fox_rabbit**). Предполагается, что численности этих популяций взаимосвязаны между собой и описываются моделью векторной авторегрессии VAR(p).

- (а) Определите оптимальный порядок лага для VAR-модели, опираясь на информационный критерий Шварца.
- (б) Оцените параметры модели. Запишите полученные уравнения регрессии в стандартной форме, указав коэффициенты детерминации и (в скобках под соответствующими коэффициентами) стандартные ошибки. Укажите, какие из переменных являются значимыми при уровне значимости 1%.
- (в) Постройте прогноз популяции кроликов на один период вперед, укажите 95%-ный доверительный интервал прогноза.
- (г) Допустим, в вашем распоряжении нет исходных данных, а есть только уравнения, оцененные в пункте (б). Вычислите равновесный уровень (оценку безусловного математического ожидания) численности лис в регионе. Приведите необходимые расчеты.
- (д) Постройте график отклика функции на импульс, характеризующий динамику численности лис в ответ на импульс численности кроликов. Дайте содержательную интерпретацию такого вида графика.
- (е) Для двух переменных, имеющихся в вашем распоряжении, осуществите тест Грейнджера на причинно-следственную связь и интерпретируйте его результаты.

Решение

(а) Нужно выбирать VAR(1), так как для первого лага критерий Шварца достигает наименьшего значения (табл. 5.1).

Таблица 5.1

Выбор лагов для VAR-модели

Порядок лагов	Логарифм правдоподобия	Критерий Акаике (AIC)	Критерий Шварца (BIC)	Критерий Хеннана–Куинна (HQC)
1	6268,83655	-12,677807	-12,648076*	-12,666500*
2	6273,27877	-12,678702*	-12,629151	-12,659858
3	6275,062	-12,674215	-12,604843	-12,647833

Окончание табл. 5.1

Порядок лагов	Логарифм правдоподобия	Критерий Акаике (AIC)	Критерий Шварца (BIC)	Критерий Хенна—Куинна (HQC)
4	6276,37646	-12,668778	-12,579586	-12,634859
5	6277,06235	-12,66207	-12,553056	-12,620613
6	6278,17609	-12,656227	-12,527393	-12,607232
7	6278,52001	-12,648826	-12,500172	-12,592294
8	6281,43205	-12,646624	-12,478149	-12,582554
9	6283,21265	-12,642131	-12,453835	-12,570524
10	6284,22161	-12,636076	-12,42796	-12,556931
11	6286,86979	-12,63334	-12,405403	-12,546657
12	6287,40858	-12,626333	-12,378576	-12,532113

(б) Оценим модель с помощью программного пакета Gretl (см. табл. 5.2).

Таблица 5.2

Метод оценки - МНК, наблюдения 1961-2959 (T = 999)					
Уравнение 1: зависимая переменная - <i>fox</i>					
Робастные оценки стандартных ошибок (с поправкой на гетероскедастичность), вариант HC1					
	Коэффициент	Ст. ошибка	t-статистика	P-значение	
<i>const</i>	3,89441	0,322224	12,0861	<0,0001	***
<i>fox_{t-1}</i>	0,519154	0,0208494	24,9002	<0,0001	***
<i>rabbit_{t-1}</i>	0,47292	0,021555	21,9402	<0,0001	***
Среднее зав. перемен	13,99979		Ст. откл. зав. перемен	0,014166	
Сумма кв. остатков	0,101140		Ст. ошибка модели	0,010077	
R-квадрат	0,494987		Испр. R-квадрат	0,493972	
F(2, 996)	537,2291		P-значение (F)	5,4e-159	
Параметр rho	-0,013491		Стат. Дарбина-Вотсона	2,026794	
F-тесты для нулевых ограничений:					
Все лаги для <i>fox</i> F(1, 996) = 620,02 [0,0000]					
Все лаги для <i>rabbit</i> F(1, 996) = 481,37 [0,0000]					
Уравнение 2: зависимая переменная - <i>rabbit</i>					
Робастные оценки стандартных ошибок (с поправкой на гетероскедастичность), вариант HC1					
	Коэффициент	Ст. ошибка	t-статистика	P-значение	
<i>const</i>	10,2148	0,333848	30,5971	<0,0001	***
<i>fox_{t-1}</i>	-0,512723	0,0218686	-23,4457	<0,0001	***
<i>rabbit_{t-1}</i>	0,493842	0,021801	22,6523	<0,0001	***

Окончание табл. 5.2

Среднее зав. перемен	5,999624	Ст. откл. зав. перемен	0,014549
Сумма кв. остатков	0,104892	Ст. ошибка модели	0,010262
R-квадрат	0,503492	Испр. R-квадрат	0,502495
F(2, 996)	535,4875	P-значение (F)	1,2e-158
Параметр rho	0,065493	Стат. Дарбина-Вотсона	1,864389
F-тесты для нулевых ограничений:			
Все лаги для <i>fox</i> F(1, 996) = 549,7 [0,0000]			
Все лаги для <i>rabbit</i> F(1, 996) = 513,13 [0,0000]			

В обоих уравнениях все объясняющие переменные значимы на 1%-ном уровне значимости.

Запишем уравнения в стандартной форме:

$$\left\{ \begin{array}{l} \widehat{fox}_t = 3,89 + 0,52 fox_{t-1} + 0,47 rabbit_{t-1} \\ \quad (0,32) \quad (0,02) \quad (0,02) \\ \quad R^2 = 0,494 \\ \widehat{rabbit}_t = 10,21 - 0,51 fox_{t-1} + 0,49 rabbit_{t-1} \\ \quad (0,33) \quad (0,02) \quad (0,02) \\ \quad R^2 = 0,5 \end{array} \right.$$

(в) Прогноз популяции кроликов на один год вперед = 6,004.

95%-ный доверительный интервал: (5,984; 6,024).

(г) Одно из условий состоятельности оценивания VAR-модели – стационарность всех временных рядов. Используемые нами временные ряды являются также стационарными. Из этого следует, что:

$$E(fox_t) = E(fox_{t-1}) = f, \quad E(rabbit_t) = E(rabbit_{t-1}) = r.$$

Тогда в равновесии наша система примет вид:

$$\begin{cases} f = 3,89 + 0,52f + 0,47r \\ r = 10,21 - 0,51f + 0,49r. \end{cases}$$

Откуда равновесные уровни численности лис и кроликов в регионе равны $f = 13,999; r = 6,02$.

(д) График отклика функции на импульс, характеризующий динамику численности лис в ответ на импульс численности кроликов, представлен на рис. 5.1.

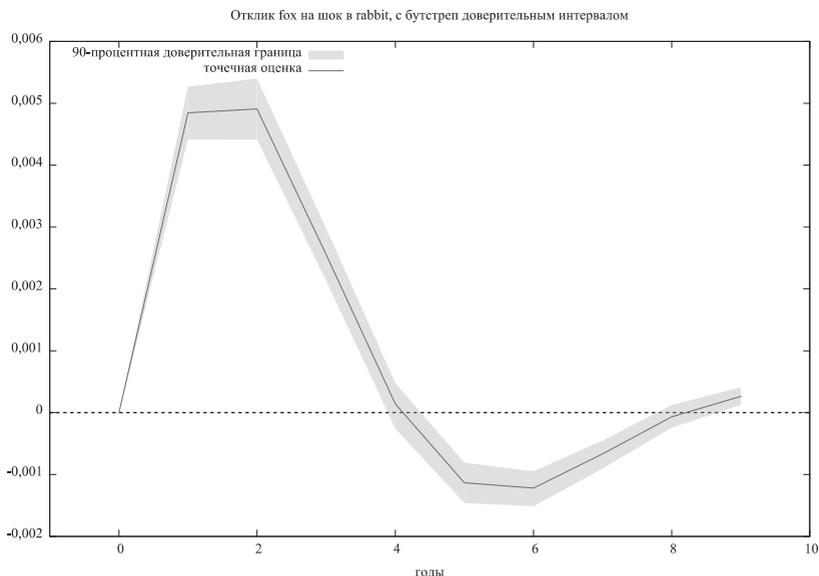


Рис. 5.1. График импульсной функции отклика

Интерпретация: рост численности кроликов сначала вызывает рост численности лис, так как у лис становится много еды и они начинают быстрее плодиться и размножаться. Однако в какой-то момент лис становится слишком много, из-за этого популяция кроликов критически сокращается, лисам становится нечего есть, и их численность падает ниже первоначального уровня, потом происходит возвращение к равновесному значению.

(e) Исходя из построенной нами VAR(1)-модели, мы видим, что и численность лис влияет на численность кроликов, и численность кроликов влияет на численность лис. Таким образом, принимаются обе гипотезы: «fox является причиной по Грейнджеру для rabbit» и «rabbit является причиной по Грейнджеру для fox».

ЗАДАЧА 6

КРИВАЯ ФИЛЛИПСА

В вашем распоряжении (файл **FillipsCurve.xlsx**) месячные данные по core inflation, денежной массе ($M1$), средней зарплате ($Wage$), ценам на импорт ($ImPrices$) по США с марта 2006 по февраль 2014 г.

Вы хотите проверить, соответствуют ли ваши данные гипотезе о расширенной кривой Филлипса. Будем рассматривать следующую ее модификацию:

$$\pi_t = const + \phi(u_{t-1} - u^f) + \alpha m_{t-1} + \beta w_{t-1} + \gamma ImPrice_{t-1} + \pi_t^e + \varepsilon_t,$$

где π_t — core inflation (инфляция, не включающая нефтегазовый и пищевой сектора);

$$m_t = \frac{M1_t - M1_{t-1}}{M1_{t-1}} \text{ — прирост денежной массы;}$$

$$w_t = \frac{Wage_t - Wage_{t-1}}{Wage_{t-1}} \text{ — прирост средней заработной платы;}$$

$$ImPrice_t = \frac{ImPrices_t - ImPrices_{t-1}}{ImPrices_{t-1}} \text{ — прирост цен на импорт;}$$

u_t — фактический уровень безработицы, u^f — естественный уровень безработицы, π_t^e — ожидаемый уровень инфляции.

