Эконометрическое моделирование

Лабораторная работа № 11

Модели стационарных временных рядов

Оглавление

Одномерные временные модели

При анализе пространственных данных, как правило, предпринимаются попытки объяснения изменений одного фактора в зависимости от изменений значений каких-то других факторов. При анализе временных рядов встречаются модели, базирующиеся на другом подходе. Такие модели называются одномерными временным моделями. В моделях этого класса предпринимается попытка смоделировать и спрогнозировать значения некоторых временных параметров, опираясь исключительно на информацию о прошлых значениях исследуемого параметра.

Обычно при использовании подобных моделей не строятся какие-то теоретические конструкции, объясняющие изменения исследуемого ряда. Априори полагается, что используемые данные уже содержат в себе такие модели.

Одна из причин использования временных моделей заключается в том, что они могут быть особенно полезны, когда невозможно найти объясняющие факторы, измеряемые с той же частотой, что и исследуемые переменные. Например, если исследуются дневные биржевые доходности, то в качестве объясняющих переменных могли бы выступать некоторые макроэкономические величины, которые измеряются не чаще, чем раз в месяц.

При работе с временными моделями необходимо знать, являются ли рассматриваемые ряды стационарными. Этот вопрос важен потому, что стационарные и нестационарные ряды обладают разными статистическими характеристиками, поэтому должны оцениваться разными способами.

Стационарные временные ряды

На общем понятийном уровне стационарный временной ряд представляется как ряд, который имеет постоянную среднюю. А значения ряда колеблются вокруг этой средней.

Свойства стационарных рядов не меняются при изменении начала отсчета времени.

Временной ряд $\{x_t\}$ называют стационарным (в широком смысле0, если он обладает следующими свойствами:

1. Среднее значение уровней ряда не изменяется во времени, т.е. математическое ожидание постоянно $Mx_t = const$.
2. Автоковариация (т.е. ковариация между уровнями одного и того же ряда) зависит только от того, насколько сильно они удалены друг от друга во времени, и не зависит от того, находятся ли они в начале или в конце временного ряда, т.е. $cov(x_t, x_{t+h}) = \Upsilon(h)$. Величина h, характеризующая разницу во времени между элементами временного ряда, называется *лагом* или *запаздыванием*.
3. Так как $\Upsilon(0) = cov(x_t, x_t)$, то дисперсия стационарного временного ряда также не меняется со временем.

Коэффициенты корреляции между разными элементами стационарного временного ряда с лагом h называются коэффициентами автокорреляции порядка h.

Функция ρ(h) = corr(x_t, x_{t+h}) называется *автокорреляционной функцией* (AFC) стационарного временного ряда. Для коэффициента автокорреляции справедливо также следующее: corr(x_t, x_s) = ρ(s-t).

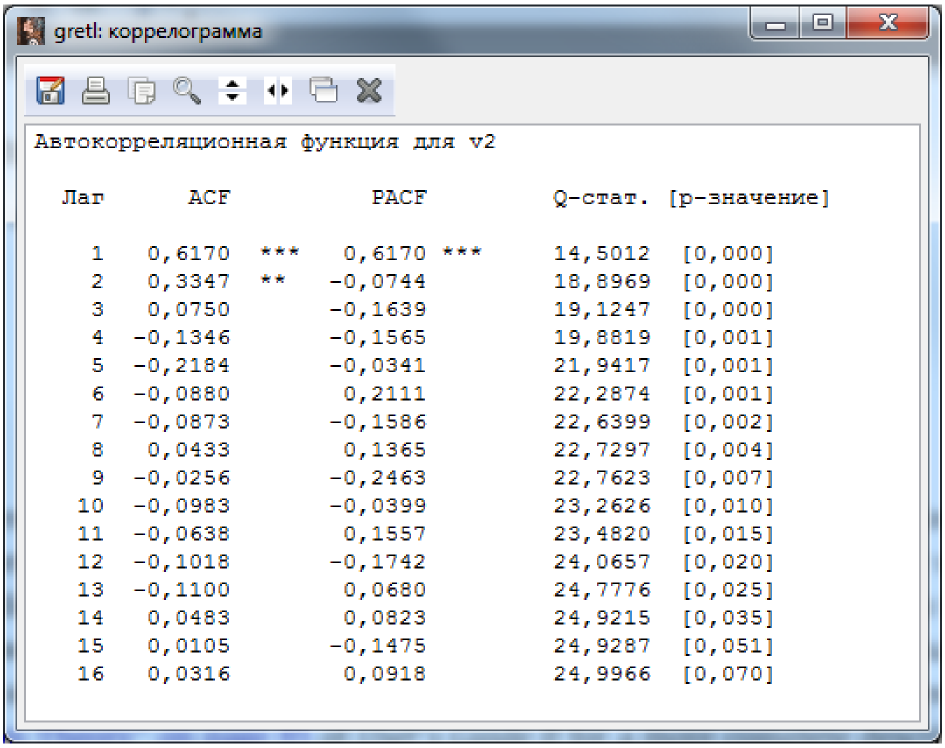
Коэффициенты автокорреляции вычисляется по следующей формуле:

