Методика построения моделей согласно методологии Бокса-Дженкинса (ARMA/ARIMA)

Откройте в пакете Gretl файл из раздела примеров Data4-2.gdt, выберете для анализа временной ряд С (real consumption expenditure (billions of 1992 dollars)), добавив переменную его первых разностей. Проведем предварительный анализ исследуемых временных рядов с помощью автокорреляционных функций и тестов «единичного корня».

Таблицы значений автокорреляционной и частной автокорреляционной функции временного ряда С ряда его первых разностей:

Автокорреляционная функция для Ct

Лаг	ACF		PACF		Q-стат.	[p-значение]
1	0,9181	***	0,9181	***	32,9487	[0,000]
2	0,8356	***	-0,0469		61,0437	[0,000]
3	0,7528	***	-0,0471		84,5345	[0,000]
4	0,6721	***	-0,0333		103,8450	[0,000]
5	0,5853	***	-0,0877		118,9607	[0,000]
6	0,4987	***	-0,0527		130,3032	[0,000]
7	0,4145	**	-0,0427		138,4074	[0,000]

Автокорреляционная функция для d_Ct

Лаг	ACF		PACF		Q-стат.	[p-значение]
1	0,3521	**	0,3521	**	4,7208	[0,030]
2	-0,0865		-0,2402		5,0143	[0,081]
3	-0,2933	*	-0,2071		8,4954	[0,037]
4	-0,1862		-0,0162		9,9443	[0,041]
5	-0,1000		-0,1075		10,3762	[0,065]
6	-0,0565		-0,0953		10,5188	[0,104]
7	0,0274		0,0217		10,5535	[0,159]

Первое выдвигаемое нами предположение заключается в том, что временной ряд С относится к нестационарным временным рядам и является интегрированным первого порядка, т.к. временной ряд первых разностей d_C стационарен. Приведем результаты тестирования обоих рядов с помощью теста ADF на наличие «единичного корня» (красным маркером выделены строки, указывающие на спецификацию теста и принимаемую гипотезу). Для ряда С в первых разностях возможно использовать обычный тест Дики-Фуллера (Gretl-ом так определяется расширенный тест Дики-Фуллера при l=0), поскольку на коррелограмме вторых разностей ряда отсутствует значимая корреляция.

Расширенный тест Дики-Фуллера (ADF-тест) для Ct
включая один лаг для (1-L)Ct
объем выборки 34
нулевая гипотеза единичного корня: a = 1

тест без константы
модель: (1-L)y = (a-1)*y(-1) + ... + e
коэф. автокорреляции 1-го порядка для e: 0,074
оценка для (a - 1): 0,0178152
тестовая статистика: tau_nc(1) = 3,13591
асимпт. p-значение 0,9996
Регрессия расширенного теста Дики-Фуллера
Зависимая переменная: d_Ct

	Коэффициент	Ст. ошибка	t-статистика	P-значение	
Ct_1	0,0178152	0,00568103	3,136	0,9996	
d_Ct_1	0,412709	0,166828	2,474	0,0189	**

тест с константой
модель: (1-L)y = b0 + (a-1)*y(-1) + ... + e
коэф. автокорреляции 1-го порядка для e: 0,068
оценка для (a - 1): 0,00570085
тестовая статистика: tau_c(1) = 0,587917
асимпт. p-значение 0,9895
Регрессия расширенного теста Дики-Фуллера
Зависимая переменная: d_Ct

	Коэффициент	Ст. ошибка	t-статистика	P-значение	
const	44,1691	28,9481	1,526	0,1372	
Ct_1	0,00570085	0,00969669	0,5879	0,9895	
d_Ct_1	0,337138	0,170808	1,974	0,0574	*

Тест Дики-Фуллера (DF-тест) для d_Ct
объем выборки 34
нулевая гипотеза единичного корня: a = 1