- (а) Ограничьте выборку: используйте данные начиная с июля 2009 г. Проверьте ряды инфляции и безработицы на стационарность (используйте модели, содержащие 6 лагов). Если они оказались нестационарными, то приведите их к стационарным разностям.
- (б) Проверьте причинно-следственную связь между инфляцией и безработицей на основе теста Грейнджера (используйте модели, содержащие 12 лагов).
- (в) Предположим, что ожидания относительно инфляции в вашем случае описываются авторегрессионной моделью. Подберите подходящую авторегрессионную модель для π_t .
- (г) Оцените предложенную выше модификацию кривой Филлипса. Интерпретируйте полученные результаты, проверьте качество построенной модели.

Решение

(а) Нужно перейти к первым разностям, они окажутся стационарными на уровне значимости 1%: P -значение для инфляции равно 0,0005, P -значение для безработицы составляет $4,475 \cdot 10^{-7}$.

(б) В программном пакете Gretl тест Грейнджера автоматически осуществляется при построении VAR-модели. Построив VAR(12), получим следующие результаты.

Таблица 6.1

Направление влияния	P -значение для теста нулевых ограничений
Инфляция->Безработица	0,3997
Безработица->Инфляция	0,0054

Делаем вывод: безработица является причиной по Грейнджеру для инфляции.

(в) С помощью процедуры подбора ARIMA-модели (см. задачу 1) выберем авторегрессионную модель для инфляции, что приведет нас к модели AR(1). Результаты оценки модели представлены в табл. 6.2.

Таблица 6.2

Модель: ARMA, использованы наблюдения 2009:07-2014:02 (T = 56)				
Зависимая переменная: $\Delta\pi_t$				
Стандартные ошибки рассчитаны на основе Гессииана				
	Коэффициент	Ст. ошибка	Z	P-значение
<i>const</i>	-0,00626862	0,0246844	-0,2540	0,7995
$\Delta\pi_{t-1}$	0,445835	0,121731	3,6625	0,0002 ***
Среднее зав. перемен	-0,002857	Ст. откл. зав. перемен	0,116490	
Среднее инноваций	0,001542	Ст. откл. инноваций	0,103599	
Лог. правдоподобие	47,39354	Крит. Акаике	-88,78708	
Крит. Шварца	-82,71103	Крит. Хеннана-Куинна	-86,43141	

(г) Оценим модификацию кривой Филлипса, перейдя к нужным переменным по денежной массе, средней заработной плате и приросту цен на импорт. Результаты оценки модели представлены в табл. 6.3.

Таблица 6.3

Модель: ARMAX, использованы наблюдения 2009:08-2014:02 (T = 55)					
Зависимая переменная: $\Delta\pi_t$					
Стандартные ошибки рассчитаны на основе Гессииана					
	Коэффициент	Ст. ошибка	Z	P-значение	
<i>const</i>	-0,00173162	0,0277333	-0,0624	0,9502	
$\Delta\pi_{t-1}$	0,468623	0,126575	3,7023	0,0002	***
Δu_{t-1}	-0,00346397	0,0839866	-0,0412	0,9671	
Δm_{t-1}	-1,29173	1,30609	-0,9890	0,3227	
Δw_{t-1}	1,69993	5,79886	0,2931	0,7694	
$\Delta \ln Price_{t-1}$	4,56631	1,71723	2,6591	0,0078	***
Среднее зав. перемен	0,000727	Ст. откл. зав. перемен	0,114404		
Среднее инноваций	0,000500	Ст. откл. инноваций	0,094410		
Лог. правдоподобие	51,64058	Крит. Акаике	-89,28116		
Крит. Шварца	-75,22983	Крит. Хеннана-Куинна	-83,84740		

В полученной модели автокорреляции нет (так как P -значение для соответствующего теста составляет 0,35).

В соответствии с полученными результатами гипотеза о воздействии уровня безработицы на уровень инфляции не подтверждается.

ЗАДАЧА 7

ВОЛАТИЛЬНОСТЬ ИНФЛЯЦИИ И РОСТ ЭКОНОМИКИ

В вашем распоряжении годовые данные по core inflation (инфляция, не включающая нефтегазовый и пищевой сектора) и темпам роста ВВП по США с 1961 по 2014 г. (файл **InfGrowth.xlsx**). Ваша цель – проверить, существует ли негативное влияние волатильности инфляции на темпы роста ВВП страны.

- (а) Проверьте ряды инфляции и роста на стационарность. Если они нестационарны, перейдите к их стационарным разностям.
- (б) Используя однопроцентный уровень значимости, осуществите тест Грейнджера на причинно-следственную связь между инфляцией и темпами роста ВВП. Самостоятельно определите оптимальный порядок лагов.
- (в) Оцените волатильность инфляции с помощью модели GARCH(q,n). Сохраните предсказываемую этой моделью дисперсию ошибок в качестве переменной, характеризующую волатильность инфляции.
- (г) Подберите оптимальную ARMA-модель для темпов роста ВВП.
- (д) Для проверки негативного влияния волатильности инфляции на темпы роста ВВП страны создайте переменную, равную произведению $\Delta \ln$ на переменную волатильности инфляции. Как можно интерпретировать такую переменную? Добавьте ее в подобранную вами в предыдущем пункте модель (используйте саму переменную и ее первый лаг). Интерпретируйте полученные результаты.

Решение

(а) Ряд темпов роста ВВП стационарен (соответствующее P -значение равно 0,0001). Ряд инфляции стационарен только в первых разностях (P -значение = $4,6 \cdot 10^{-10}$).

(б) В программном пакете Gretl тест Грейнджера автоматически осуществляется при построении VAR-модели. Соответственно, для определения оптимального порядка лага нужно воспользоваться встроенной процедурой «Выбор порядка лагов для VAR». Оптимальная модель согласно всем критериям — модель VAR(1). Используем ее для исследования направления причинно-следственной связи между переменными (табл. 7.1).

Таблица 7.1

Направление влияния	<i>P</i> -значение для теста нулевых ограничений
Инфляция->Темпы роста ВВП	0,0084
Темпы роста ВВП->Инфляция	0,0114

На 1%-ном уровне значимости инфляция является причиной по Грейнджеру для темпов роста ВВП.

(в) На основе информационных критериев наилучшей является модель GARCH(1,0).

Таблица 7.2

Модель: GARCH, использованы наблюдения 1962-2014 (T = 53)				
Зависимая переменная: $\Delta\pi_t$				
Стандартные ошибки - QML				
	Коэффициент	Ст. ошибка	Z	P-значение
<i>const</i>	-0,0285475	0,157496	-0,1813	0,8562
α_0	0,773304	0,515602	1,4998	0,1337
α_1	1	0,593068	1,6861	0,0918 *
Среднее зав. перемен	0,010322	Ст. откл. зав. перемен	1,737289	
Лог. правдоподобие	-96,43399	Крит. Акаике	200,8680	
Крит. Шварца	208,7492	Крит. Хеннана-Куинна	203,8987	

Сохраним предсказываемую этой моделью дисперсию ошибок в качестве переменной, характеризующую волатильность инфляции.

(г) Оптимальной ARMA-моделью для темпов роста ВВП является AR(1).

(д) Новая переменная характеризует суммарное влияние инфляции и ее волатильности. Если в какой-то год инфляция была высока, это не обязательно негативно скажется на темпах роста ВВП. Однако если это высокое значение инфляции сопровождается ее высокой волатильностью (т.е. резким изменением по отношению к прошлым значениям), такие колебания могут вызывать негативный эффект на ВВП.

Итоговая модель представлена в табл. 7.3.

Таблица 7.3

Модель: ARMAX, использованы наблюдения 1963-2014 (T = 52)					
Зависимая переменная: y_t					
Стандартные ошибки рассчитаны на основе Гессмана					
	Коэффициент	Ст. ошибка	Z	P-значение	
$const$	3,09808	0,374149	8,2803	<0,0001	***
y_{t-1}	0,32327	0,133966	2,4131	0,0158	**
$\Delta\pi_t \cdot InfVol_t$	0,00170422	0,0166188	0,1025	0,9183	
$\Delta\pi_{t-1} \cdot InfVol_{t-1}$	-0,0379481	0,01651	-2,2985	0,0215	**
Среднее зав. перемен	3,113581	Ст. откл. зав. перемен	2,089035		
Среднее инноваций	-0,008129	Ст. откл. инноваций	1,841199		
Лог. правдоподобие	-105,5817	Крит. Акаике	221,1634		
Крит. Шварца	230,9196	Крит. Хеннана-Куинна	224,9037		

В полученной модели автокорреляции нет (P -значение для соответствующего теста 0,78).

На основе построенной модели можно сделать вывод, что высокая инфляция в совокупности с высокой волатильностью инфляции отрицательно сказывается на темпах роста ВВП.

ЗАДАЧА 8

СТОИМОСТЬ АВИАБИЛЕТОВ

В этом задании вам предлагается исследовать связь между стоимостью авиабилетов, ценами на топливо и доходами населения на примере внутренних рейсов США по данным с 1-го квартала 1995 г. по 4-й квартал 2011 г.

По данным Международной ассоциации воздушного транспорта¹, затраты авиакомпаний на авиационное топливо составляют значительную долю от эксплуатационных расходов (26% в 2010 г.). В то же время в условиях высокой конкуренции среди авиакомпаний США динамика цен на авиабилеты может следовать за динамикой доходов граждан США как основных потребителей данных услуг (поскольку речь идет о внутренних рейсах). Поэтому можно сформулировать следующие гипотезы:

Гипотеза 1. Рост цен на авиационное топливо может привести к росту стоимости авиабилетов;

Гипотеза 2. Рост доходов населения положительно сказывается на динамике цен авиабилетов.

