ВЗАИМОСВЯЗАННЫЕ ВРЕМЕННЫЕ РЯДЫ

1. Понятие взаимосвязанных временных рядов.
2. Методы исключения тенденции: метод отклонений от тренда, метод последовательных разностей, включение в модель фактора времени.
3. Автокорреляция в остатках. Критерий Дарбина-Уотсона.
4. Оценивание уравнения регрессии при автокорреляции в остатках.
5. Коинтеграция временных рядов.

Модели на основе рядов динамики

- Модели изолированного динамического ряда
- Модели взаимосвязанных рядов динамики
- Модели автрегрессии
- Модели с распределенным лагом

Многомерные модели временных рядов

Ранее мы рассматривали модели для единственного временного ряда.

Теперь мы будем анализировать модели, включающие несколько рядов.

Мотивация:

- Такой подход может улучшить качество прогнозов;
- Такой подход позволяет отвечать на вопросы о динамических причинно-следственных связях.

Взаимосвязанные временные ряды

При изучении развития явления во времени иногда возникает необходимость *оценить степень взаимосвязи* в изменениях уровней двух и более временных рядов различного содержания, но связанных один с другим.

Связанные временные ряды – это временные ряды, показывающие зависимость результативного признака от одного или нескольких факторных признаков.

Корреляция уровней временных рядов

Степень тесноты связи между уровнями временных рядов рядов y_t и x_t можно оценить с помощью линейного коэффициенты корреляции Пирсона.

Однако в уровнях временных рядов y_t и x_t может присутствовать систематическая составляющая, они могут оказаться нестационарными.

Модели нестационарных временных рядов

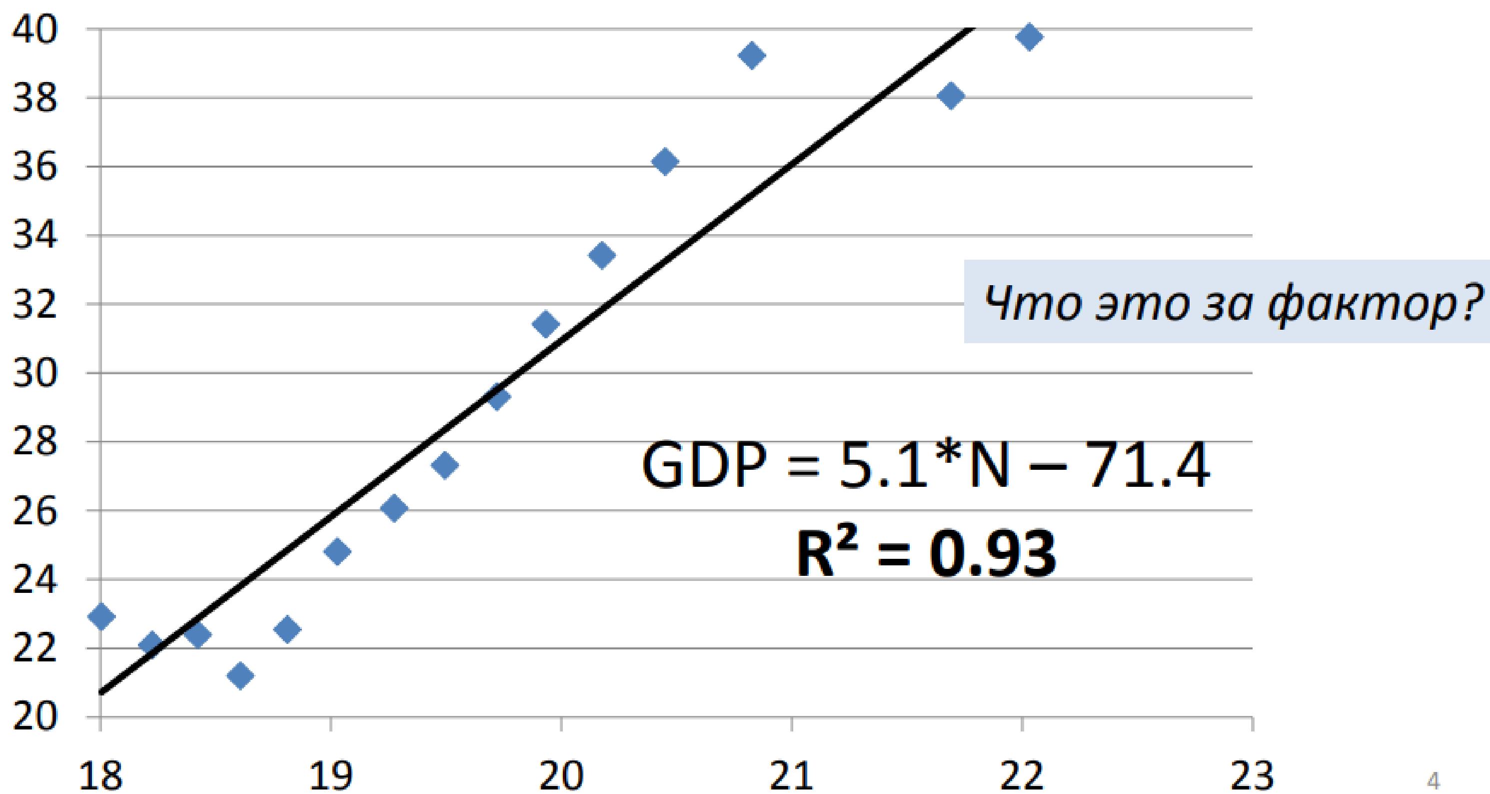
Чем опасно строить регрессии, включающие нестационарные временные ряды?

Риском возникновения ложной регрессии.

Ложная регрессия — ситуация, когда между регрессором и зависимой переменной в действительности нет никакой связи, однако уравнение регрессии характеризуется хорошей значимостью и высоким значением R^2.