$$ACorr_h = \frac{\sum_{t=h+1}^{T}((X_t - \bar{X})(X_{t-h} - \overline{X_{t-h}}))/(T-h)}{\sum_{t=1}^{T}(X_t - \bar{X})^2 /T}$$

В эконометрических пакетах (Gretl, Eviews) используется несколько упрощенная формула для расчетов коэффициентов автокорреляции:

$$ACorr_h = \frac{\sum_{t=h+1}^{T}(X_t - \bar{X})(X_{t-h} - \bar{X})}{\sum_{t=1}^{T}(X_t - \bar{X})^2}$$

Рассмотрим коэффициенты автокорреляции, полученные для варианта 0 с помощью эконометрического пакета Gretl.



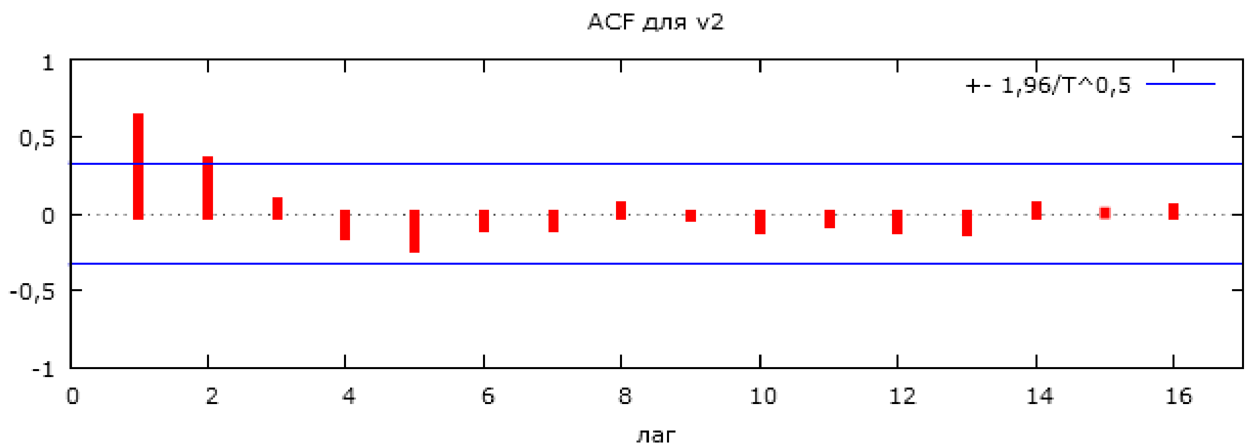
gretl: коррелограмма

Автокорреляционная функция для v2

Лаг	ACF		PACF		Q-стат.	[p-значение]
1	0,6170	***	0,6170	***	14,5012	[0,000]
2	0,3347	**	-0,0744		18,8969	[0,000]
3	0,0750		-0,1639		19,1247	[0,000]
4	-0,1346		-0,1565		19,8819	[0,001]
5	-0,2184		-0,0341		21,9417	[0,001]
6	-0,0880		0,2111		22,2874	[0,001]
7	-0,0873		-0,1586		22,6399	[0,002]
8	0,0433		0,1365		22,7297	[0,004]
9	-0,0256		-0,2463		22,7623	[0,007]
10	-0,0983		-0,0399		23,2626	[0,010]
11	-0,0638		0,1557		23,4820	[0,015]
12	-0,1018		-0,1742		24,0657	[0,020]
13	-0,1100		0,0680		24,7776	[0,025]
14	0,0483		0,0823		24,9215	[0,035]
15	0,0105		-0,1475		24,9287	[0,051]
16	0,0316		0,0918		24,9966	[0,070]

График функции ρ(h) называется *коррелограммой* временного ряда.

Эконометрический пакет также позволяет построить и график автокорреляционной функции.



Частная автокорреляционная функция

Частная автокорреляционная функция $(PACF)_{hh}$ определяет корреляцию между текущим и произошедшим h периодов назад наблюдением после удаления косвенного влияния (x_{t-h+1}, x_{t-h+2}, … x_{t-1})-х наблюдений. То есть $PACorr_4$ измеряет корреляцию между x_t и x_{t-4} без учета влияния x_{t-1}, x_{t-2}, x_{t-3} –го лага.

Для первого лага значения автокорреляционной функции и частной автокорреляционной функции совпадают, так как отсутствует влияние промежуточных лагов. Для второго лага частная автокорреляционная функция рассчитывается по формуле:

$$PACorr_2 = \frac{ACorr_2 - ACorr_1^2}{1 - ACorr_1^2}$$

Задание 1. Расчет автокорреляционной функции