Исходные данные — Tickets.xls:

Fare — средняя² стоимость билетов на внутренних рейсах США, долл.;

Income — реальный среднедушевой квартальный личный располагаемый доход населения США, тыс. долл.³;

Fuel — средняя цена на авиационное топливо в США на начало квартала, долл. за галлон⁴.

- (а) Осуществите для рассматриваемых показателей тест Грейнджера на причинно-следственную связь, перейдя к стационарным рядам и используя первый и второй лаги для каждой из переменных. Интерпретируйте результаты теста.
- (б) На основании результатов тестов постройте модель динамики стоимости авиабилетов в зависимости от изменения цен на топливо и/или среднедушевого дохода. Дайте содержательную интерпретацию.

¹ http://www.iata.org/whatwedo/economics/Pages/industry_outlook.aspx

² Bureau of Transportation Statistics, Airline Origin & Destination Survey (10% Sample) <http://www.transtats.bts.gov/AverageFare/>

³ <http://research.stlouisfed.org/fred2/series/DSPI>

⁴ <http://www.indexmundi.com/commodities/?commodity=jet-fuel&months=300>

Решение

(а) Рассмотрим динамику цен на авиабилеты, цен на топливо и среднедушевых доходов населения (см. рис. 8.1).

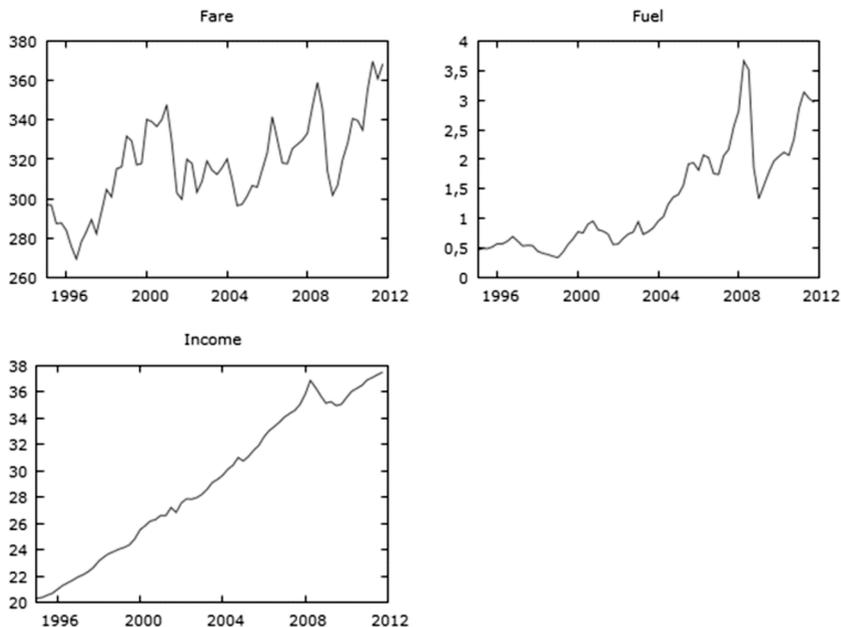


Рис. 8.1. Динамика стоимости авиабилетов, цен на топливо и реального среднедушевого располагаемого дохода в США с 1-го квартала 1995 г. по 4-й квартал 2011 г.

В целом можно говорить о росте с циклическими колебаниями (особенно заметен кризисный спад 2008–2009 гг.) цен на авиабилеты и цен на топливо. В динамике реального среднедушевого располагаемого дохода виден явный повышательный тренд, за исключением 2008–2009 гг. Ряды явно нестационарные, однако необходимо провести формальные тесты.

Таблица 8.1

Результаты расширенного теста Дики—Фуллера

Расширенный тест Дики-Фуллера для Fare
включая 2 лага(-ов) для $(1-L)Fare$
(max was 10, criterion Крит. Акаике)
объем выборки 65
нулевая гипотеза единичного корня: $a = 1$

Окончание табл. 8.1

тест с константой
 модель: $(1-L)y = b_0 + (a-1)y(-1) + \dots + e$
 оценка для $(a - 1)$: -0,0673112
 тестовая статистика: $\tau_{a_c}(1) = -1,05055$
 асимпт. р-значение 0,7371
 коэф. автокорреляции 1-го порядка для e : 0,039
 лаг для разностей: $F(2, 61) = 7,052 [0,0018]$

Расширенный тест Дики-Фуллера для Fuel
 включая 2 лага(-ов) для $(1-L)Fuel$
 (max was 10, criterion Крит. Акаике)
 объем выборки 65
 нулевая гипотеза единичного корня: $a = 1$

тест с константой
 модель: $(1-L)y = b_0 + (a-1)y(-1) + \dots + e$
 оценка для $(a - 1)$: -0,021977
 тестовая статистика: $\tau_{a_c}(1) = -0,558816$
 асимпт. р-значение 0,8772
 коэф. автокорреляции 1-го порядка для e : 0,033
 лаг для разностей: $F(2, 61) = 11,145 [0,0001]$

Расширенный тест Дики-Фуллера для Income
 включая один лаг для $(1-L)Income$
 (max was 10, criterion Крит. Акаике)
 объем выборки 66
 нулевая гипотеза единичного корня: $a = 1$

тест с константой
 модель: $(1-L)y = b_0 + (a-1)y(-1) + \dots + e$
 оценка для $(a - 1)$: -0,00448662
 тестовая статистика: $\tau_{a_c}(1) = -0,668291$
 асимпт. р-значение 0,8527
 коэф. автокорреляции 1-го порядка для e : -0,013

По результатам расширенного теста Дики—Фулера нулевая гипотеза о наличии единичного корня не отвергается для всех трех временных рядов. Соответствующие p -значения составляют 0,74, 0,87 и 0,85 и позволяют сделать вывод о нестационарности рядов. Поэтому для дальнейшего анализа необходимо перейти к первым разностям. Для каждой переменной создадим первую разность (см. рис. 8.2).

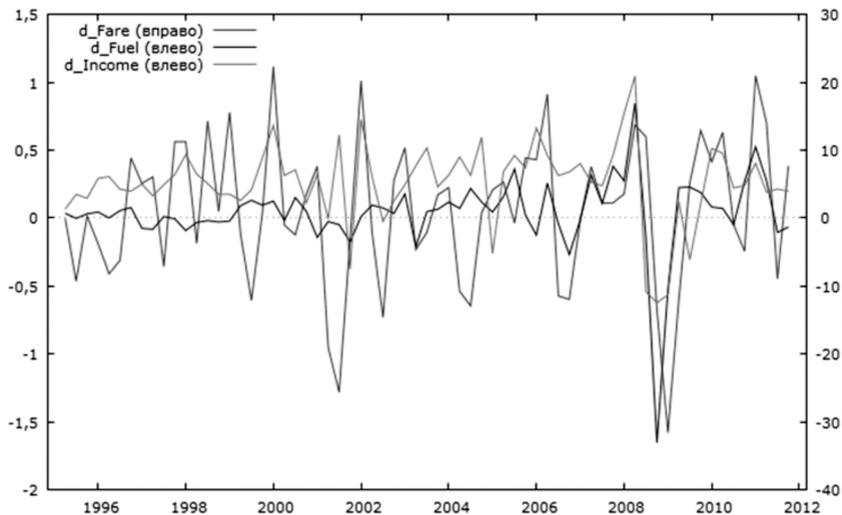


Рис. 8.2. Динамика изменений цен на авиабилеты (d_Fare), цен на авиационное топливо (d_Fuel) и реального среднедушевого дохода (d_Income) в США в 1995–2011 гг.

Для первых разностей также проведем расширенный тест Дики—Фулера.

Таблица 8.2

Результаты расширенного теста Дики—Фулера для первых разностей

Расширенный тест Дики-Фуллера для d_Fare

включая один лаг для $(1-L)d_Fare$
(max was 10, criterion Крит. Акаике)

объем выборки 65

нулевая гипотеза единичного корня: $a = 1$

тест с константой

модель: $(1-L)y = b_0 + (a-1)y(-1) + \dots + e$

оценка для $(a - 1)$: -1,11263

тестовая статистика: $\tau_{a-c}(1) = -7,68171$

асимпт. р-значение 4,181e-012

коэф. автокорреляции 1-го порядка для e : 0,031

Расширенный тест Дики-Фуллера для d_Fuel

включая один лаг для $(1-L)d_Fuel$
(max was 10, criterion Крит. Акаике)

объем выборки 65

нулевая гипотеза единичного корня: $a = 1$

Окончание табл. 8.2

тест с константой
 модель: $(1-L)y = b_0 + (a-1)y(-1) + \dots + e$
 оценка для $(a - 1)$: -0,978971
 тестовая статистика: $\tau_{c(1)} = -7,31626$
 асимпт. р-значение 4,497e-011
 коэф. автокорреляции 1-го порядка для e : 0,027

Расширенный тест Дики-Фуллера для d_Income
 включая 0 лага(-ов) для $(1-L)d_Income$
 (max was 10, criterion Крит. Акаике)
 объем выборки 66
 нулевая гипотеза единичного корня: $a = 1$

тест с константой
 модель: $(1-L)y = b_0 + (a-1)y(-1) + e$
 оценка для $(a - 1)$: -0,789719
 тестовая статистика: $\tau_{c(1)} = -6,48274$
 Р-значение 4,165e-007
 коэф. автокорреляции 1-го порядка для e : -0,011

По результатам теста нулевая гипотеза о наличии единичного корня отвергается, поскольку p -значение меньше любого разумного уровня значимости (близко к нулю). Временной ряд изменений цен на авиабилеты стационарен. Аналогичный результат получен для первых разностей цен на авиационное топливо и первых разностей реального среднедушевого располагаемого дохода. Ряды стационарны, для них можно проводить тест Грейнджера.