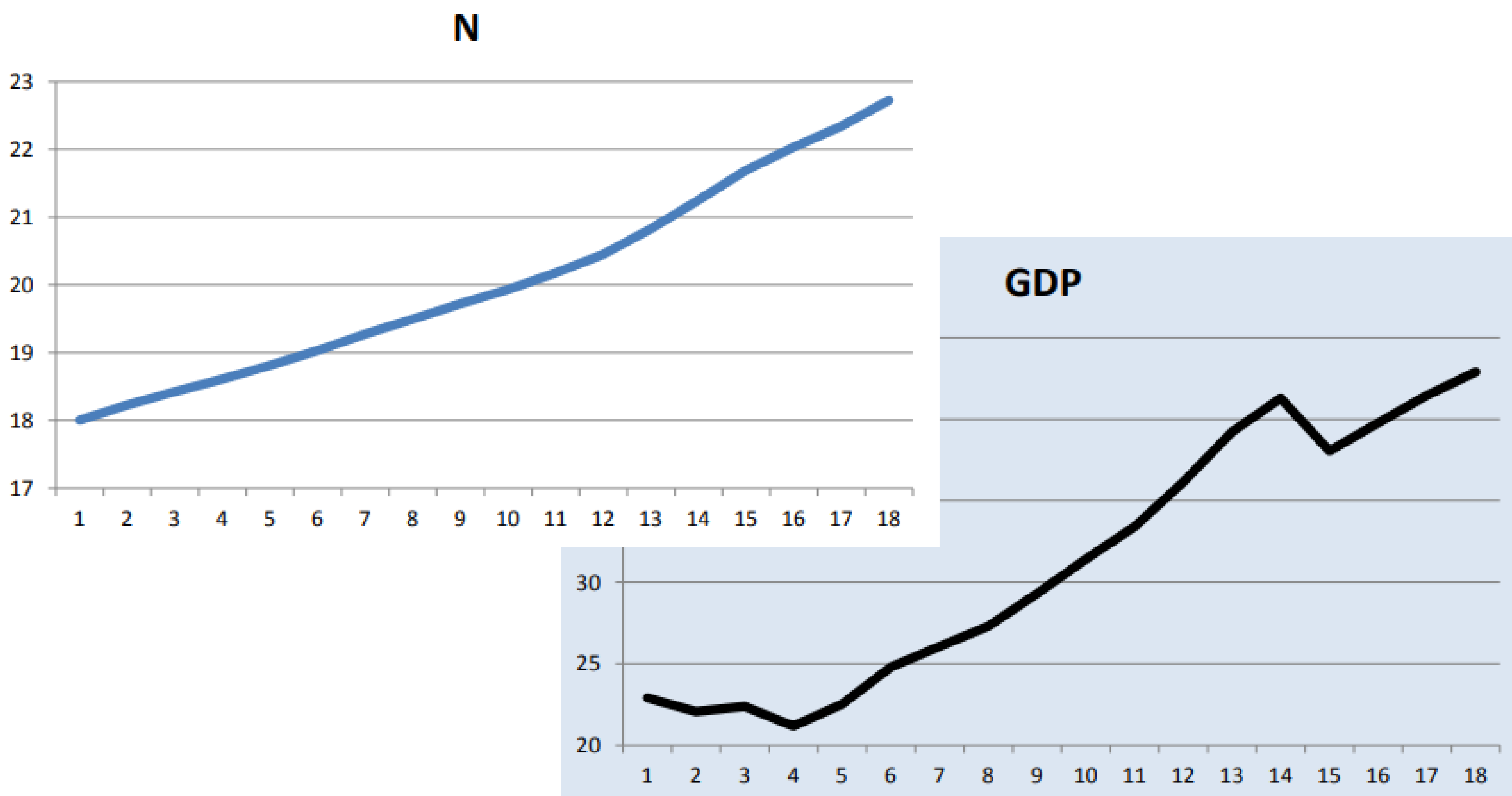
Ложная регрессия

Зависимость ВВП России от некоторого фактора в 1995–2012 гг. (реальный ВВП в ценах 2008 г., трлн руб.)



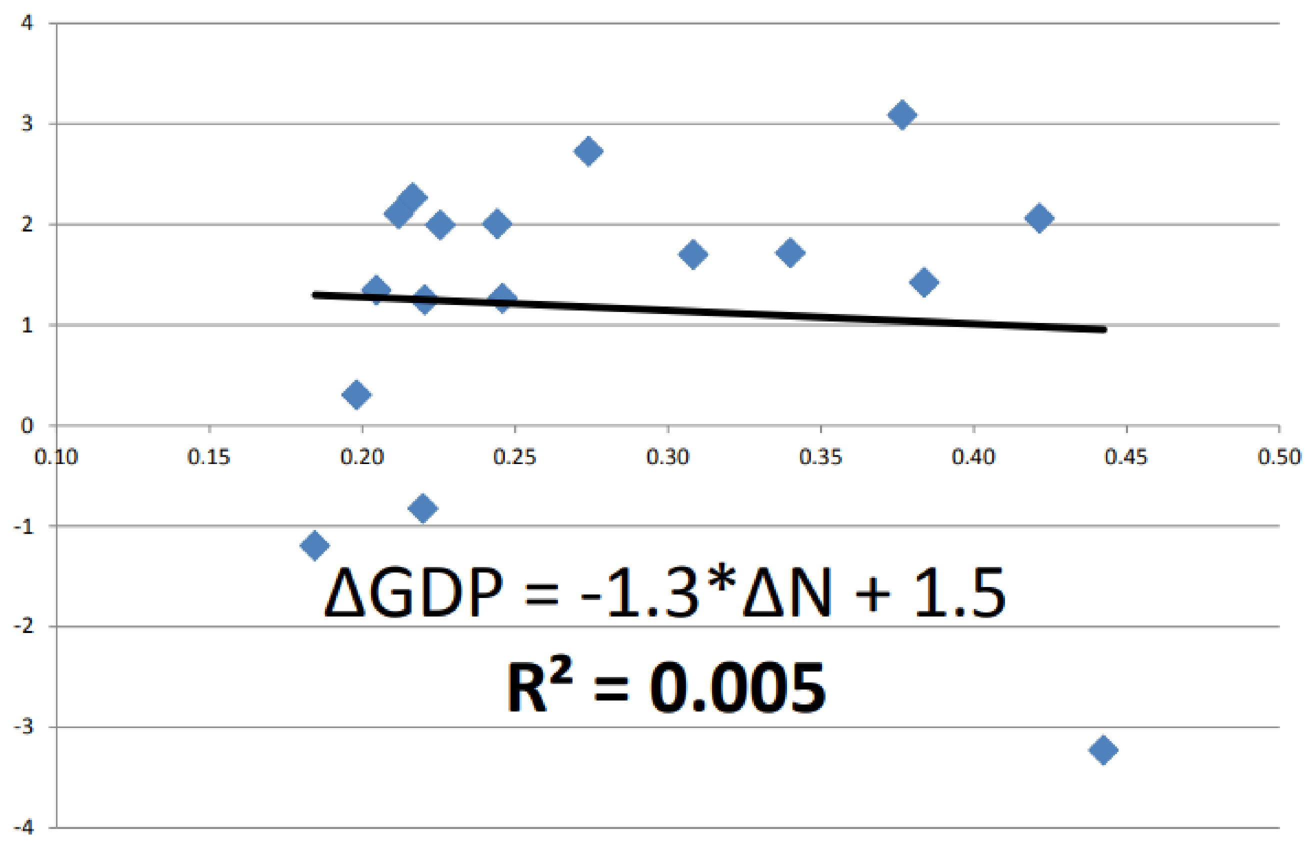
Ложная регрессия

Все дело в нестационарности



Ложная регрессия

Если оценивать регрессию для первых разностей, то связь становится незначимой:



Мнимая регрессия

Если рассматриваемые временные ряды y_t и x_t содержат тенденцию, то коэффициент корреляции, характеризующий степень зависимости между y_t и x_t будет иметь высокое значение. Такая же ситуация будет иметь место тогда, когда y_t и x_t зависят от переменной времени t.

Как в первом, так и во втором случае имеет место ложная корреляция, которая приводит при построении регрессии y_t на x_t к автокорреляции в остатках и нестационарности ряда остатков регрессии (ложная регрессия), то есть к нарушению предпосылок МНК.

Методы исключения тренда

Для получения регрессии со стационарным временным рядом остатков ε_t может быть использован метод последовательных разностей, когда переход к некоторым *k*-м разностям уровней ряда позволяет получить стационарный ряд остатков.

Другими методами исключения тренда являются методы включения фактора времени и отклонений от тренда.

Метод отклонения уровней ряда от основной тенденции (метод отклонений от тренда)

Если каждый из рядов y_t и x_t содержит тренд, то аналитическим выравниванием по каждому из рядов можно найти параметры тренда и определить расчетные по тренду уровни рядов $\widehat{y_t}$ и $\widehat{x_t}$.

Влияние тенденции можно устранить путем вычитания расчетных значений тренда из фактических.

Дальнейший регрессионный анализ проводят с отклонениями от тренда $y_t - \widehat{y_t}$ и $x_t - \widehat{x_t}$.