Для своего варианта (данные приведены в конце лабораторной работы) с помощью MS Excel рассчитайте автокорреляционную функцию для 5 лагов и постройте график автокорреляционной функции. Рассчитайте частную автокорреляционную функцию второго порядка.

Проделайте то же самое в Gretl и сравните полученные результаты.

Модель авторегрессии

Рассмотрим одну из основных моделей, используемых для анализа стационарных временных рядов, - модель авторегрессии.

Модель авторегрессии – это модель временного ряда, в которой текущее значение моделированной переменной задается функцией от прошлых значений самой переменной.

Модель вида

$$x_t = \mu + a_1 x_{t-1} + \dots + a_p x_{t-p} + \varepsilon_t$$

называется моделью авторегрессии *AR(p)* порядка *p*. Параметры a_i – называются коэффициентами авторегрессии, ε_t – *«белый шум».*

Согласно этой модели, значения временного ряда x_t складывается из суммы прошлых *p* значений временного ряда и величины ε_t, отвечающей за влияние внешних факторов в момент времени t.

Модель авторегрессии может описывать как стационарный процесс, так и нестационарный. Существуют определенные условия, накладываемые на параметры модели, при выполнении которых процесс, описываемый моделью авторегрессии, будет стационарным.

Чтобы определить эти условия, рассмотрим понятия лагового оператора и характеристического уравнения.

Лаговый оператор определяется соотношением:

$$L^h x_t = x_{t-h}$$

Модель авторегрессии

$$x_t = \mu + a_1 x_{t-1} + a_2 x_{t-2} + \varepsilon_t$$

с помощью лагового оператора можно выразить следующим образом:

$$x_t = \mu + a_1 + Lx_t + a_2 L^2 x_t + \varepsilon_t$$

или

$$(1 - a_1 L - a_2 L^2) x_t = \mu + \varepsilon_t.$$

Для модели авторегрессии AR(p) можно определить так называемое «характеристическое» уравнение. Оно называется «характеристическим», потому что его корни определяют характеристики временного ряда x_t.

Характеристическое уравнение в данном случае будет иметь вид

$$1 - a_1 x - a_2 z^2 = 0$$

В общем виде

$$1 - a_1 z - \cdots - a_p z^p = 0$$

где p – порядок модели регрессии.