Тест Грейнджера для изменения стоимости билетов и изменения цен на топливо представлен в табл. 8.3. По его результатам нулевая гипотеза об отсутствии влияния лагов переменной d_Fuel на переменную d_Fare отвергается на любом уровне значимости (p -значение равно 0,0001), а гипотеза об отсутствии влияния лагов переменной d_Fare на переменную d_Fuel не отвергается (p -значение равно 0,6945). Таким образом, можно сделать вывод о том, что изменение цен на топливо является причиной по Грейнджеру для изменения стоимости авиабилетов.

Таблица 8.3

Тест Грейнджера для изменения стоимости билетов и цен на топливо

Метод оценки - МНК, наблюдения 1995:4-2011:4 (T = 65)				
Уравнение 1: d_Fare				
	Коэффициент	Ст. ошибка	t-статистика	P-значение
Const	0,874718	1,12939	0,7745	0,4417
d_Fare_1	0,111771	0,12582	0,8883	0,3779
d_Fare_2	-0,393048	0,128821	-3,0511	0,0034 ***

Окончание табл. 8.3

d_Fuel_1	12,5477	3,59751	3,4879	0,0009	***
d_Fuel_2	5,12289	4,18632	1,2237	0,2258	
Среднее зав. перемен	1,244308	Ст. откл. зав. перемен		10,85767	
Сумма кв. остатков	5090,263	Ст. ошибка модели		9,210739	
R-квадрат	0,325337	Испр. R-квадрат		0,280360	
F(4, 60)	14,66956	P-значение (F)		2,06e-08	
Параметр rho	0,019558	Стат. Дарбина-Вотсона		1,927982	
F-тесты для нулевых ограничений:					
Все лаги для d_FareF(2, 60) = 4,6969 [0,0127]					
Все лаги для d_FuelF(2, 60) = 11,012 [0,0001]					
Уравнение 2: d_Fuel					
	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>t-статистика</i>	<i>P-значение</i>	
Const	0,0384713	0,0272772	1,4104	0,1636	
d_Fare_1	-0,00175053	0,00211395	-0,8281	0,4109	
d_Fare_2	-0,000418936	0,00193304	-0,2167	0,8292	
d_Fuel_1	0,47089	0,167847	2,8055	0,0068	***
d_Fuel_2	-0,404416	0,279349	-1,4477	0,1529	
Среднее зав. перемен	0,038000	Ст. откл. зав. перемен		0,285663	
Сумма кв. остатков	3,783463	Ст. ошибка модели		0,251113	
R-квадрат	0,275562	Испр. R-квадрат		0,227266	
F(4, 60)	2,415940	P-значение (F)		0,058494	
Параметр rho	0,030020	Стат. Дарбина-Вотсона		1,939413	
F-тесты для нулевых ограничений:					
Все лаги для d_FareF(2, 60) = 0,36678 [0,6945]					
Все лаги для d_FuelF(2, 60) = 4,1987 [0,0197]					

Таблица 8.4

Тест Грейнджера для изменения стоимости билетов и изменения доходов

Метод оценки - МНК, наблюдения 1995:4-2011:4 (T = 65)					
Уравнение 1: d_Fare					
	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>t-статистика</i>	<i>P-значение</i>	
Const	-1,98422	1,88562	-1,0523	0,2969	
d_Fare_1	0,194417	0,122213	1,5908	0,1169	
d_Fare_2	-0,459341	0,119538	-3,8426	0,0003	***
d_Income_1	7,72514	4,27904	1,8053	0,0760	*
d_Income_2	5,96002	4,38223	1,3600	0,1789	

Окончание табл. 8.4

Среднее зав. перемен	1,244308	Ст. откл. зав. перемен	10,85767		
Сумма кв. остатков	5474,040	Ст. ошибка модели	9,551649		
R-квадрат	0,274472	Испр. R-квадрат	0,226103		
F(4, 60)	4,897670	P-значение (F)	0,001746		
Параметр rho	0,033908	Стат. Дарбина-Вотсона	1,892812		
F-тесты для нулевых ограничений:					
Все лаги для $d_FareF(2, 60) = 7,6825 [0,0011]$					
Все лаги для $d_IncomeF(2, 60) = 2,9737 [0,0587]$					
Уравнение 2: d_Income					
	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>t-статистика</i>	<i>P-значение</i>	
Const	0,197793	0,0776708	2,5466	0,0135	**
d_Fare_1	0,000184767	0,00398805	0,0463	0,9632	
d_Fare_2	-0,000372504	0,00466983	-0,0798	0,9367	
d_Income_1	0,197961	0,264907	0,7473	0,4578	
d_Income_2	0,0454037	0,212047	0,2141	0,8312	
Среднее зав. перемен	0,260855	Ст. откл. зав. перемен	0,295955		
Сумма кв. остатков	5,350659	Ст. ошибка модели	0,298626		
R-квадрат	0,045501	Испр. R-квадрат	-0,018132		
F(4, 60)	0,229420	P-значение (F)	0,920835		
Параметр rho	0,006843	Стат. Дарбина-Вотсона	1,984458		
F-тесты для нулевых ограничений:					
Все лаги для $d_FareF(2, 60) = 0,0050269 [0,9950]$					
Все лаги для $d_IncomeF(2, 60) = 0,33309 [0,7180]$					

По результатам теста Грейнджера нулевая гипотеза об отсутствии влияния лагов переменной d_Income на переменную d_Fare не отвергается на 5%-ном уровне значимости (p -значение равно 0,0587). Гипотеза об отсутствии влияния лагов переменной d_Fare на переменную d_Income также не отвергается на 5%-ном уровне значимости (p -значение равно 0,995). Таким образом, нельзя сделать вывод о том, что изменение среднедушевого располагаемого дохода является причиной по Грейнджеру для изменения стоимости авиабилетов, что говорит против использования лагов среднедушевого дохода при моделировании динамики стоимости авиабилетов и косвенно свидетельствует против гипотезы № 2 из условия задачи.

(б) На основании результатов тестов построим модель, описывающую изменение стоимости авиабилетов в зависимости от изменения цен на топливо в текущем и двух предшествующих периодах (табл. 8.5).

Таблица 8.5

Модель 3: МНК, использованы наблюдения 1995:4-2011:4 (T = 65)				
Зависимая переменная: d_Fare				
Стандартные ошибки HAC, ширина окна 3 (Ядро Бартлетта (Bartlett))				
	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>t-статистика</i>	<i>P-значение</i>
Const	0,123661	0,984429	0,1256	0,9004
d_Fuel	11,9433	3,64426	3,2773	0,0017 ***
d_Fuel_1	11,4849	3,83926	2,9914	0,0040 ***
d_Fuel_2	5,32934	2,843	1,8746	0,0656 *
Среднее зав. перемен	1,244308	Ст. откл. зав. перемен	10,85767	
Сумма кв. остатков	5498,699	Ст. ошибка модели	9,494346	
R-квадрат	0,271203	Испр. R-квадрат	0,235361	
F(3, 61)	23,88270	P-значение (F)	2,41e-10	
Лог. Правдоподобие	-236,4621	Крит. Акаике	480,9242	
Крит. Шварца	489,6217	Крит. Хеннана-Куинна	484,3559	
Параметр rho	0,083777	Стат. Дарбина-Вотсона	1,822512	

Все коэффициенты в модели, кроме константы, являются значимыми. По результатам теста Дарбина—Уотсона нулевая гипотеза об отсутствии автокорреляции в остатках не отвергается, поскольку расчетная статистика равна 1,82 и лежит между «верхним Дарбином» (равным 1,69) и 2, т.е. в «зоне отсутствия автокорреляции». Модель можно считать удовлетворительной. Это косвенно свидетельствует в пользу гипотезы № 1.

Дадим интерпретацию коэффициентов. Рост цен на авиационное топливо на 1 долл. за галлон приводит к мгновенному (т.е. в том же периоде) росту цен на авиабилеты на 11,9 долл.; росту цен на авиабилеты на 11,5 долл. по итогам следующего квартала; росту цен на авиабилеты на 5,33 долл. спустя два квартала.

ЗАДАЧА 9

СПИСЫВАНИЕ

Вам предлагается реплицировать результаты исследования, которое проводилось на данных, собранных на экономическом факультете одного из ведущих российских университетов.

Исходный файл с данными: **cheating**. В этом файле также доступно краткое описание переменных. Их развернутое описание приводится далее.

Основой для исследования послужили результаты двух потоковых контрольных работ по одному из предметов, написанных второкурсниками в весеннем семестре 2016 г. Обе контрольные работы состояли из двух частей (максимальная оценка за каждую из частей составляла 10 баллов):

- пять теоретических вопросов по два балла каждый (общий список вопросов был известен заранее, и ответы на них присутствовали в доступных студентам учебных материалах);
- пять практических задач по два балла каждая (для их решения нужно было применить ряд идей и алгоритмов, разбиравшихся на лекциях и семинарах).

Таким образом, для списывания первой части мог пригодиться конспект лекций, учебник, составленная заранее шпаргалка, а также любое электронное устройство, на которое скачаны учебные материалы. Для списывания второй части могли помочь конспекты лекций и семинаров с образцами решенных задач, но без соответствующей подготовки студента они были почти бесполезны. Продвинутые методы списывания с использованием современной техники и третьих лиц, находящихся вне аудитории, были практически недоступны ввиду отсутствия приема сотового сигнала и отсутствия wi-fi в аудиториях, где проводились работы.

Каждая контрольная работа проводилась в двух аудиториях, студенты делились на две части по алфавитному признаку. Те, чьи фамилии начинаются на буквы А-Л (они составили экспериментальную (treatment) группу), писали в одной аудитории, остальные (они составили контрольную (control) группу) – в другой. Общее число студентов составило 116 человек, по 58 человек в каждой группе.

На первой контрольной работе преподаватели прилагали максимум усилий для предотвращения списывания:

- в аудиторию не допускались студенты в верхней одежде;
- личные вещи студентов (сумки и т.п.) складывались на лекторском подиуме;

- рассадка студентов осуществлялась преподавателями;
- тетради для написания работы раздавались преподавателями;
- во время написания работы преподаватели внимательно наблюдали за студентами.