Метод отклонений от тренда

$$e_{y_t} = y_t - \hat{y}_t$$

$$e_{x_t} = x_t - \hat{x}_t$$

$$e_{y_t} = a + b \cdot e_{x_t}$$

Метод отклонений от тренда

$$(y_t - \hat{y}_t) = b(x_t - \hat{x}_t)$$

$$y_t = \hat{y}_t + b(x_t - \hat{x}_t)$$

$$y_p = \hat{y}_{t=p} + b(x_p - \hat{x}_{t=p})$$

Мет од от клонений от т ренда

y_p - прогнозное значение y_t

$\hat{y}_{t=p}$ - прогноз *у* по тренду при *t=p*

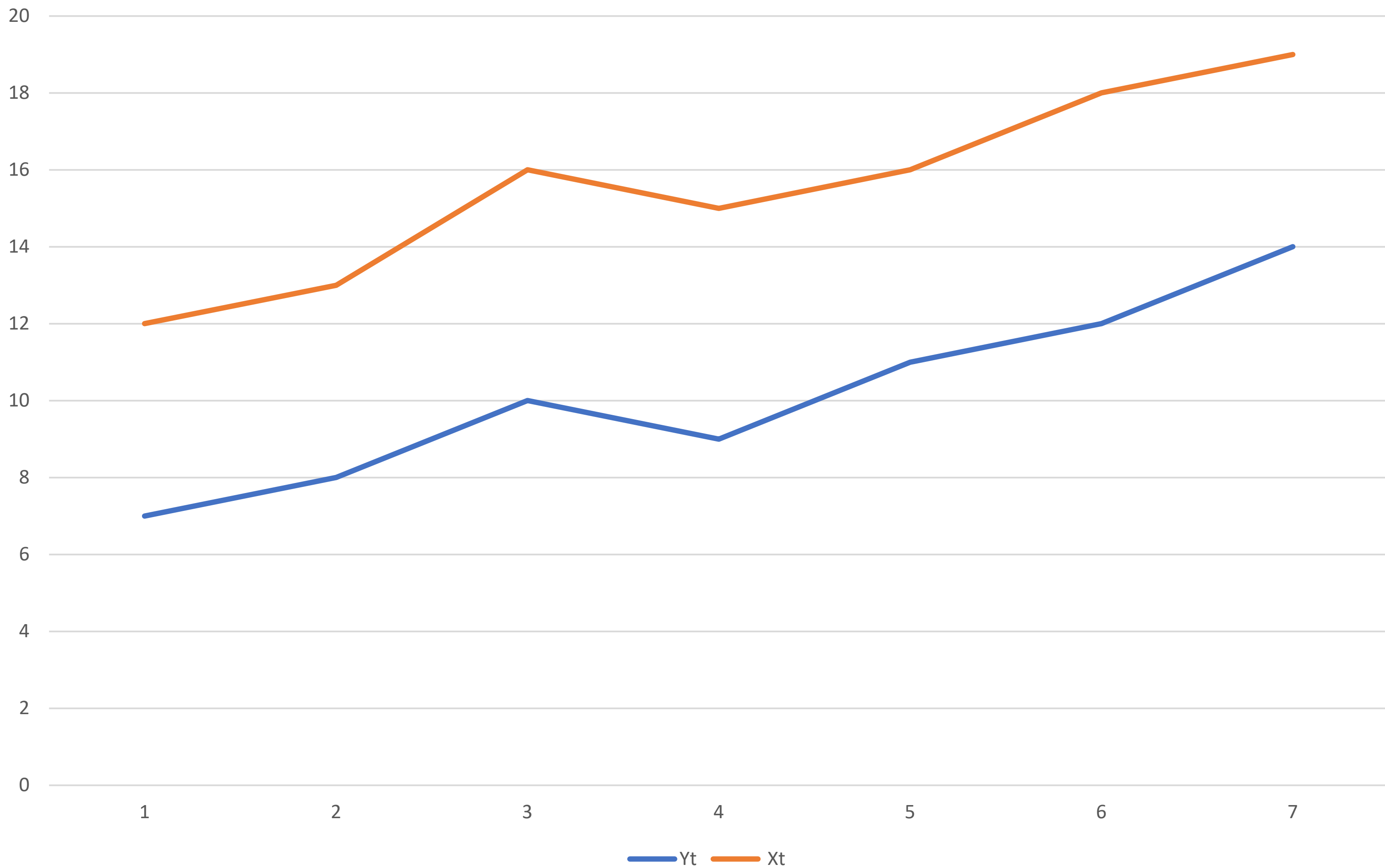
x_p - прогнозное значение x_t

$\hat{x}_{t=p}$ - прогноз x_t исходя из уравнения тренда при *t=p*

Метод отклонений от тренда

t	Расходы на потребление, Y_t	Доходы, X_t
1	7	12
2	8	13
3	10	16
4	9	15
5	11	16
6	12	18
7	14	19

Метод отклонений от тренда



Метод отклонений от тренда

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,960276599
R-квадрат	0,922131148
Нормированный R-квадрат	0,906557377
Стандартная ошибка	0,736788398
Наблюдения	7

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	32,14285714	32,14285714	59,21052632	0,000591241
Остаток	5	2,714285714	0,542857143		
Итого	6	34,85714286			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	5,857142857	0,622699849	9,406045089	0,00022912	4,256441936	7,457843778
t	1,071428571	0,139239919	7,694837641	0,000591241	0,713500964	1,429356179

$$\hat{y}_t = 5{,}88 + 1{,}07t$$

Метод отклонений от тренда

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,953959055
R-квадрат	0,910037879
Нормированный R-квадрат	0,892045455
Стандартная ошибка	0,823754471
Наблюдения	7

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	34,32142857	34,32142857	50,57894737	0,000852153
Остаток	5	3,392857143	0,678571429		
Итого	6	37,71428571			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	11,14285714	0,696199596	16,00526229	1,73378E-05	9,353219107	12,93249518
t	1,107142857	0,155674962	7,11188775	0,000852153	0,706967627	1,507318087

$$\hat{x}_t = 11{,}14 + 1{,}11t$$

Мет од от клонений от т ренда

t	y_t	X_t	e_y	e_x
1	7	12	0,071428571	-0,25
2	8	13	8,88178E-16	-0,357142857
3	10	16	0,928571429	1,535714286
4	9	15	-1,142857143	-0,571428571
5	11	16	-0,214285714	-0,678571429
6	12	18	-0,285714286	0,214285714
7	14	19	0,642857143	0,107142857

Мет од от клонений от т ренда

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,729664287
R-квадрат	0,532409972
Нормированный R-квадрат	0,438891967
Стандартная ошибка	0,503819994
Наблюдения	7

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	1,445112782	1,445112782	5,693127959	0,06269521
Остаток	5	1,269172933	0,253834587		
Итого	6	2,714285716			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	1,12316E-16	0,190426059	5,89815E-16	1	-0,489505767	0,489505767
ex	0,652631579	0,273522219	2,386027653	0,06269521	-0,05047967	1,355742828

$$\hat{e}_{yt} = 0{,}65\hat{e}_{xt}$$

Метод первых разностей

Для устранения влияния времени на результат и факторы при изучении взаимосвязанных временных рядов используется метод первых разностей.

Для этого рассчитываются первые разности (цепные абсолютные приросты) и строится регрессия разностей по y_t на разности по x_t.

Метод первых разностей

$$\Delta_{y_t} = y_t - y_{t-1}$$

$$\Delta_{x_t} = x_t - x_{t-1}$$

$$\Delta_{y_t} = a + b \cdot \Delta_{x_t}$$

Метод первых разностей

$$(y_t - y_{t-1}) = a + b(x_t - x_{t-1})$$

$$y_t = y_{t-1} + a + b(x_t - x_{t-1})$$

$$y_p = y_n + a + b(x_p - x_n)$$

Метод первых разностей

y_p - прогнозное значение уровня ряда y_t

y_n - конечный уровень динамического ряда y_t

x_p и x_n - то же по ряду x_t

Метод первых разностей

t	Y_t	X_t	ΔY_t	ΔX_t
1	7	12	-	-
2	8	13	1	1
3	10	16	2	3
4	9	15	-1	-1
5	11	16	2	1
6	12	18	1	2
7	14	19	2	1

Метод первых разностей

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,75082406
R-квадрат	0,563736769
Нормированный R-квадрат	0,454670962
Стандартная ошибка	0,863297758
Наблюдения	6

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	3,852201258	3,852201258	5,168776371	0,085397475
Остаток	4	2,981132075	0,745283019		
Итого	5	6,833333333			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	0,396226415	0,48893051	0,810394129	0,463172202	-0,961262307	1,753715137
Переменная X 1	0,660377358	0,290468006	2,273494309	0,085397475	-0,146091115	1,466845832

$$\widehat{\Delta y_t} = 0{,}40 + 0{,}66\widehat{\Delta x_t}$$

Метод включения фактора времени

Для устранения влияния времени на результат и факторы при изучении взаимосвязанных временных рядов используется прием включения времени t в качестве независимой переменной в модель регрессии, что позволяет зафиксировать воздействие фактора t.

Достоинством такого подхода является использование всей имеющейся выборки в отличие от метода последовательных разностей, который приводит к потере некоторого числа наблюдений.