Временной ряд, заданный моделью авторегрессии AR(p), является стационарным тогда и только тогда, когда все корни характеристического уравнения по модулю больше единицы.

Пример

Рассмотрим модель авторегрессии первого порядка AR(1), задаваемую уравнением

$$x_t = \mu + \alpha x_{t-1} + \varepsilon_t$$

Преобразуем модель с помощью лагового оператора

$$x_t = \mu + \alpha L x_t + \varepsilon_t$$

или

$$(1 - \alpha L) x_t = \mu + \varepsilon_t$$

Для модели характеристическое уравнение (1 – αz)=0 имеет единственный корень $z = \frac{1}{\alpha}$, и, следовательно, она определяет стационарный временной ряд только в случае $|z| \geq 1 \leftrightarrow |\alpha| < 1.$

При α = 1 модель определяет нестационарный временной ряд

$$x_t = x_{t-1} + \varepsilon_t$$

называемый *«случайным блужданием».*

Для любого стационарного авторегрессионного процесса автокорреляционная функция будет уменьшаться по экспоненте.

Задание 2. Базовые модели временных рядов

Рассмотрите самостоятельно две базовые модели временных рядов: белый шум и случайное блуждание: опишите, какой процесс демонстрируют данные модели, какие параметры используются в этих моделях, как выглядят процессы, описываемый данными моделями на графиках.

Пример

Рассмотрим модель авторегрессии AR(2), задаваемую уравнением

$$x_t = 3 + 2x_{t-1} - x_{t-2} + \varepsilon_t$$

Для нее характеристическое уравнение имеет вид $1 - 2z - z^2 = 0$ и имеет единственный корень z = 1. Следовательно, уравнение авторегрессии задает нестационарный временной ряд.

Задание 3. Стационарные временные ряды

Является ли временной ряд, заданный авторегрессионным разностным уравнением, стационарным? Задание выполняется полностью (без вариантов).

1. $x_t = 7 + 0{,}5x_{t-1} + \varepsilon_t$
2. $x_t = 10 + 0{,}25x_{t-2} + \varepsilon_t$
3. $x_t = \frac{3}{2}x_t - \frac{3}{4}x_{t-2} + \frac{1}{8}x_{t-3} + \varepsilon_t$
4. $x_t = 3 + 0{,}54 - 0{,}04x_{t-2} + \varepsilon_t$
5. $x_t = 5 - 3x_{t-1} - 3x_{t-2} - x_{t-3} + \varepsilon_t$
6. $x_t = -2x_{t-1} - 1{,}25x_{t-2} - 0{,}25x_{t-3} + \varepsilon_t$

Эконометрические методы исследования стационарных временных рядов

Можно выделить три основных этапа исследования авторегрессионных моделей стационарных временных рядов:

1. Идентификация модели: оценивание порядка модели.

2. Оценивание модели: получение оценок параметров модели.

3. Диагностика модели: проверка адекватности и соответствия основным требованиям регрессионного анализа.

Идентификация модели

Рассмотрим один из подходов к оценке порядка регрессии, который основа на идее информационных критериев.

Процедура применения информационных критериев состоит в следующем: оценивается несколько моделей авторегрессии разных порядков и для каждой модели считается «количество информации», содержащееся в этой модели. Выбирается та модель, для которой «количество информации» максимально или минимально в зависимости от используемого критерия.

Информационный критерий Акаике: порядок p выбирается из условия

$$AIC(m) = lns^2(m) + \frac{2m}{n} \to min$$

Здесь $s^2(m)$ – оценка дисперсии остатков в модели AR(m). Исторически это первый информационный критерий. Его основной недостаток – при больших объемах выборочных данных он завышает порядок авторегрессии.

Информационный критерий Шварца: порядок p выбирается из условия:

$$SIC(m) = lns^2(m) + \frac{mlnn}{n} \to min$$

Пример

По временному ряду длины n=100 были оценены авторегрессионные модели до четвертого порядка и для них получены следующие значения дисперсии остатков: $s^2(1)=0{,}9$, $s^2(2)=0{,}7$, $s^2(3)=0{,}5$ и $s^2(4)=0{,}46$. Выберем порядок модели авторегрессии с помощью информационных критериев.