На второй работе преподаватели, дежурившие в контрольной группе, вели себя так же строго, как обычно. При этом преподаватели, дежурившие в экспериментальной группе, следили за студентами лишь формально:

- допускалось присутствие в верхней одежде/ее складирование рядом со студентом;
- допускался пронос личных вещей (сумок и т.п.) за парты;
- при осуществлении рассадки преподаватели следили лишь за чередованием вариантов, не уделяя внимания наличию пространства между студентами и между рядами студентов;
- разрешалось использовать свои тетради;
- во время написания работы преподаватели занимались своими делами за кафедрой (чтением, проверкой других работ), осуществляя за студентами лишь визуальный контроль «с места».

По результатам каждой контрольной работы были собраны следующие данные:

- *балл за контрольную работу*
Возможные значения — от 0 до 20. Характеризует общий результат за контрольную работу;
- *балл за теоретическую часть контрольной работы*
Возможные значения — от 0 до 10. Характеризует результат теоретической (более пригодной для списывания) части работы. **Эта переменная используется в качестве зависимой;**
- *пол студента*
Бинарная переменная, характеризующая пол студента (равна единице для девушек и нулю для юношей);
- *преподаватель, проверявший работу/номер варианта*
Каждая работа была представлена в четырех вариантах и проверялась соответственно одним из четырех преподавателей по единому критерию. Студенты распределялись по вариантам случайным образом. Чтобы учесть возможный дисбаланс между различными вариантами, а также стилями проверки, было решено использовать в качестве контрольной переменной номер варианта/проверяющего. Для этого применяются три фиктивные переменные: *professor_1*, *professor_2* и *professor_3*, соответствующие трем из четырех преподавателей;
- *номер группы*
Возможные значения — от одного до шести. Студенты экономического факультета делятся по группам по результатам внутреннего экзамена по математике. Соответственно, номер группы от-

ражает исходный уровень подготовки студентов. Кроме того, его использование позволяет учесть различие в преподавателях, которые вели семинарские занятия.

Следует отметить также, что академическая группа № 1 характеризуется специфическими свойствами. Это так называемая группа повышенной академической нагрузки («Первая группа»). В ней собраны наиболее сильные и мотивированные студенты курса, которые изучают ряд дисциплин по программам повышенной сложности. Кроме того, в этой группе в процессе обучения делается особый акцент на необходимость соблюдения стандартов академической этики. Предполагается, что в этой группе студенты более мотивированы к учебе и потому ведут себя честно даже в условиях, когда есть возможность для мошенничества.

Результирующая выборка представляет собой двухпериодную панель. Нулевой период — это первая контрольная работа, первый период — вторая контрольная работа. Наблюдения, соответствующие студентам, посетившим только одну из контрольных работ, были отброшены. Таким образом, итоговая выборка представляет собой сбалансированную панель для 116 наблюдений и двух периодов, индекс i соответствует номеру студента, индекс t соответствует периоду времени.

Выберите подходящую эмпирическую стратегию, осуществите необходимые преобразования данных и эконометрические расчеты для того, чтобы ответить на два исследовательских вопроса:

- *Верно ли, что студенты списывают, если предоставить им такую возможность? (Иными словами, верно ли, что создание более удобных для списывания условий является причиной получения более высокого балла по теории?)*
- *Верно ли что студенты Первой группы не списывают?*

Решение

Создадим дополнительную переменную $treat_effect$, которая равна произведению переменных t и $A_L_student$.

Также создадим фиктивные переменные для $gr1$, $gr2$, ..., $gr6$, соответствующие каждой из шести академических групп.

Будем использовать метод «разность разностей». Сначала оценим модель на полной выборке студентов без использования контрольных переменных.

Таблица 9.1

Модель 1: Объединенный (pooled) МНК, использовано наблюдений – 232					
Включено 116 пространственных объектов					
Длина временного ряда = 2					
Зависимая переменная: theory					
Робастные стандартные ошибки (НАС)					
	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>t-статистика</i>	<i>P-значение</i>	
Const	4,90517	0,423987	11,5692	<0,0001	***
treat_effect	0,715517	0,589077	1,2146	0,2258	
T	1,37069	0,4061	3,3753	0,0009	***
A_L_student	-0,0862069	0,55764	-0,1546	0,8773	
Среднее зав. перемен	5,726293	Ст. откл. зав. перемен	3,132271		
Сумма кв. остатков	2081,392	Ст. ошибка модели	3,021409		
R-квадрат	0,081618	Испр. R-квадрат	0,069534		
F(3, 228)	11,81748	P-значение (F)	3,18e-07		
Лог. правдоподобие	-583,7041	Крит. Акаике	1175,408		
Крит. Шварца	1189,195	Крит. Хеннана-Куинна	1180,968		

Затем оценим аналогичную модель с включением контрольных переменных (табл. 9.2).

Таблица 9.2

Модель 2: Объединенный (pooled) МНК, использовано наблюдений - 232					
Включено 116 пространственных объектов					
Длина временного ряда = 2					
Зависимая переменная: theory					
Робастные стандартные ошибки (НАС)					
	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>t-статистика</i>	<i>P-значение</i>	
Const	4,30813	0,603592	7,1375	<0,0001	***
treat_effect	0,732293	0,572036	1,2802	0,2018	
T	1,42248	0,383727	3,7070	0,0003	***
A_L_student	-0,142731	0,486853	-0,2932	0,7697	
professor_1	1,11305	0,491874	2,2629	0,0246	**
professor_2	2,36615	0,566298	4,1783	<0,0001	***
professor_3	1,86851	0,480743	3,8867	0,0001	***
gr2	-1,12917	0,67168	-1,6811	0,0942	*
gr3	-0,806949	0,717496	-1,1247	0,2620	
gr4	-1,85715	0,634792	-2,9256	0,0038	***
gr5	-2,76597	0,684855	-4,0388	<0,0001	***
gr6	-1,89795	0,856007	-2,2172	0,0276	**
Female	1,06176	0,4108	2,5846	0,0104	**

Окончание табл. 9.2

Среднее зав. перемен	5,726293	Ст. откл. зав. перемен	3,132271
Сумма кв. остатков	1667,619	Ст. ошибка модели	2,759474
R-квадрат	0,264189	Испр. R-квадрат	0,223871
F(12, 219)	10,56767	P-значение (F)	2,03e-16
Лог. правдоподобие	-557,9938	Крит. Акаике	1141,988
Крит. Шварца	1186,795	Крит. Хеннана-Куинна	1160,058

Наконец, оценим модель, включив в нее фиксированные эффекты для каждого студента. В этом случае контролировать на переменные $A_L_student$, $gr2$, $gr3$, $gr4$, $gr5$, $gr6$ и $female$ не нужно, так как они неизменны во времени (см. табл. 9.3).

Таблица 9.3

Модель 3: Фиксированные эффекты, использовано наблюдений - 232					
Включено 116 пространственных объектов					
Длина временного ряда = 2					
Зависимая переменная: theory					
Робастные стандартные ошибки (НАС)					
Пропущены из-за совершенной коллинеарности: A_L_student					
gr2 gr3 gr4 gr5 gr6 female					
	Коэффициент	Ст. ошибка	t-статистика	P-значение	
const	3,75066	0,463103	8,099	7,86e-013	***
treat_effect	0,748723	0,569065	1,316	0,1910	
t	1,39753	0,386153	3,619	0,0004	***
professor_1	1,24073	0,555161	2,235	0,0274	**
professor_2	1,79626	0,751576	2,390	0,0185	**
professor_3	1,35048	0,612354	2,205	0,0295	**
Среднее зав. перемен	5,726293	Ст. откл. зав. перемен	3,132271		
Сумма кв. остатков	542,2890	Ст. ошибка модели	2,210314		
LSDV R-squared	0,760723	В пределах R-квадрат	0,290546		
Лог. правдоподобие	-427,6849	Крит. Акаике	1097,370		
Крит. Шварца	1514,425	Крит. Хеннана-Куинна	1265,564		
Joint test on named regressors -					
Тестовая статистика: F(5, 111) = 8,72925					
p-значение = P(F(5, 111) > 8,72925) = 5,23931e-007					
Robust test for differing group intercepts -					
Нулевая гипотеза: Группы имеют общие константы					
Тестовая статистика: Welch F(115, 39,3) = 52,0154					
p-значение = P(F(115, 39,3) > 52,0154) = 6,40327e-026					

Для ответа на первый исследовательский вопрос, сформулированный в задаче, нас интересует значимость коэффициента при переменной *treat_effect*. Если она значима, то можно заключить, что создание более благоприятных условий для списывания само по себе приводит к росту результатов студентов.

Отметим, что во всех моделях этот эффект воздействия оказывается незначимым. Возможно, дело в том, что студенты вовсе не списывают. Однако есть и альтернативное объяснение: возможно, выборка, с которой мы работаем, является сильно неоднородной. Например, из-за того, что поведение студентов Первой группы сильно отличается от поведения остальных студентов.

Для проверки этого предположения оценим три аналогичные модели, используя только данные для студентов групп со второй по пятую, т.е. исключим из выборки Первую группу (см. табл. 9.4, 9.5 и 9.6).