Метод включения фактора времени

В качестве такой модели можно, например, использовать модель вида:

$$y_{\mathrm{t}} = \alpha + \beta_1 x_{\mathrm{t}} + \beta_2 t + \varepsilon_{\mathrm{t}},$$

которая относится к моделям с включенным фактором времени. Параметры модели определяются обычным МНК.

Метод включения фактора времени

Помимо линейной модели можно использовать нелинейные, например функцию Кобба-Дугласа

$$P = aK^{b_1}L^{b_2}e^{ct}$$

$$\ln P = \ln a + b_1 \ln K + b_2 \ln L + ct$$

Метод включения фактора времени

Y_t	X_t	t
7	12	1
8	13	2
10	16	3
9	15	4
11	16	5
12	18	6
14	19	7

Метод включения фактора времени

ВЫВОД ИТОГОВ							
Регрессионная статистика							
Множественный R	0,981625846						
R-квадрат	0,963589301						
Нормированный R-квадрат	0,945383952						
Стандартная ошибка	0,563287878						
Наблюдения	7						
Дисперсионный анализ							
	df	*SS*	*MS*	*F*	*Значимость F*		
Регрессия	2	33,58796992	16,79398496	52,92890995	0,001325739		
Остаток	4	1,269172932	0,317293233				
Итого	6	34,85714286					
	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	
Y-пересечение	-1,415037594	3,440659669	-0,411269271	0,701955603	-10,96784029	8,137765102	
Xt	0,652631579	0,305807138	2,134128012	0,09974384	-0,196425152	1,50168831	
t	0,34887218	0,354912705	0,982980254	0,38127029	-0,636523461	1,334267822	

$$\hat{y}_t = -1{,}42 + 0{,}65x_t + 0{,}35t$$

Автокорреляция в остатках

Условие отсутствия автокорреляции

$$M(\varepsilon_i \varepsilon_j) = 0, \qquad \text{при } i \neq j$$

или

$$M(\varepsilon^T \varepsilon) = \sigma_\varepsilon^2 I$$

где I – единичная матрица

Причины автокорреляции

Стохастические зависимости между значениями случайных ошибок – <u>автокорреляция остатков</u>.

Автокорреляция в остатках – типичное явление в моделях временных рядов.

Последствия автокорреляции:

Стандартные ошибки коэффициентов будут оценены неправильно, т.е. оценки коэффициентов будут неэффективны.

Выводы о значимости оценок коэффициентов будут некорректны.

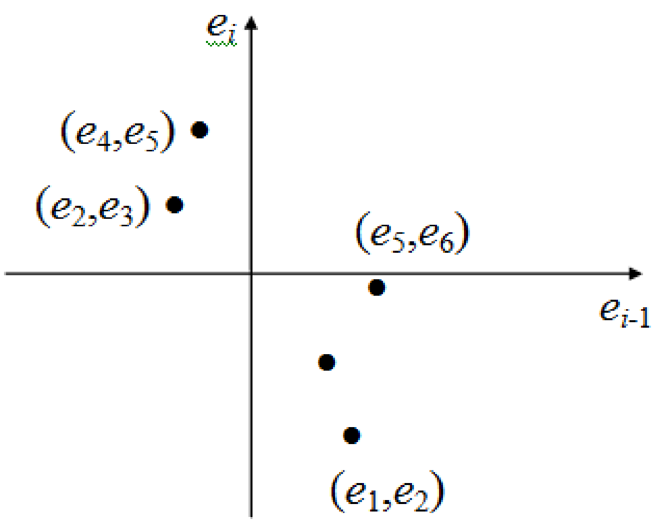
Можно рассматривать так называемую <u>корреляцию сериями</u> (автокорреляцию), когда зависимость между ошибками, отстоящими на некоторое количество шагов s, называемое <u>порядком корреляции</u> (в частности, на один шаг, s=1), остается одинаковой, что хорошо проявляется визуально на графике в системе координат (e_i; e_{i-s}).

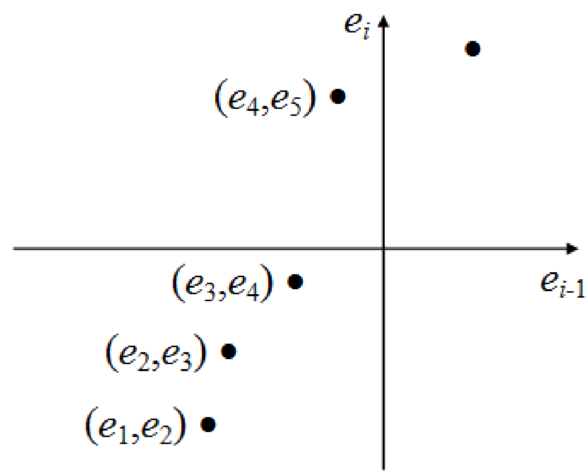
Например, для s=1 показаны отрицательная и положительная автокорреляция остатков. В экономических исследованиях чаще всего встречается положительная автокорреляция.