1. $SIC(1) = ln0{,}9 + \frac{ln100}{100} \approx -0{,}059.$
2. $SIC(2) = ln0{,}7 + \frac{2ln100}{100} \approx -0{,}265.$
3. $SIC(3) = ln0{,}5 + \frac{3ln100}{100} \approx -0{,}555.$
4. $SIC(4) = ln0{,}46 + \frac{4ln100}{100} \approx -0{,}592.$

По критерию Шварца следует выбрать для прогнозирования модель четвертого порядка, так как значение критерия для нее минимально.

1. $AIC(1) = ln0{,}9 + \frac{2}{100} \approx -0{,}085$
2. $AIC(2) = ln0{,}7 + \frac{4}{100} \approx -0{,}317$
1. $AIC(3) = ln0{,}5 + \frac{6}{100} \approx -0{,}633$
1. $AIC(4) = ln0{,}46 + \frac{8}{100} \approx -0{,}697$

По критерию Акаике также следует выбрать для прогнозирования модель четвертого порядка, так как значение критерия для нее минимально.

После выбора порядка авторегрессии проверяется значимость коэффициентов и модель можно улучшать, последовательно исключая незначимые коэффициенты.

Проверка адекватности модели

Проверка адекватности, т.е. проверка согласованности выбранной и оцененной модели с наблюдениями, как и в регрессионном анализе основана на исследовании остатков.

В данном случае, для исследования остатков на автокорреляцию не подходит тест Дарбина-Уотсона. Вместо него используется Q-статистика.

Q-Statistics

Последние два столбца, представленные в расчетах автокорреляционной функции в Gretl, содержат предложенные Боксов и Льюингом (Box, Ljung) значения Q-статистик и соответствующие им p-значения. Q-статистика для лага h является тестом для нулевой гипотезы: не существует автокорреляции до порядка h в модели AR(p). Она вычисляется по следующей формуле:

$$Q = n(n+2)\sum_{j=1}^{h}\frac{ACorr_j^2}{n-j}$$

где n – длина временного ряда.

Величина Q распределена как χ^2 с (h-p) степенями свободы. Если Q> χ^2, то нулевая гипотеза отвергается.

Пример

Для временно ряда длины n=100 была оценена модель второго порядка AR(2) и вычислены коэффициенты автокорреляции остатков.

$ACorr_1 = 0{,}001$; $ACorr_2 = 0{,}001$; $ACorr_3 = 0{,}0006$; $ACorr_4 = 0{,}0004$; $ACorr_5 = 0{,}0003$.

Вычислим Q-статистику:

$$Q = 100(100+2)\left(\frac{0{,}001^2}{100-1} + \frac{0{,}001^2}{100-2} + \frac{0{,}0006^2}{100-3} + \frac{0{,}0004^2}{100-4} + \frac{0{,}0003^2}{100-5}\right) \approx 0{,}00027$$

Критическое значение распределения χ^2 с (h–2) = 5-2 =3 степенями свободы при уровне значимости α=0,05 составляет 7,8.

Так как Q< $\chi^2_{критич}$, то остатки в модели удовлетворяют условию отсутствия автокорреляции.