Таблица 9.4

Модель 4: Объединенный (pooled) МНК, использовано наблюдений - 184					
Включено 92 пространственных объектов					
Длина временного ряда = 2					
Зависимая переменная: <i>theory</i>					
Робастные стандартные ошибки (HAC)					
	<i>Коэффициент</i>	<i>Ст. ошибка</i>	<i>t-статистика</i>	<i>P-значение</i>	
Const	4,65217	0,482165	9,6485	<0,0001	***
<i>treat_effect</i>	1,42391	0,663344	2,1466	0,0332	**
T	1,30435	0,481824	2,7071	0,0074	***
A_L_student	-0,445652	0,610139	-0,7304	0,4661	
Среднее зав. перемен	5,437500	Ст. откл. зав. перемен	3,167916		
Сумма кв. остатков	1622,940	Ст. ошибка модели	3,002721		
R-квадрат	0,116301	Испр. R-квадрат	0,101573		
F(3, 180)	14,42377	P-значение (F)	1,84e-08		
Лог. правдоподобие	-461,3741	Крит. Акаике	930,7482		
Крит. Шварца	943,6080	Крит. Хеннана-Куинна	935,9604		

Таблица 9.5

Модель 5: Объединенный (pooled) МНК, использовано наблюдений – 184					
Включено 92 пространственных объектов					
Длина временного ряда = 2					
Зависимая переменная: theory					
Робастные стандартные ошибки (НАС)					
	Коэффициент	Ст. ошибка	t-статистика	P-значение	
Const	1,97265	0,94448	2,0886	0,0382	**
treat_effect	1,49289	0,642993	2,3218	0,0214	**
T	1,28426	0,455865	2,8172	0,0054	***
A_L_student	-0,357	0,551601	-0,6472	0,5184	
professor_1	1,39053	0,534556	2,6013	0,0101	**
professor_2	2,42236	0,646372	3,7476	0,0002	***
professor_3	1,92298	0,532074	3,6141	0,0004	***
gr2	0,902092	0,876729	1,0289	0,3050	
gr3	1,21098	0,901909	1,3427	0,1811	
gr4	0,12719	0,825162	0,1541	0,8777	
gr5	-0,781423	0,92301	-0,8466	0,3984	
Female	1,45255	0,459519	3,1610	0,0019	***
Среднее зав. перемен	5,437500	Ст. откл. зав. перемен	3,167916		
Сумма кв. остатков	1285,963	Ст. ошибка модели	2,734325		
R-квадрат	0,299787	Испр. R-квадрат	0,255006		
F(11, 172)	13,44915	P-значение (F)	2,43e-18		
Лог. правдоподобие	-439,9628	Крит. Акаике	903,9257		
Крит. Шварца	942,5049	Крит. Хеннана-Куинна	919,5623		

Таблица 9.6

Модель 6: Фиксированные эффекты, использовано наблюдений - 184					
Включено 92 пространственных объектов					
Длина временного ряда = 2					
Зависимая переменная: theory					
Робастные стандартные ошибки (НАС)					
Пропущены из-за совершенной коллинеарности: gr2 gr3 gr4 gr5 female					
	Коэффициент	Ст. ошибка	t-статистика	P-значение	
const	3,20939	0,464268	6,913	7,53e-010	***
treat_effect	1,49891	0,642449	2,333	0,0219	**
t	1,29047	0,459737	2,807	0,0062	***
professor_1	1,41917	0,575786	2,465	0,0157	**

Окончание табл. 9.6

professor_2	1,84959	0,765888	2,415	0,0178	**
professor_3	1,44583	0,641661	2,253	0,0268	**
Среднее зав. перемен	5,437500		Ст. откл. зав. перемен	3,167916	
Сумма кв. остатков	425,5663		Ст. ошибка модели	2,211689	
LSDV R-squared	0,768277		В пределах R-квадрат	0,370348	
Лог. правдоподобие	-338,2253		Крит. Акаике	870,4506	
Крит. Шварца	1182,299		Крит. Хеннана-Куинна	996,8468	
Joint test on named regressors -					
Тестовая статистика: $F(5, 87) = 11,4392$					
p-значение = $P(F(5, 87) > 11,4392) = 1,68605e-008$					
Robust test for differing group intercepts -					
Нулевая гипотеза: Группы имеют общие константы					
Тестовая статистика: Welch $F(91, 31,3) = 27,5234$					
p-значение = $P(F(91, 31,3) > 27,5234) = 8,71689e-017$					

Во всех трех моделях коэффициент при переменной *treat_effect* оказывается положительным, статистически значимым и составляет около полутора баллов. Таким образом, если анализировать выборку студентов, не имеющих отношения к Первой группе, то следует сделать вывод о том, что они списывают.

Теперь оценим три указанные модели, анализируя выборку только из студентов Первой группы (см. табл. 9.7, 9.8 и 9.9).

Таблица 9.7

Модель 7: Объединенный (pooled) МНК, использовано наблюдений - 48					
Включено 24 пространственных объектов					
Длина временного ряда = 2					
Зависимая переменная: theory					
Робастные стандартные ошибки (HAC)					
	Коэффициент	Ст. ошибка	t-статистика	P-значение	
Const	5,875	0,827322	7,1012	<0,0001	***
treat_effect	-2	0,998697	-2,0026	0,0514	*
T	1,625	0,659137	2,4653	0,0177	**
A_L_student	1,29167	1,05649	1,2226	0,2280	
Среднее зав. перемен	6,833333		Ст. откл. зав. перемен	2,750886	
Сумма кв. остатков	337,9583		Ст. ошибка модели	2,771438	
R-квадрат	0,049789		Испр. R-квадрат	-0,014998	
F(3, 44)	2,153702		P-значение (F)	0,107048	
Лог. Правдоподобие	-114,9504		Крит. Акаике	237,9007	
Крит. Шварца	245,3855		Крит. Хеннана-Куинна	240,7293	

Таблица 9.8

Модель 8: Объединенный (pooled) МНК, использовано наблюдений - 48					
Включено 24 пространственных объектов					
Длина временного ряда = 2					
Зависимая переменная: theory					
Робастные стандартные ошибки (НАС)					
	Коэффициент	Ст. ошибка	t-статистика	P-значение	
Const	5,09365	0,755577	6,7414	<0,0001	***
treat_effect	-2,08873	1,04612	-1,9966	0,0527	*
T	1,89914	0,642875	2,9541	0,0052	***
A_L_student	1,00011	0,931908	1,0732	0,2896	
Female	-0,324315	0,881354	-0,3680	0,7148	
professor_1	0,798984	1,16968	0,6831	0,4985	
professor_2	1,91149	0,90764	2,1060	0,0415	**
professor_3	1,48811	1,11287	1,3372	0,1887	
Среднее зав. перемен	6,833333	Ст. откл. зав. перемен	2,750886		
Сумма кв. остатков	305,0893	Ст. ошибка модели	2,761744		
R-квадрат	0,142204	Испр. R-квадрат	-0,007910		
F(7, 40)	2,443041	P-значение (F)	0,034917		
Лог. правдоподобие	-112,4947	Крит. Акаике	240,9895		
Крит. Шварца	255,9591	Крит. Хеннана-Куинна	246,6465		

Таблица 9.9

Модель 9: Фиксированные эффекты, использовано наблюдений - 48					
Включено 24 пространственных объектов					
Длина временного ряда = 2					
Зависимая переменная: theory					
Робастные стандартные ошибки (НАС)					
Пропущены из-за совершенной коллинеарности: A_L_student female					
	Коэффициент	Ст. ошибка	t-статистика	P-значение	
const	5,88547	1,25591	4,686	0,0002	***
treat_effect	-2,03925	0,967079	-2,109	0,0485	**
t	1,67111	0,736259	2,270	0,0351	**
professor_1	1,02943	1,84500	0,5580	0,5834	
professor_2	1,07064	1,87140	0,5721	0,5740	
professor_3	0,770801	1,35831	0,5675	0,5770	
Среднее зав. перемен	6,833333	Ст. откл. зав. перемен	2,750886		
Сумма кв. остатков	69,41533	Ст. ошибка модели	1,911397		

Окончание табл. 9.9

LSDV R-squared	0,804830	В пределах R-квадрат	0,215646
Лог. правдоподобие	-76,96281	Крит. Акаике	211,9256
Крит. Шварца	266,1904	Крит. Хеннана-Куинна	232,4324
Joint test on named regressors -			
Тестовая статистика: $F(5, 19) = 1,7574$			
p-значение = $P(F(5, 19) > 1,7574) = 0,170063$			
Robust test for differing group intercepts -			
Нулевая гипотеза: Группы имеют общие константы			
Тестовая статистика: Welch $F(23, 8,6) = 25,653$			
p-значение = $P(F(23, 8,6) > 25,653) = 1,78373e-005$			

Все три модели позволяют уверенно утверждать, что в аудитории, где были созданы условия для списывания, средний балл студентов Первой группы не вырос (последняя из моделей фиксирует даже падение среднего балла, которое значимо при уровне значимости 5%, однако в остальных моделях переменная *treat_effect* не значима при этом уровне значимости).

Таким образом, ответ на второй исследовательский вопрос, сформулированный в этой задаче, состоит в том, что студенты Первой группы не склонны списывать. Это может объясняться либо тем, что в Первой группе, действительно, не принято так себя вести, либо тем, что студенты Первой группы были настолько хорошо готовы к контрольной, что у них не было нужды мошенничать, либо обоими этими эффектами одновременно.

ЗАДАЧА 10

ЗАРАБОТНАЯ ПЛАТА В МЕКСИКЕ

Файл с данными **Mexico.Dat**

Переменные, значения которых меняются во времени

Wave	время (1-й кв. 1992 – 1-й кв. 1993 г.)
Age	возраст респондента
Employ	=1 если работал во время проведения интервью
Unemp	=1 если был безработным
Nonpa	=1 если не участвовал в рынке труда
Formm	=1 если работал легально
Informm	=1 если работал нелегально
Earn	месячный заработок, в песо
Nch6	количество детей (в возрасте до 6 лет)
Nadu	количество взрослых в семье
Ofinc	доходы других членов семьи
Singl	=1 если холост
Marrid	=1 если женат
Divo	=1 если в разводе
Learn	логарифм месячного заработка
Lwage	логарифм часовой ставки заработной платы

переменные, значения которых постоянны во времени

Pid	код индивида
Vstart	первый квартал, когда респондента интервьюировали
Vend	последний квартал, когда респондента интервьюировали
Noschl	=1 если нет образования
Basicl	=1 если учился не более 6 лет
Mexico	=1 если проживает в Мехико
Juare	=1 если проживает в Суидад-Хуарез
Tijua	=1 если проживает в Тихуане
Guada	=1 если проживает в Гвадалахаре
Monte	=1 если проживает в Монтерее
Midd	=1 если учился от 7 до 12 лет
Higg	если учился более 12 лет

- (а) Оцените простую полную регрессию логарифма часовой ставки заработной платы по переменным, характеризующим уровень образования и место проживания респондента.
- (б) Оцените уравнение из предыдущего пункта, используя модель со случайными эффектами. Сравните полученную модель с моделью

из пункта (а), опираясь на результаты соответствующего теста. Что можно сказать о воздействии образования на доход?