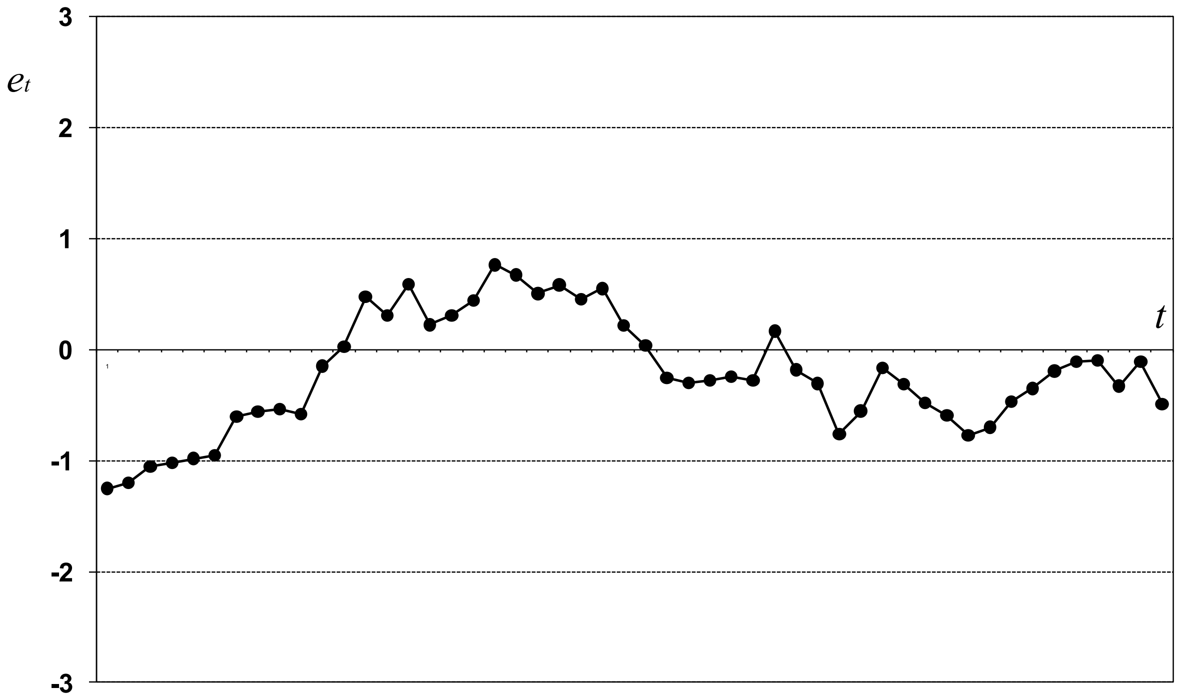
система координат

$(e_i; e_{i-s})$

Отрицательная и положительная автокорреляция остатков

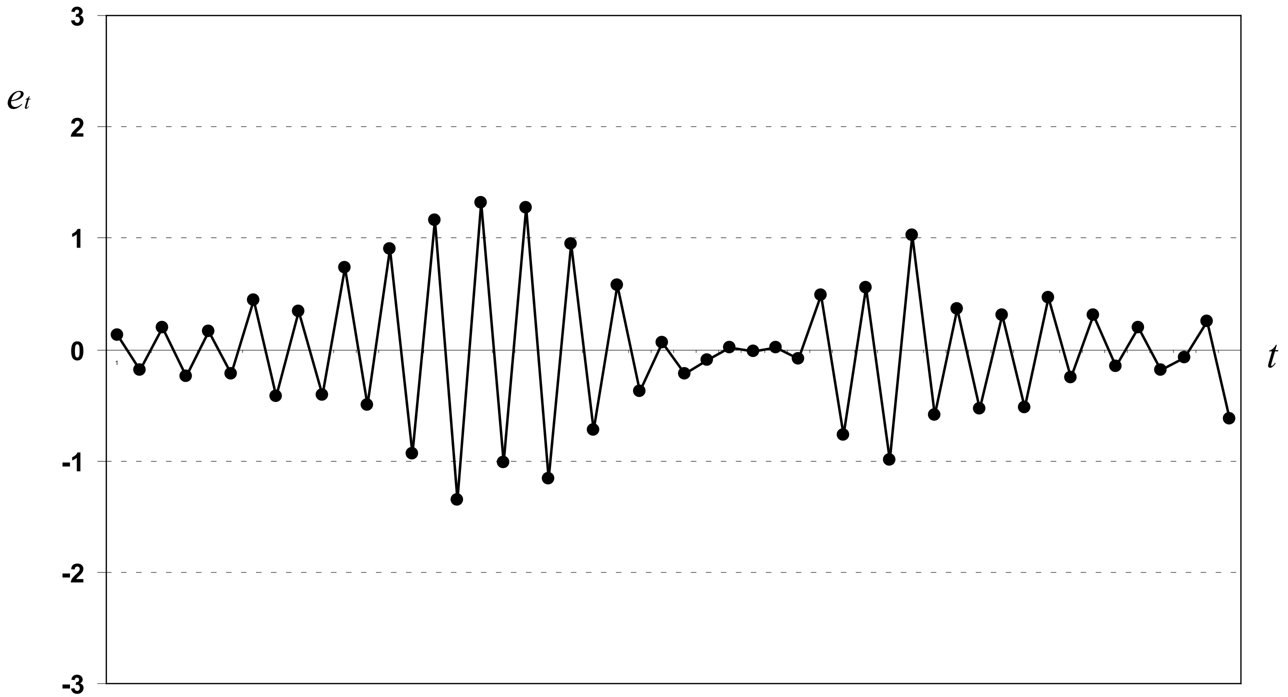






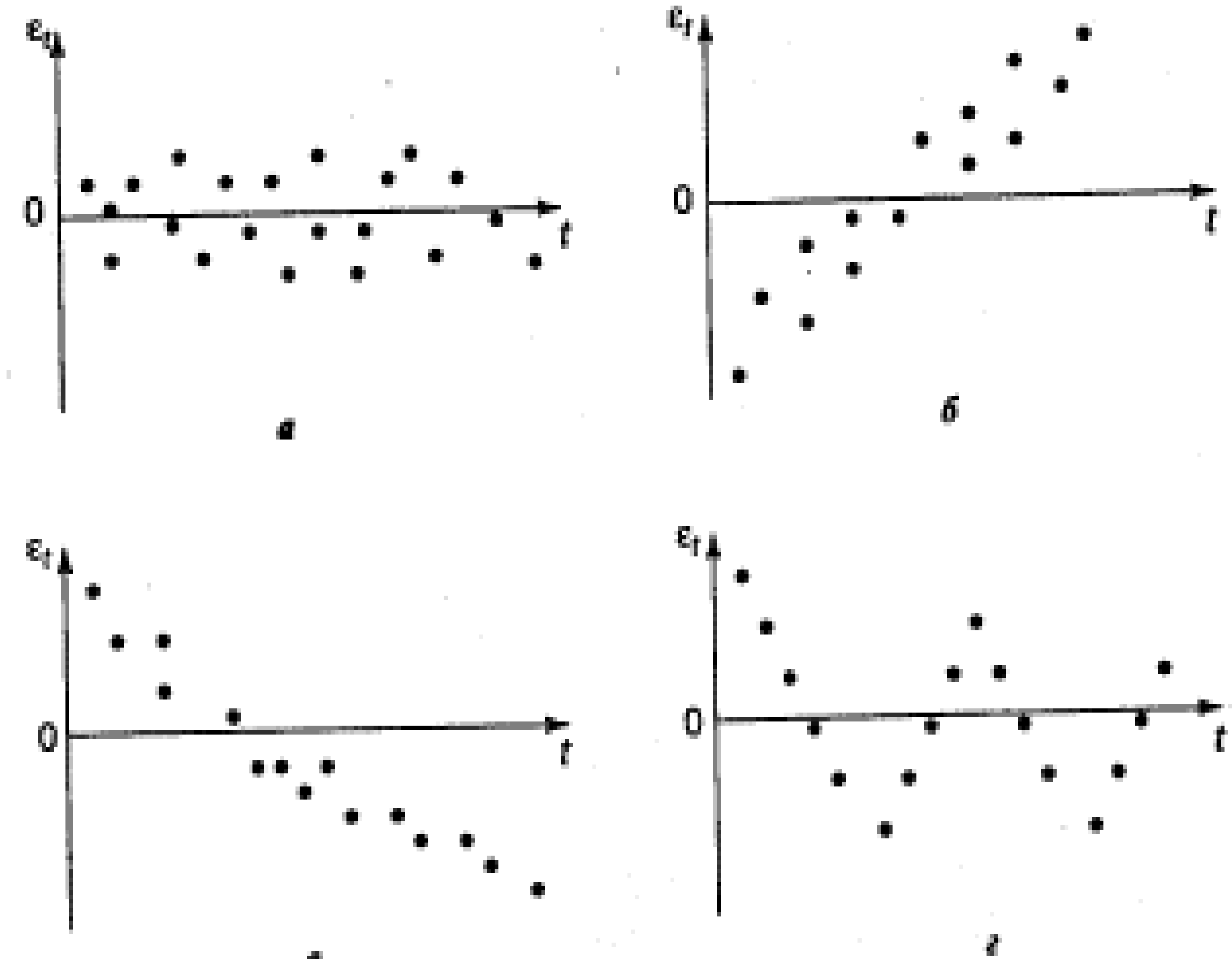
$$e_t = 0,9e_{t-1} + \varsigma_t$$

Пример графика остатков по наблюдениям при положительной автокорреляции



$$e_t = -0,9e_{t-1} + \varsigma_t$$

Пример графика остатков по наблюдениям при отрицательной автокорреляции



ε_t
0
t
а
ε_t
0
t
б
ε_t
0
t
в
ε_t
0
t
г

Тестирование на наличие автокорреляции.

Для проверки гипотезы о существовании линейной автокорреляции **первого порядка**, которая чаще всего имеет место на практике, используется критерий Дарбина-Уотсона, основанный на статистике

$$d = \frac{\sum_{i=2}^{n} \left(e_i - e_{i-1}\right)^2}{\sum_{i=1}^{n} e_i^2}$$

Критерий Дарбина-Уотсона.

$$d = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}$$

<u>Тестирование на наличие автокорреляции</u>.

Критерий Дарбина-Уотсона предназначен для моделей с детерминированными регрессорами X и не применим в случаях, когда среди объясняющих переменных есть лагированные значения переменной Y.

Для больших выборок	$d \to 2 - 2\rho$
Автокорреляция отсутствует	$d \to 2$
Положительная автокорреляция	$d \to 0$
Отрицательная автокорреляция	$d \to 4$

$$d \approx 2(1 - r_e)$$

Коэффициент автокорреляции первого порядка

$$r_{a_e} = \frac{\overline{e_t e_{t-1}} - \overline{e_t} \cdot \overline{e_{t-1}}}{\sigma_{e_t} \cdot \sigma_{e_{t-1}}}$$

$$r_e = \frac{\sum_{i=2}^{n} e_i e_{i-1}}{\sum_{i=1}^{n} (e_i)^2}$$

По таблице Дарбина-Уотсона определяются две критические точки: верхняя d_U и нижняя d_L.