тест без константы
модель: (1-L)y = (a-1)*y(-1) + e
коэф. автокорреляции 1-го порядка для e: -0,084
оценка для (a - 1): -0,136855
тестовая статистика: tau_nc(1) = -1,43252
P-значение 0,1391
Регрессия теста Дики-Фуллера
Зависимая переменная: d_d_Ct

	Коэффициент	Ст. ошибка	t-статистика	P-значение
d_Ct_1	-0,136855	0,0955345	-1,433	0,1391

тест с константой
модель: (1-L)y = b0 + (a-1)*y(-1) + e
коэф. автокорреляции 1-го порядка для e: 0,059
оценка для (a - 1): -0,6392
тестовая статистика: tau_c(1) = -3,89061
P-значение 0,005284
Регрессия теста Дики-Фуллера
Зависимая переменная: d_d_Ct

	Коэффициент	Ст. ошибка	t-статистика	P-значение	
const	***58,1043***	***16,4478***	***3,533***	***0,0013***	*********
d_Ct_1	-0,639200	0,164293	-3,891	0,0053	***

с константой и трендом модель: (1-L)y = b0 + b1*t + (a-1)*y(-1) + ... + e коэф. автокорреляции 1-го порядка для e: 0,046 оценка для (a - 1): -0,268747 тестовая статистика: tau_ct(1) = -2,74803 ***асимпт. p-значение 0,2171*** Регрессия расширенного теста Дики-Фуллера Зависимая переменная: d_Ct Коэффициент Ст. ошибка t-статистика P-значение --- const 323,721 102,607 3,155 0,0036 *** Ct_1 -0,268747 0,0977966 -2,748 0,2171 d_Ct_1 0,436097 0,158344 2,754 0,0099 *** ***time 24,6892 8,76232 2,818 0,0085 ******	**с константой и трендом** модель: (1-L)y = b0 + b1*t + (a-1)*y(-1) + e коэф. автокорреляции 1-го порядка для e: 0,071 оценка для (a - 1): -0,668988 тестовая статистика: tau_ct(1) = -3,95579 P-значение 0,02023 Регрессия теста Дики-Фуллера Зависимая переменная: d_d_Ct Коэффициент Ст. ошибка t-статистика P-значение --- const 46,8979 21,4716 2,184 0,0366 ** d_Ct_1 -0,668988 0,169116 -3,956 0,0202 ** time 0,707052 0,864349 0,8180 0,4196

Согласно проведенному тестированию, исходный временной ряд действительно является интегрированным первого порядка и построение модели ARMA (ARIMA) будем проводить для временного ряда его первых разностей. Очевидно, согласно значениям автокорреляционных функций значения показателей p и q выбираются соответственно: p=1, q=3 (см. на коррелограмме максимальный порядок значимых лагов для ACF – выбор порядка MA, на PACF – выбор порядка AR). Строим соответствующую модель, используя в меню пакета GRETL: Модель->Временные ряды->Авторегрессия интегрированного скользящего среднего (ARIMA), где в поле «порядок AR и MA» отмечаем «or specific lags» и ставим выбранные нами значения. В данном случае в качестве зависимой переменной выступает временной ряд d_C, поэтому для него речь идет о построении модели ARMA (если в качестве зависимой переменной вы выбираете исходный ряд C, то строится модель ARIMA и вам потребует так же поставить значение «Разность», под которым подразумевается порядок интегрирования временного ряда»).

Модель 1: Метод оценки - ARMA, использовано наблюдений - 35 (1960-1994)
Зависимая переменная: d_Ct
Стандартные ошибки, рассчитанные на основе гессиана

	Коэффициент	*Ст. ошибка*	*z-статистика*	*P-значение*	
const	88,0233	7,34667	11,9814	<0,00001	***
phi_1	0,320919	0,163951	1,9574	0,05030	*
theta_3	-0,358717	0,206966	-1,7332	0,08306	*

Среднее зав. перемен	87,92857	Ст. откл. зав. перемен	50,70967
Mean of innovations	-0,922492	S.D. of innovations	44,32864
Лог. правдоподобие	-182,6190	Крит. Акаике	373,2379
Крит. Шварца	379,4593	Крит. Хеннана-Куинна	375,3855

		Real	*Imaginary*	*Modulus*	*Частота*
AR					
	Root 1	3,1161	0,0000	3,1161	0,0000
MA					
	Root 1	-0,7037	-1,2188	1,4074	-0,3333
	Root 2	-0,7037	1,2188	1,4074	0,3333
	Root 3	1,4074	0,0000	1,4074	0,0000

Оцените статистическую значимость построенной модели самостоятельно.