- (в) Почему на основе доступных данных не удастся оценить воздействие образования на доход при помощи модели с фиксированными эффектами?
- (г) В рамках модели со случайными эффектами оцените различия в уровне заработной плате легальных и нелегальных рабочих с разным уровнем образования.

Решение

(а) Оценка модели с использованием обычного МНК дает следующие результаты (табл. 10.1).

Таблица 10.1

Модель 1: Объединенный (pooled) МНК, использовано наблюдений - 4901					
Включено 1346 пространственных объектов					
Длина временного ряда: минимум 1, максимум 5					
Зависимая переменная: <i>lwage</i>					
Робастные стандартные ошибки (HAC)					
	Коэффициент	Ст. ошибка	<i>t</i> -статистика	<i>P</i> -значение	
const	1,37789	0,0565678	24,3582	<0,0001	***
basic1	0,229812	0,0531658	4,3225	<0,0001	***
midd	0,416238	0,0553758	7,5166	<0,0001	***
higg	1,24362	0,0672305	18,4978	<0,0001	***
mexio	-0,23718	0,0407407	-5,8217	<0,0001	***
juare	0,0391107	0,0480464	0,8140	0,4157	
tijua	0,297931	0,0506449	5,8828	<0,0001	***
guada	-0,141957	0,0436949	-3,2488	0,0012	***
Среднее зав. перемен	1,794326	Ст. откл. зав. перемен	0,766095		
Сумма кв. остатков	2079,172	Ст. ошибка модели	0,651865		
R-квадрат	0,277016	Испр. R-квадрат	0,275982		
F(7, 4893)	89,04119	P-значение (F)	1,6e-122		
Лог. правдоподобие	-4852,988	Крит. Акаике	9721,977		
Крит. Шварца	9773,954	Крит. Хеннана-Куинна	9740,213		
Параметр rho	0,517254	Стат. Дарбина-Вотсона	0,751460		

Мы не стали включать в модель переменную *Nosch1*, чтобы избежать чистой мультиколлинеарности. Поэтому интерпретировать коэффициенты при всех остальных переменных уровня образования следует в терминах сравнения с работниками, которые не имеют образования вовсе.

Например, тот факт, что коэффициент при переменной *basic1* является положительным и статистически значимым, говорит о том, что работники, учившиеся в пределах шести лет, характеризуются более высоким уровнем дохода, чем работники без образования.

(б) Оценка модели со случайными эффектами дает следующие результаты (см. табл. 10.2).

Таблица 10.2

Модель 2: Случайные эффекты (GLS), использовано наблюдений - 4901					
Включено 1346 пространственных объектов					
Длина временного ряда: минимум 1, максимум 5					
Зависимая переменная: <i>lwage</i>					
Пропущены из-за совершенной коллинеарности: <i>monte</i>					
	Коэффициент	Ст. ошибка	t-статистика	P-значение	
const	1,39108	0,0884598	15,73	2,13e-054	***
<i>basic1</i>	0,225120	0,0865465	2,601	0,0093	***
<i>midd</i>	0,419046	0,0872226	4,804	1,60e-06	***
<i>higg</i>	1,22751	0,0907192	13,53	5,57e-041	***
<i>mexio</i>	-0,233051	0,0418386	-5,570	2,68e-08	***
<i>juare</i>	0,0408743	0,0500561	0,8166	0,4142	
<i>tijua</i>	0,302747	0,0520069	5,821	6,21e-09	***
<i>guada</i>	-0,152082	0,0488277	-3,115	0,0019	***
Среднее зав. перемен	1,794326	Ст. откл. зав. перемен	0,766095		
Сумма кв. остатков	2080,017	Ст. ошибка модели	0,651931		
Лог. правдоподобие	-4853,984	Крит. Акаике	9723,968		
Крит. Шварца	9775,946	Крит. Хеннана-Куинна	9742,204		
Межгрупповая дисперсия = 0,235124					
Внутригрупповая дисперсия = 0,219036					
$\text{meantheta} = 0,531081$					
$\text{corr}(y, \hat{y})^2 = 0,276918$					
Тест Бриша-Пэгана (Breusch-Pagan) -					
Нулевая гипотеза: Дисперсия специфических для наблюдений ошибок = 0					
Асимптотическая тестовая статистика: Хи-квадрат(1) = 1537,73					
p-значение = 0					

Результаты теста Бреуша—Пагана говорят в пользу модели со случайными эффектами, так как гипотеза о равенстве дисперсии случайных эффектов нулю уверенно отвергается (соответствующее *P*-значение существенно меньше одной сотой). Коэффициенты при переменных об-

разования являются значимыми и положительными, а также растут по абсолютной величине по мере роста уровня образования. Таким образом, можно заключить, что более высокий уровень образования ассоциируется с более высоким уровнем дохода.

(в) Оценка коэффициентов при переменных образования в модели с фиксированными эффектами невозможна, так как эти переменные являются неизменными во времени (и, следовательно, коллинеарны с переменными индивидуальных эффектов).

(г) Чтобы оценить различия в уровне заработной плате легальных и нелегальных рабочих с разным уровнем образования, добавим в модель переменную $form$, характеризующую работу в легальном секторе, а также переменные $fb = form * basic$, $fm = form * midd$, $fh = form * higg$.

Таблица 10.3

Модель 3: Случайные эффекты (GLS), использовано наблюдений - 4901					
Включено 1346 пространственных объектов					
Длина временного ряда: минимум 1, максимум 5					
Зависимая переменная: lwage					
Пропущены из-за совершенной коллинеарности: monte					
	Коэффициент	Ст. ошибка	t-статистика	P-значение	
const	1,50957	0,107843	14,00	1,11e-043	***
basic1	0,198251	0,107374	1,846	0,0649	*
midd	0,379939	0,109878	3,458	0,0005	***
higg	0,920306	0,125771	7,317	2,94e-013	***
mexio	-0,241970	0,0414232	-5,841	5,51e-09	***
juare	0,0239125	0,0495956	0,4821	0,6297	
tijua	0,291980	0,0515033	5,669	1,52e-08	***
guada	-0,161773	0,0484338	-3,340	0,0008	***
formm	-0,198182	0,112687	-1,759	0,0787	*
fb	0,0574282	0,117073	0,4905	0,6238	
fm	0,101627	0,118872	0,8549	0,3926	
fh	0,424301	0,133786	3,171	0,0015	***
Среднее зав. перемен	1,794326		Ст. откл. зав. перемен	0,766095	
Сумма кв. остатков	2050,551		Ст. ошибка модели	0,647561	
Лог. правдоподобие	-4819,021		Крит. Акаике	9662,043	
Крит. Шварца	9740,009		Крит. Хеннана-Куинна	9689,396	

Коэффициент при переменной fh оказался значимым и положительным. Это означает, что те работники с высшим образованием, которые заняты в легальном секторе, получают более высокую зарплату по сравнению с работниками с высшим образованием, занятыми в теневом секторе. Для всех работников с более низким уровнем образования не наблюдается различий в уровне доходов между легальным и нелегальным секторами, так как все соответствующие переменные не значимы при уровне значимости 5%.

ЗАДАЧА 11

ГОСУДАРСТВЕННЫЙ ДОЛГ И ЭКОНОМИЧЕСКИЙ РОСТ

Среди макроэкономистов хорошо известна работа С. М. Reinhart & К. S. Rogoff «Growth in a time of debt» (2010)¹, где на основе корреляции между темпами роста ВВП и уровнем долга был найден «пороговый уровень долга» 90% ВВП, при превышении которого темпы роста ВВП начинают значительно падать. Примечательно, что сами R&R сделали оговорку, что полученный результат является приближенным и неодинаков для разных групп стран. Несмотря на это, значение 90% ВВП получило широкое распространение, во многих эмпирических исследованиях тестируется и находит подтверждение нелинейная зависимость темпов роста ВВП от долга.

Однако в апреле 2013 г. Herndon, Ash, Pollin нашли ошибку в расчетах R&R и опровергли основной результат их исследования². На основании тех же самых данных делается вывод о том, что влияние государственного долга на темпы роста реального ВВП отрицательное и одинаково для любых значений долга (монотонная зависимость). Таким образом, был поставлен под сомнение один из аргументов в пользу необходимости «политики затягивания поясов», которым выступает исследование R&R. Но эта статья не отменяет результатов многочисленных исследований, нашедших нелинейную зависимость для разных групп стран.

В этом задании вам предлагается проверить наличие перевернутой U-образной (квадратичной) зависимости темпов роста ВВП от уровня государственного долга для 18 стран ОЭСР за 1980–2009 гг.

Данные содержаться в файле *debtOECD.xls* (описание данных см. также на листе Info).

Country — порядковый номер страны

Year — год

realGDP — реальный ВВП на душу населения, в постоянных ценах, долл. США

¹ Reinhart C. M., Rogoff K. S. Growth in a time of debt // NBER, Working Paper № 15639, 2010.

² Herndon T., Ash M., Pollin R. Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff // University of Massachusetts Amherst, Political Economy Research Institute, Working Paper series. — 2013. — № 322 [URL: http://www.peri.umass.edu/fileadmin/pdf/working_papers/working_papers_301-350/WP322.pdf].