Границы интервала (d_l и d_u) критических значений критерия Дарбина-Уотсона при уровне значимости α=0,05
(n - объем выборки, m - число объясняющих переменных

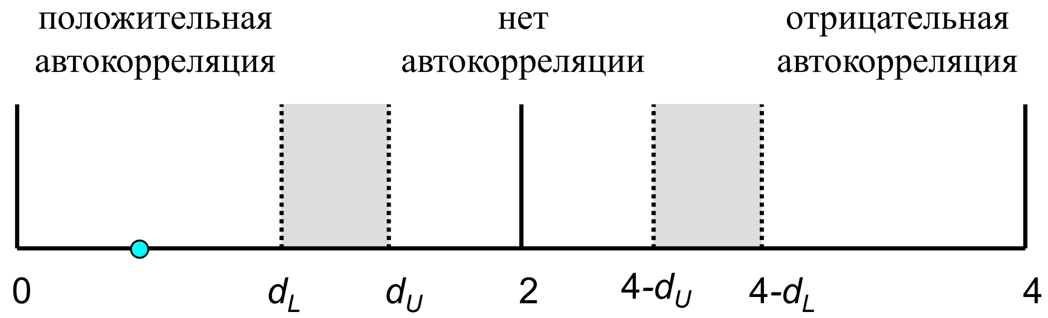
n	m=1		m =2		m =3		m =4	
	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u
6	0,610	1,400						
7	0,7000	1,356	0,467	1,896				
8	0,763	1,332	0,359	1,777	0,368	2,287		
9	0,824	1,320	0,629	1,699	0,435	2,128	0,296	2,388
10	0,879	1,320	0,697	1,641	0,525	2,016	0,356	2,414
11	0,927	1,324	0,658	1,604	0,595	1,928	0,444	2,283
12	0,971	1,331	0,812	1,576	0,658	1,864	0,512	2,177
13	1,010	1,340	0,861	1,562	0,715	1,816	0,574	2,094
14	1,045	1,330	0,905	1,551	0,767	1,779	0,632	2,030
15	1,077	1,361	0,946	1,543	0,814	1,750	0,685	1,977
16	1,106	1,371	0,982	1,539	0,857	1,728	0,734	1,935

Области статистических решений для критерия Дарбина-Уотсона

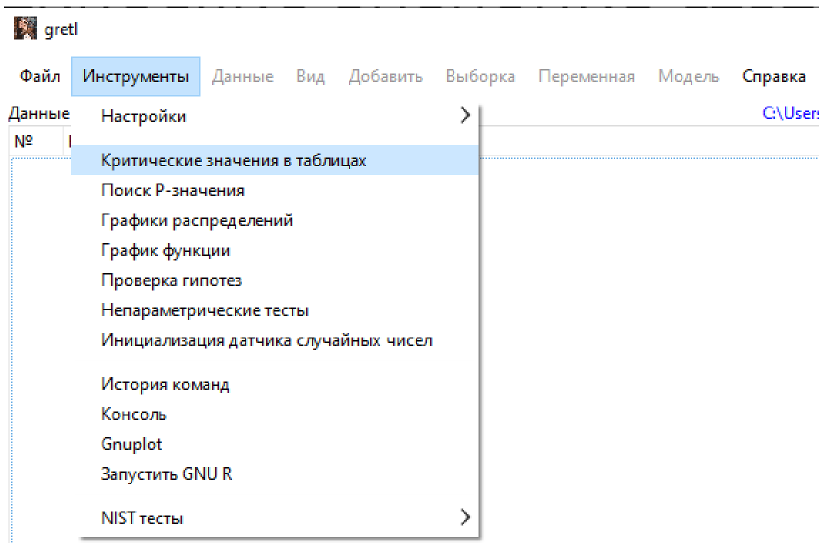
H_0: $\rho=0$ (автокорреляции нет)

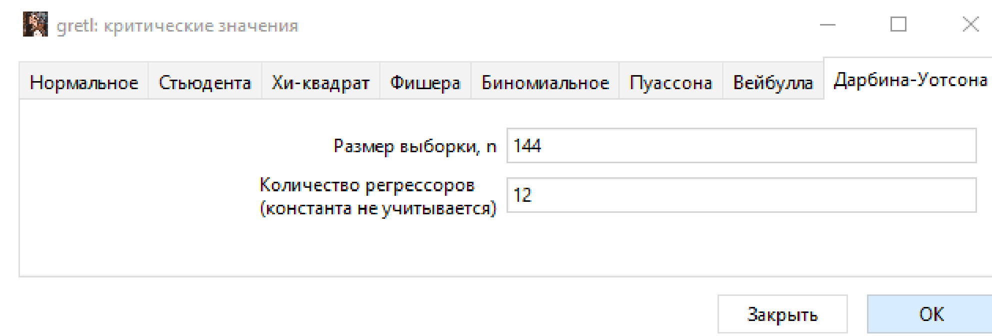
H_1: $\rho\neq0$ (автокорреляция есть)

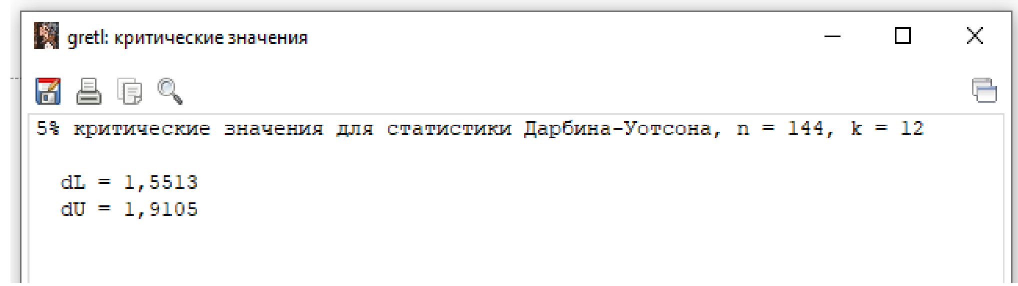
$d<d_L$	$d_L<d<d_U$	$d_U<d<2$; $2<d<(4-d_U)$	$(4-d_U)<d<(4-d_L)$	$d>(4-d_L)$
Отвергаем H_0 в пользу гипотезы о положительной автокорреляции	H_0 не принимается и не отвергается	Принимается H_0	H_0 не принимается и не отвергается	Отвергаем H_0 в пользу гипотезы об отрицательной автокорреляции



Критические значения статистики Дарбина-Уотсона в Gretl







Тест Бройша — Годфри

Тест Бройша — Годфри, называемый также LM-тест Бройша — Годфри на автокорреляцию (англ. Breusch-Godfrey serial correlation LM-test) — применяемая в эконометрике процедура проверки автокорреляции произвольного порядка в случайных ошибках регрессионных моделей. Тест является асимптотическим, то есть для достоверности выводов требуется большой объём выборки.

Особенность данного теста заключается в том, что его можно использовать практически всегда, в отличие от, например, критерия Дарбина — Уотсона или h-теста Дарбина. Кроме того, указанные тесты проверяют только автокорреляцию первого порядка, тогда как тест Бройша — Годфри позволяет проверить автокорреляцию любого порядка.

Тест Бройша — Годфри

Для проверки автокорреляции порядка p тест использует вспомогательную регрессию МНК-остатков исходной модели на факторы этой модели и лаговые значения остатков:

$$e_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{k-1} X_{k-1} + \beta_k \mathrm{X}_k + \sum_{i=1}^{p} \alpha_i \, e_{t-i} + u_t$$

Далее для этой вспомогательной регрессии проверяется гипотеза об одновременном равенстве нулю всех коэффициентов при лаговых остатках. Проверка осуществляется с помощью соответствующей LM-статистики, равной
nR^2 ,
где R^2 — коэффициент детерминации вспомогательной модели,
а n — объём выборки (этот объём выборки на p меньше объёма выборки для исходной модели, так как из-за лаговых значений остатков во вспомогательной регрессии первые p наблюдений не учитываются).
Статистика теста имеет асимптотическое распределение $\chi^2(p)$. Если значение статистики превышает критическое значение, то автокорреляция признаётся значимой, в противном случае она незначима.