Таблица – Критические точки распределения Хи-квадрат

Число степеней свободы	Уровень значимости α					
	0,01	0,05	0,1	0,90	0,95	0,99
1	6,6	3,8	2,71	0,02	0,004	0,0002
2	9,2	6,0	4,61	0,21	0,1	0,02
3	11,3	7,8	6,25	0,58	0,35	0,12
4	13,3	9,5	7,78	1,06	0,71	0,30
5	15,1	11,1	9,24	1,61	1,15	0,55
6	16,8	12,6	10,6	2,20	1,64	0,87
7	18,5	14,1	12,0	2,83	2,17	1,24
8	20,1	15,5	13,4	3,49	2,73	1,65
9	21,7	16,9	14,7	4,17	3,33	2,09
10	23,2	18,3	16,0	4,87	3,94	2,56
11	24,7	19,7	17,3	5,58	4,57	3,05
12	26,2	21,0	18,5	6,30	5,23	3,57
13	27,7	22,4	19,8	7,04	5,89	4,11
14	29,1	23,7	21,1	7,79	6,57	4,66
15	30,6	25,0	22,3	8,5	7,26	5,23
16	32,0	26,3	23,5	9,31	7,98	5,81
17	33,4	27,6	24,8	10,1	8,67	6,41
18	34,8	28,9	26,0	10,9	9,39	7,01
19	36,2	30,1	27,2	11,7	10,1	7,63
20	37,6	31,4	28,4	12,4	10,9	8,26
21	38,9	32,7	29,6	13,2	11,6	8,90
22	40,3	33,9	30,6	14,0	12,63	9,54
23	41,6	35,2	32,0	14,8	13,1	10,2
24	43,0	36,4	33,2	15,7	13,8	10,9
25	44,3	37,7	34,4	16,5	14,6	11,5
26	45,6	38,9	35,6	17,3	15,4	12,2
27	47,0	40,1	36,7	18,1	16,2	12,9
28	48,3	41,3	37,9	18,9	16,9	13,6
29	49,6	42,6	39,1	19,8	17,7	14,3
30	50,9	43,8	40,3	20,6	18,5	15,0

Задание 4. Построение авторегрессионной модели

Для временного ряда, соответствующего вашему варианту, оцените модели AR(1), AR(2) и AR(3). Используя информационные критерии, выберете одну из этих моделей и проверьте выполняется ли условие отсутствия автокорреляции остатков.

t	Вар 0	Вар 1	Вар 2	Вар 3	Вар 4	Вар 5
1	-1,98	0,67	4,22	5,84	1,88	19,05
2	0,75	2,57	4,75	5,92	0,78	20,11
3	1,41	-0,70	8,02	4,64	2,87	22,42
4	1,25	4,71	5,27	4,77	-0,26	18,80
5	1,27	-1,79	3,61	4,64	2,50	20,64
6	1,7	4,34	2,65	6,38	0,25	18,36
7	0,48	-0,57	4,99	6,71	0,02	20,72
8	0,2	2,71	7,67	7,47	3,59	21,12
9	-0,28	1,70	6,08	7,58	-0,52	19,90
10	0,25	2,84	3,81	6,47	2,47	19,84
11	2,51	0,47	3,78	5,42	1,40	19,96
12	2,37	1,78	4,63	4,70	-0,81	19,78
13	3,65	2,02	6,45	5,08	5,07	21,11
14	2,81	1,69	5,97	4,23	-4,03	20,06
15	0,51	1,60	6,09	4,03	4,60	21,64
16	1,44	1,18	9,07	2,17	-2,30	21,27
17	0,7	0,17	5,96	3,61	1,54	21,84
18	-0,73	2,39	5,17	5,02	3,52	21,19
19	-1,14	2,34	4,41	5,79	-1,10	20,97
20	0,38	1,45	4,15	5,65	2,60	20,54
21	1,5	3,48	3,48	5,91	2,57	19,47
22	3,33	2,05	4,19	6,56	-1,24	20,02
23	2,15	-0,16	5,51	6,31	4,39	20,19
24	3,95	2,99	4,30	5,99	-1,08	18,67
25	3,1	0,48	6,03	3,93	4,01	21,48
26	0,68	2,34	4,19	2,96	1,11	17,94
27	-0,32	-0,81	3,66	2,98	0,83	19,75
28	1,41	4,44	3,34	1,70	1,85	17,89
29	-0,58	-2,16	4,65	3,26	0,18	19,36
30	1,00	4,77	6,21	6,56	1,38	19,55
31	0,5	0,04	6,37	8,26	0,34	19,88
32	-1,23	4,03	4,84	9,21	-0,21	19,19
33	-3,08	3,07	3,60	7,08	3,05	18,98
34	-2,61	0,41	4,11	4,94	-0,80	19,26
35	-2,89	0,41	7,36	2,90	1,71	21,60