- Growth* — темп прироста реального ВВП на душу населения, %
Pop — численность населения, тыс. чел.
Ngs — валовые национальные сбережения, % ВВП
School — среднее число лет обучения среди населения старше 15 лет
Openness — степень открытости экономики (экспорт + импорт к ВВП), % ВВП
Infl — уровень инфляции (ИПЦ), %
Total_dep — демографическая нагрузка детьми и пожилыми, %
Bankcrisis — фиктивная переменная, равна единице, если в данной стране в этот год произошел банковский кризис, и равна нулю в противном случае
Debtgov — валовой государственный долг, % ВВП

- (а) Оцените регрессию переменной *Growth* на *Ngs*, **логарифм** реального ВВП на душу населения, **темп прироста** населения, *Openness*, *School*, *Total_dep*, *Infl*, *Bankcrisis*, *Debtgov* и **квадрат** долга, используя три подхода:
1) обычный МНК (pooledOLS),
2) модель с фиксированными эффектами,
3) модель со случайными эффектами.
При этом, чтобы избежать потенциальной проблемы эндогенности, вместо самих регрессоров возьмите их первые лаги (периода T-1).

Представьте результаты в виде единой таблицы. Укажите коэффициенты и (в скобках под ними) их стандартные ошибки. Отметьте звездочками значимые переменные.

- (б) Выберите среди оцененных моделей наилучшую. Приведите результаты всех тестов, которые вы использовали для этого. Поясните, как именно на основе полученных результатов тестов осуществляется выбор.
- (в) На основе полученных **по выбранной модели** оценок рассчитайте «пороговое значение» уровня государственного долга (как вершину параболы), если соответствующие переменные оказались значимыми. С помощью теста на линейные ограничения на 5%-ном уровне значимости проверьте гипотезу о том, что «порог» равен 0,9 (90% ВВП)¹.
- (г) Для выбранной модели дайте содержательную интерпретацию коэффициента при переменной *Bankcrisis* (не забудьте, что она включена в модель с лагом).

¹ Вспомните, как выразить это соотношение как линейное. При необходимости воспользуйтесь справкой для диалогового окна теста на линейные ограничения.

Решение

(а) Во-первых, данные удобнее всего импортировать, используя индексные переменные *Year* и *Country*. Должно получиться 18 наблюдений за 31 период.

Во-вторых, создадим новые переменные:

1. *l_realGDP* — логарифм реального ВВП (в программе Gretl это можно сделать через «добавить»---«логарифмы выделенных переменных»);
2. *Pop_1*, т.е. первый лаг для *Pop* (поскольку без него Gretl откажется создавать темпы прироста);
3. *Popgrowth* — темп прироста населения (с помощью «добавить»---«добавить новую переменную» по формуле $popgrowth = (Pop/Pop_{1-1}) * 100$).

Примечание: если не умножить на 100, то из-за разных размерностей (слева — проценты, справа — доли) коэффициент при этой переменной в модели будет слишком большим по сравнению с остальными коэффициентами. Это не представляет проблемы, но разный порядок коэффициентов в модели может вызвать у неопытного читателя вопросы касательно интерпретации коэффициентов¹;

4. *Sq_Debtgov* — квадрат переменной *Debtgov* (создается через «добавить»---«квадраты выделенных переменных»).

Теперь можно переходить к оценке моделей:

1. Обычный МНК (pooledOLS), не забудем про робастные стандартные ошибки;
2. Модель с фиксированными эффектами (fixedeffects).
3. Модель со случайными эффектами (randomeffects).

Сводные результаты представлены в табл. 11.1.

Таблица 11.1

Результаты оценки трех моделей

	Обычная МНК-модель	Модель с фиксированными эффектами	Модель со случайными эффектами
Константа	47,39 *** (6,94)	80,29*** (6,04)	47,39*** (6,04)
Валовые национальные сбережения (лаг)	0,06* (0,037)	0,11** (0,05)	0,06*** (0,02)
Реальный ВВП на душу населения (логарифм, лаг)	-5,24*** (0,74)	-7,67*** (1,69)	-5,23*** (0,63)

¹ Хорошая заметка с подробными объяснениями, что «самый большой» коэффициент при переменной не делает ее «самой главной» в модели, есть в блоге Дэйва Джайлса, профессора Университета Виктории в Канаде [URL: <http://davegiles.blogspot.ru/2013/08/large-and-small-regression-coefficients.html> (дата обращения: 16.09.2016)].

Окончание табл. 11.1

	Обычная МНК-модель	Модель с фиксированными эффектами	Модель со случайными эффектами
Темп прироста населения (лаг)	-0,13 (0,34)	-0,56* (0,29)	-0,13 (0,22)
Открытость экономики (лаг)	-0,009** (0,004)	0,016 (0,014)	-0,009** (0,003)
Среднее число лет обучения среди населения старше 15 (лаг)	0,31*** (0,099)	-0,22 (0,39)	0,31*** (0,007)
Демографическая нагрузка (детьми и пожилыми, лаг)	0,07 (0,06)	-0,05 (0,067)	0,07** (0,03)
Инфляция (лаг)	-0,23*** (0,035)	-0,32*** (0,06)	-0,23*** (0,03)
Банковские кризисы (лаг)	-1,82*** (0,33)	-1,88*** (0,3)	-1,82*** (0,25)
Государственный долг (лаг)	5,1*** (1,654)	5,48** (2,33)	5,1*** (1,32)
Квадрат государственного долга (лаг)	-2,88*** (0,75)	-2,54*** (0,9)	-2,88*** (0,69)
Расчетное значение «порога» отношения госдолга к ВВП	88%	108%	89%
Тест на различие констант в группах (<i>p</i> -значение)	-	$2,3 \cdot 10^{-11}$ (ноль)	-
Тест Бреуша—Пагана (<i>p</i> -значение)	-	-	$2,25 \cdot 10^{-6}$ (ноль)
Тест Хаусмана (<i>p</i> -значение)	-	-	$5,9 \cdot 10^{-16}$ (ноль)
Число наблюдений	522	522	522

Источник: составлено авторами на основе расчетов, полученных в программе Gretl.

(б) Выберем среди оцененных моделей лучшую.

Тест «на различие констант в группах» сравнивает обычную МНК-модель и модель с фиксированными эффектами. По результатам этого теста (см. табл. 11.1) нулевая гипотеза об отсутствии фиксированных эффектов отвергается: *p*-значение около нуля, т.е. меньше любого разумного уровня значимости. Поэтому модель с фиксированными эффектами предпочтительнее, чем обычная МНК-модель.

Тест Бреуша—Пагана сравнивает обычную МНК-модель и модель со случайными эффектами. По его результатам нулевая гипотеза об отсутствии случайных эффектов отвергается (*p*-значение около нуля, т.е. меньше любого разумного уровня значимости), поэтому модель со случайными эффектами предпочтительнее, чем обычная МНК-модель.

Тест Хаусмана сравнивает модель с фиксированными эффектами и модель со случайными эффектами. По результатам этого теста нуле-

вая гипотеза о состоятельности ОМНК-оценок, получаемых в модели со случайными эффектами, отвергается, поскольку p -значение снова меньше любого разумного уровня значимости. Поэтому модель с фиксированными эффектами предпочтительнее, чем модель со случайными эффектами.

Таким образом, по результатам трех тестов выбираем модель с фиксированными эффектами, что совпадает и с содержательными соображениями: страны ОЭСР – уникальные объекты, а не случайные наблюдения из некой генеральной совокупности.

(в) На основе полученных **по выбранной модели** оценок рассчитаем «пороговое значение» уровня государственного долга (как вершину параболы).

Расчетное значение составляет приблизительно $5,48/(2*2,54) = 1,08$, т.е. «порог» государственного долга около 108% ВВП.

Гипотеза о равенстве порогового значения 90% проверяется с помощью теста на линейные ограничения. Гипотеза «вершина параболы равна 90%» может быть записана в виде: $-b/2a = 0,9$, а для теста на линейные ограничения нулевая гипотеза переписывается как $1,8*a + b = 0$. В данном тесте было получено p -значение 0,24, что больше чем 0,1, поэтому нулевая гипотеза принимается на 10%-ном уровне значимости. Можно сказать, что «пороговое значение» государственного долга значимо не отличается от 90% ВВП.

(г) Содержательная интерпретация коэффициента при переменной *Bankcrisis* (с лагом): если в какой-то стране ОЭСР случился банковский кризис, то при прочих равных в следующем году темп роста реального ВВП на душу населения в этой стране снизится на 1,88 процентного пункта.

ЛИТЕРАТУРА

1. *Вербик М.* Путеводитель по современной эконометрике / Пер. с англ. В. А. Банникова. Научн. ред. и предисл. С. А. Айвазяна. — М.: Научная книга, 2008.
2. *Доугерти К.* Введение в эконометрику: Учебник. 3-е изд. / Пер. с англ. — М.: ИНФРА-М, 2009.
3. *Кэмерон Э. Колин; Триведи Правин К.* Микроэконометрика: методы и их применение. — М.: Издательский дом «Дело», РАНХиГС, 2015.
4. *Магнус Я. Р., Катышев П. К., Пересецкий А. А.* Эконометрика. Начальный курс: Учеб. — 6-е изд., перераб. и доп. — М.: Дело, 2004.
5. *Сток Дж., Уотсон М.* Введение в эконометрику. — М.: Издательский дом «Дело», РАНХиГС, 2015.
6. *Greene W. H., William.* Econometric analysis. 5th Ed. — Prentice Hall, 2003.
7. Hayashi. Econometrics. Princeton University Press, 2000.
8. *Wooldridge J. M.* Econometric Analysis of Cross Section and Panel Data. — The MIT Press, 2010.

Научное электронное издание

Картаев Ф. С., Ключкова О. А., Ромашова В. М., Сучкова О. В.

**СБОРНИК ЗАДАЧ
ПО ЭКОНОМЕТРИКЕ ВРЕМЕННЫХ РЯДОВ
И ПАНЕЛЬНЫХ ДАННЫХ**

ISBN 978-5-906783-48-6



9 785906 783486