Обобщенный мет од наименьших квадрат ов при пост роении модели регрессии по временным рядам (ОМНК)

Алгоритм ОМНК

1. Преобразование исходных переменных

$$x_t^* = x_t - r_e x_{t-1} \qquad y_t^* = y_t - r_e y_{t-1}$$

2. Применение обычного МНК к уравнению и определение *a** и *b*

$$y_t^* = a^* + bx_t^* + V_t$$

3. Расчет параметра *a*

$$a = \frac{a^*}{1 - r_e}$$

4. *Переход к исходному уравнению*

$$y_t = a + bx_t + e_t$$

Поправка Прайса-Винст ена

$$x_1^* = x_1 \cdot \sqrt{1 - r_e^2}$$

$$y_1^* = y_1 \cdot \sqrt{1 - r_e^2}$$

$$r_e = 1 - \frac{d}{2}$$

Пример

По данным за 1995-2003 гг. по Тамбовской области рассматривается зависимость потребления растительного масла на душу населения (y, кг) от потребления овощей (x, кг)

Годы	y_t	x_t	e_t	e_{t-1}	y_t^*	x_t^*
1	2	3	4	5	6	7
1995	7,2	100	-0,272	-	5,97	82,916
1996	7,5	101	-0,108	-0,272	11,525	156,9
1997	7,9	103	0,021	0,108	12,093	159,46
1998	8,6	105	0,450	0,021	13,016	162,578
1999	9,5	121	-0,818	0,450	14,307	179,696
2000	10,9	121	0,582	-0,818	16,210	188,64
2001	10,3	120	0,117	0,582	16,393	187,64
2002	10,1	119	0,053	0,117	15,858	186,081
2003	10,7	124	-0,025	0,053	16,346	190,522

$$\hat{y}_t = -6{,}0791 + 0{,}1355x_t$$

$$t \qquad -3{,}5 \qquad +8{,}9$$

$$R^2 = 0{,}9188 \qquad F = 79{,}2$$

$$r_e = \frac{\sum e_t e_{t-1}}{\sum e_t^2} = \frac{-0{,}734}{1{,}314} = -0{,}55901$$

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,958560316
R-квадрат	**0,9188**
Нормированный R-квадрат	0,907243291
Стандартная ошибка	0,433247629
Наблюдения	9

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	14,87496433	14,87496433	**79,25**	4,58001E-05
Остаток	7	1,313924554	0,187703508		
Итого	8	16,18888889			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	**-6,0791**	1,721168111	**-3,53**	0,009572817	-10,14898307	-2,009151358
Переменная X 1	**0,1355**	0,01522277	**8,90**	4,58001E-05	0,099518272	0,171510534

ВЫВОД ОСТАТКА

Наблюдение	*Предсказанное Y*	*Остатки*				
1	7,472373114	-0,272373114		0,074187113		
2	7,607887517	-0,107887517	-0,272373114	0,011639716	0,029385659	
3	7,878916324	0,021083676	-0,107887517	0,000444521	-0,002274665	
4	8,14994513	0,45005487	0,021083676	0,202549386	0,009488811	
5	10,31817558	-0,818175583	0,45005487	0,669411285	-0,368223905	
6	10,31817558	0,581824417	-0,818175583	0,338519652	-0,476034532	
7	10,18266118	0,11733882	0,581824417	0,013768399	0,068270591	
8	10,04714678	0,052853224	0,11733882	0,002793463	0,006201735	
9	10,72471879	-0,024718793	0,052853224	0,000611019	-0,001306468	
				1,3139	**-0,7345**	**-0,55901**

Расчет преобразованных значений

$$y_1^* = 7{,}2\sqrt{1-(-0{,}559)^2} = 5{,}97$$

$$x_1^* = 100\sqrt{1-(-0{,}559)^2} = 82{,}916$$

$$y_2^* = 7{,}5-(-0{,}559)7{,}2 = 11{,}525$$

$$x_2^* = 101-(-0{,}559)100 = 156{,}9$$

- и т.д.

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,980278416
R-квадрат	**0,9609**
Нормированный R-квадрат	0,955366597
Стандартная ошибка	0,719703535
Наблюдения	9

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	89,21464237	89,21464237	172,2379576	3,47918E-06
Остаток	7	3,62581225	0,517973179		
Итого	8	92,84045462			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	**-2,8156**	1,26794285	-2,22058418	0,061823935	-5,813782246	0,182634578
Переменная X 1	**0,0984**	0,007498076	13,12394596	3,47918E-06	0,080674207	0,11613447

$$y^* = -2{,}8156 + 0{,}0984x^*$$

$R^2 = 0{,}961$ $F = 172{,}3$

$$a = a^* / (1 - r_e) = -1{,}806$$

$$\hat{y} = -1{,}806 + 0{,}0984x$$

$R^2 = 0{,}845$ $r_e = 0{,}239$

Что делать, если вы столкнулись с нестационарными рядами?

1. Перейти к стационарным разностям или темпам прироста и оценивать зависимости для них
2. Тестировать наличие коинтеграции между переменными, и если она будет обнаружена, оценивать коинтеграционное соотношение

Коинтеграция

Интуитивное (**неформальное**) определение:

Временные ряды называют **коинтегрированными**, если они нестационарны, и между ними существует долгосрочная устойчивая связь.

Коинтеграция

Формальное определение:

Если

1. Временные ряды x_t и y_t нестационарны,
2. Временные ряды x_t и y_t интегрированы первого порядка,
3. Существует такое число θ, что временной ряд ($y_t - \theta x_t$) является стационарным, то **временные ряды x_t и y_t коинтегрированы**.

В этом случае число θ называют **коинтеграционным коэффициентом**,

а **вектор $(1, -\theta)$ — коинтегрирующим вектором**.

Коинтеграция

Пусть ряды x_t и y_t нестационарны.

Как понять, можно ли доверять оценкам коэффициентов в регрессии $y_t = \alpha + \theta x_t$?

Если ряды коинтегрированы, то доверять можно: в этом случае коэффициенты в такой модели оцениваются (супер)состоятельно.

Если ряды не коинтегрированы, то доверять нельзя: вы имеете дело с ложной регрессией.

Примеры коинтеграции

1. Длинная ставка процента R, короткая ставка процента r: $\varepsilon_t = R_t - r_t$, вектор коинтеграции (1, –1).

2. Логарифм потребления C_t, логарифм дохода y_t: $\varepsilon_t = C_t - y_t$, вектор коинтеграции (1, –1).

3. Логарифм обменного курса D_t, логарифм внутренней цены P_t, логарифм цен мирового рынка P_t^*: $\varepsilon_t = D_t - P_t + P_t^*$, вектор коинтеграции (1, –1, 1).

Коинтеграция

- В случае коинтегрируемости временных рядов говорят о долгосрочном динамическом равновесии. Если y_t и x_t коинтегрированы, то y_t и θx_t содержат общую нестационарную компоненту – долговременную тенденцию, а разность $y_t - \theta x_t$ стационарна и совершает флуктуации около нуля.
- Таким образом, **коинтеграция временных рядов** – причинно-следственная зависимость в уровнях временных рядов, которая выражается в совпадении или противоположной направленности их тенденций и случайной колеблемости.

Коинтеграция

Возможен случай, когда ошибка $\varepsilon_t = y_t - \Theta x_t$, $t=1,\ldots,n$ является нестационарным временным рядом.

Тогда условия классической регрессионной модели не выполняются, в частности дисперсия ε_t не является постоянной.

Кроме того, МНК оценка параметра Θ не состоятельна, поэтому с ростом объема выборки увеличиваются шансы получения ложных выводов о зависимости y_t от x_t.

Такая ситуация называется ложной (мнимой) регрессией.

На практике признаками мнимой регрессии являются высокое значение R^2 и малое значение статистики Дарбина-Уотсона.

Тестирование коинтеграции

Для проверки рядов на коинтеграцию используются тесты:

1. Энгла-Гренжера;
2. Йохансена.