Для сравнения построим помимо уже имеющейся модели ARIMA(1,1,3) модель ARIMA(2,1,2):

Модель 2: Метод оценки - ARMA, использовано наблюдений - 35 (1960-1994)
Зависимая переменная: d_Ct
Стандартные ошибки, рассчитанные на основе гессиана

	Коэффициент	*Ст. ошибка*	*z-статистика*	*P-значение*	
const	88,9563	3,0202	29,4538	<0,00001	***
phi_2	0,639249	0,176961	3,6124	0,00030	***
theta_2	-1	0,143894	-6,9496	<0,00001	***

Среднее зав. перемен	87,92857	Ст. откл. зав. перемен	50,70967
Mean of innovations	-6,845285	S.D. of innovations	45,54104
Лог. правдоподобие	-184,8933	Крит. Акаике	377,7866
Крит. Шварца	384,0080	Крит. Хеннана-Куинна	379,9342

	Real	*Imaginary*	*Modulus*	*Частота*
AR				
Root 1	-1,2507	0,0000	1,2507	0,5000
Root 2	1,2507	0,0000	1,2507	0,0000
MA				
Root 1	-1,0000	0,0000	1,0000	0,5000
Root 2	1,0000	0,0000	1,0000	0,0000

Во второй модели коэффициенты даже более статистически значимы (т.е. коэффициенты значимы согласно значениям Р-вероятностей для меньшего уровня значимости), однако в пользу первой модели свидетельствуют значения информационных критериев Акайке, Шварца, Хеннана-Куинна. Приведем так же в таблице значения автокорреляционных функций для случайных отклонений сравниваемых моделей, из которых следует, что вторая модель не является адекватной с точки зрения методологии Бокса-Дженкинса, поскольку значения функций для первого лага статистически значимы и случайные отклонения не являются «белым шумом», они коррелированы.

Функция автокорреляции ошибок ARIMA(1,1,3)

Лаг	ACF	PACF	Q-стат. [p-значение]
1	0,0509	0,0509	0,0988 [0,753]
2	-0,1384	-0,1414	0,8508 [0,654]
3	0,0262	0,0424	0,8786 [0,831]
4	-0,0450	-0,0704	0,9630 [0,915]
5	-0,0752	-0,0598	1,2069 [0,944]
6	-0,0429	-0,0538	1,2890 [0,972]
7	0,0702	0,0620	1,5171 [0,982]

Функция автокорреляции ошибок ARIMA(2,1,2)

Лаг	ACF	PACF	Q-стат. [p-значение]
1	0,3460 **	0,3460 **	4,5590 [0,033]
2	0,0532	-0,0755	4,6702 [0,097]
3	-0,2153	-0,2387	6,5454 [0,088]
4	-0,0881	0,0849	6,8693 [0,143]
5	-0,0259	-0,0065	6,8982 [0,228]
6	-0,0071	-0,0692	6,9005 [0,330]
7	0,0690	0,1086	7,1204 [0,416]

Примечание по обозначениям компонент AR и MA в пакете Gretl:

Generating an ARMA(1,1)

Problem: Generate $y_t = 0.9y_{t-1} + \varepsilon_t - 0.5\varepsilon_{t-1}$, with $\varepsilon_t \sim NIID(0, 1)$.

Solution:

```
alpha = 0.9
theta = -0.5
series e = normal()
series y = 0
series y = alpha * y(-1) + e + theta * e(-1)
```