Тестирование коинтеграции
Тест Энгла — Грейнжера

1. Тестируем стационарность ряда $\boldsymbol{x_t}$
2. Тестируем стационарность ряда $\boldsymbol{y_t}$
3. Оцениваем регрессию $\boldsymbol{y_t} = \boldsymbol{\alpha} + \boldsymbol{\theta x_t}$, получаем ряд остатков $\boldsymbol{e_t}$.
4. Тестируем стационарность ряда остатков.

Если на 1 и 2 шаге — нестационарность, а на 4 шаге — стационарность, то ряды x_t и y_t коинтегрированы.

Тестирование коинтеграции
Тест Энгла — Грейнжера

Примечания:

Стационарность тестируется при помощи ADF-теста.

При этом на четвертом шаге надо использовать специальные критические значения: Davidson and MacKinnon, 1993.

Специальные критические значения: Davidson and MacKinnon, 1993

Число переменных	Тип теста	Уровень значимости		
		0,01	0,05	0,10
2	константа	-3,90	-3,34	-3,04
	константа и тренд	-4,32	-3,78	-3,50
3	константа	-4,29	-3,74	-3,45
	константа и тренд	-4,66	-4,12	-3,84
4	константа	-4,64	-4,10	-3,81
	константа и тренд	-4,97	-4,43	-4,15
5	константа	-4,96	-4,42	-4,13
	константа и тренд	-5,25	-4,72	-4,43
6	константа	-5,25	-4,71	-4,42
	константа и тренд	-5,52	-4,98	-4,70

Число переменных: количество переменных в уравнении коинтеграции. Например, в уравнении $\boldsymbol{y_t} = \boldsymbol{\alpha} + \boldsymbol{\theta x_t}$ – две переменных.

Тип теста: в уравнение, тестирующее наличие единичных корней в остатках, могут быть включены константа и временной тренд.

Тестирование коинтеграции

Тест Энгла — Грейнжера

Коинтегрированы ли потребление и располагаемый доход?

Используем квартальные данные о: совокупном потреблении (consum) и располагаемом доходе (income) в Великобритании за период с 1 квартала 1971 года по 2 квартал 1985 года, содержащиеся в файле consum.xlsx. Воспользуемся тестом Энгла-Грейнжера в пакете Gretl.

Коинтегрированы ли потребление и располагаемый доход?

Шаг 1: тестирование единичного корня для consum

```
Шаг 1: тестирование единичного корня для consum

Расширенный тест Дики-Фуллера для consum
включая 4 лага(-ов) для (1-L)consum
объем выборки 53
нулевая гипотеза единичного корня: a = 1

   тест с константой
   модель: (1-L)y = b0 + (a-1)*y(-1) + ... + e
   оценка для (a - 1): 0,00105428
   тестовая статистика: tau_c(1) = 0,150097
   асимпт. p-значение 0,9694
   коэф. автокорреляции 1-го порядка для e: -0,070
   лаг для разностей: F(4, 47) = 95,820 [0,0000]
```

Коинтегрированы ли потребление и располагаемый доход?

Шаг 2: тестирование единичного корня для income

```
Шаг 2: тестирование единичного корня для income

Расширенный тест Дики-Фуллера для income
включая 4 лага(-ов) для (1-L)income
объем выборки 53
нулевая гипотеза единичного корня: a = 1

  тест с константой
  модель: (1-L)y = b0 + (a-1)*y(-1) + ... + e
  оценка для (a - 1): 0,00498682
  тестовая статистика: tau_c(1) = 0,711855
  асимпт. p-значение 0,9925
  коэф. автокорреляции 1-го порядка для e: -0,035
  лаг для разностей: F(4, 47) = 6,425 [0,0003]
```

Коинтегрированы ли потребление и располагаемый доход?

Шаг 3: коинтеграционная регрессия

```
Шаг 3: коинтеграционная регрессия

Коинтеграционная регрессия -
МНК, использованы наблюдения 1971:1-1985:2 (T = 58)
Зависимая переменная: consum

             коэффициент    ст. ошибка    t-статистика   p-значение
  ---------------------------------------------------------------
  const       176,848      258,415           0,6844     0,4966
  income        0,869020     0,00749733     115,9       2,16e-068  ***

Среднее завис. перемен  26772,86   Ст. откл. завис. перем  13927,30
Сумма кв. остатков      45892826   Ст. ошибка модели       905,2705
R-квадрат               0,995849   Исправ. R-квадрат       0,995775
Лог. правдоподобие     -476,1583   Крит. Акаике            956,3167
Крит. Шварца            960,4376   Крит. Хеннана-Куинна    957,9219
параметр rho            0,140845   Стат. Дарбина-Уотсона   1,675463
обратите внимание на сокращенные обозначения статистики
```

Коинтегрированы ли потребление и располагаемый доход?

Шаг 4: тестирование единичного корня для uhat

Расширенный тест Дики-Фуллера для uhat
тест. начиная с 4 лагов, критерий BIC
объем выборки 53
нулевая гипотеза единичного корня: a = 1
  тест с константой
  включая 4 лага(-ов) для (1-L)uhat1
  модель: (1-L)y = b0 + (a-1)*y(-1) + ... + e
  оценка для (a - 1): -0,361661
  тестовая статистика: tau_c(1) = -1,95005
  асимпт. p-значение 0,3094
  коэф. автокорреляции 1-го порядка для e: -0,118
  лаг для разностей: F(4, 47) = 21,143 [0,0000]

> Коинтеграционная связь присутствует если:
> (а) Гипотеза единичного корня не отвергается для отдельных переменных.
> (б) Гипотеза единичного корня отвергается для остатков (uhat) коинтеграционной регрессии.

Коинтегрированы ли потребление и располагаемый доход?

Переменные consum и income – нестационарны, т.е. условия 1) и 2) выполнены, но остатки также нестационарны, т.е. не выполнено условие 4).

Это говорит о том, что регрессия consum на income может оказаться ложной, и доверять полученному на шаге 3 коинтеграционному соотношению нельзя, несмотря на высокий R^2 и значимость коэффициента регрессии.

Коинтеграция

При изучении двух взаимосвязанных временных рядов на предварительной стадии регрессионного анализа рекомендуется устранить сезонные или циклические колебания, если они имеются в исследуемых временных рядах, в соответствии с принятой аддитивной или мультипликативной моделями рядов.

Коинтеграция

При изучении двух взаимосвязанных временных рядов на предварительной стадии регрессионного анализа рекомендуется устранить сезонные или циклические колебания, если они имеются в исследуемых временных рядах, в соответствии с принятой аддитивной или мультипликативной моделями рядов.

Модель коррекции ошибок

Error correction model (ECM)

Теорема представления Грейнджера:

Если ряды x_t и y_t коинтегрированы с коинтеграционным коэффициентом θ, то связь между ними может быть представлена следующим образом:

$$\Delta y_t = \delta + \sum_{j=1}^{p} \beta_j \Delta y_{t-j} + \sum_{j=0}^{q} \varphi_j \Delta x_{t-j} - \gamma z_{t-1} + \varepsilon_t$$

$$z_t = y_t - \alpha - \theta x_t, \gamma > 0$$

Модель коррекции ошибок

$$\Delta y_t = \delta + \sum_{j=1}^{p} \beta_j \Delta y_{t-j} + \sum_{j=0}^{q} \varphi_j \Delta x_{t-j} - \gamma z_{t-1} + \varepsilon_t$$

$$z_t = y_t - \alpha - \theta x_t, \gamma > 0$$

Оценивание ECM

На практике при оценивании ECM вместо zt следует брать остатки, полученные при оценке уравнения:

$$y_t = \alpha + \theta x_t$$

Модель коррекции ошибок

Интерпретация:

- Уравнение $yt = \alpha + \theta xt$ интерпретируют как долгосрочное равновесное соотношение между переменными.
- ECM интерпретируют как модель краткосрочных колебаний переменных вокруг долгосрочного равновесного соотношения.
- Оценка параметра γ интерпретируется как характеристика скорости возвращения к долгосрочному равновесию.