Учет автокорреляции: ARIMA

Метод ARIMA (авторегрессионные модели интегрированного скользящего среднего) позволяет учитывать как влияние предыдущих уровней исходного ряда на прогнозируемый уровень, так и влияние ошибок прогнозной модели на предыдущих этапах, а также сезонную компоненту.

ARIMA применима только к **стационарным** временным рядам, т.е. таким, у которых *нет тренда, нет сезонности и дисперсия постоянна*.

В случае, когда исходный ряд является нестационарным, его приводят к стационарному путем подбора определенного типа модели ARIMA.

Если в исходном ряду нет сезонности, то строят модель **ARIMA (p,d,q)**:

- Параметр p (AR) – **авторегрессия**, т.е. реагирование на последние уровни ряда.

$$\hat{Y}_t = B_0 + \sum_{i=1}^{p} AR_i \cdot Y_{t-i} + \sum_{j=1}^{k} B_j \cdot X_j$$

- Параметр d – **порядок интегрируемости** или порядок разностей уровней ряда, вычисление которых может привести ряд к стационарному (чтобы стабилизировать дисперсию, как привило, логарифмируют уровни ряда по основанию e).

$\Delta_t^{(1)} = Y_{t+1} - Y_t$ – разности 1-го порядка (d = 1); $\Delta_t^{(2)} = \Delta_{t+1}^{(1)} - \Delta_t^{(1)}$ – разности 2-го порядка (d = 2)

- Параметр q (MA) – **скользящее среднее**, т.е. реагирование на последние ошибки.

$$\hat{Y}_t = B_0 + \sum_{i=1}^{q} MA_i \cdot e_{t-i} + \sum_{j=1}^{k} B_j \cdot X_j$$

Когда в исследуемом ряду присутствует сезонность, то ARIMA дополняется еще одним набором параметров – **ARIMA (p,d,q)(sp,sd,sq)**.

На практике, чаще всего, каждый из параметров модели ARIMA может принимать значения 0, 1 или 2.

ARIMA (p,0,0)	Модель, без тренда и сезонности. Учитывает только автокорреляцию уровней ряда на p-ом лаге (модель AR (p)). Исходный ряд стационарен.
ARIMA (0,0,q)	Модель без тренда и сезонности. Учитывает только автокорреляцию остатков на q-ом лаге (модель MA (q)). Исходный ряд стационарен.
ARIMA (p,0,q)	Модель без тренда и сезонности. Учитывает автокорреляцию на p-ом лаге и автокорреляцию остатков на q-ом лаге (модель ARMA (p,q)). Исходный ряд стационарен.
ARIMA (p,d,q)	Модель, позволяющая учитывать только тренд и автокорреляцию. Исходный ряд не стационарен.
ARIMA (p,d,q)(sp,sd,sq)	Модель, позволяющая учитывать тренд, сезонность и автокорреляцию. Исходный ряд не стационарен.

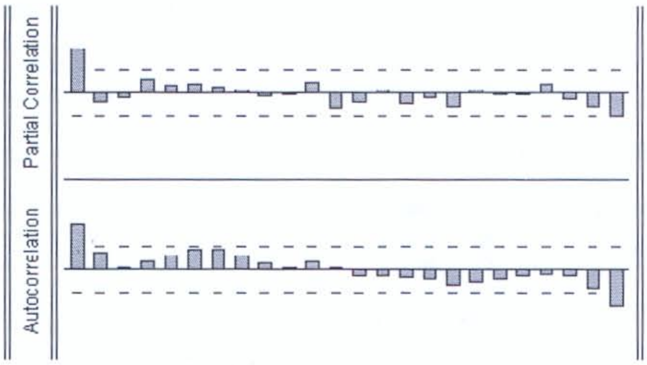
Еще раз о видах графиков автокорреляции и частной корреляции для идентификации моделей ARIMA:

Модель	Автокорреляция	Частная корреляция
«Белый шум» - «хорошие» остатки модели (нулевое среднее, постоянная дисперсия), процесс MA(0)	Отсутствует	Отсутствует
AR(1)	Тенденция к уменьшению	Всплеск на 1-ом лаге
AR(p)	Тенденция к уменьшению	Обнуление, начиная с лага p + 1
MA(1)	Всплеск на 1-ом лаге	Тенденция к уменьшению
MA(q)	Обнуление, начиная с лага q + 1	
ARMA(1,1)	Тенденция к уменьшению	Тенденция к уменьшению
ARMA(p,q)	Тенденция к уменьшению, начиная с лага q	Тенденция к уменьшению, начиная с лага p

Учет автокорреляции

Для прогнозирования железнодорожных перевозок с помощью регрессионного метода на основе LS прослеживалась остаточная автокорреляция (стр. 17).

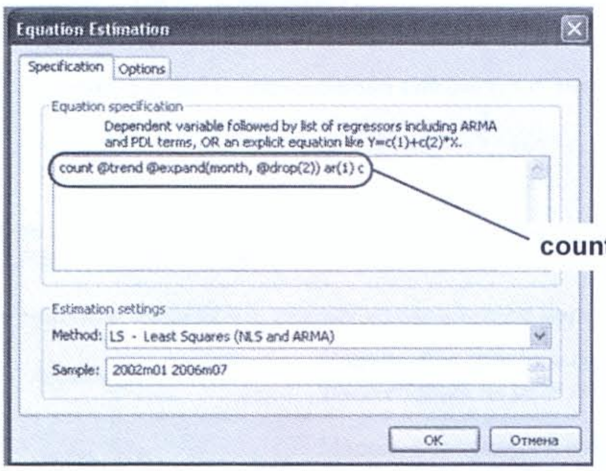
Применим метод ARIMA к данным, предварительно выбрав его параметры с помощью графиков автокорреляции и частной автокорреляции остатков.



Вид графиков указывает на спецификацию модели AR (1).

Будем строить модель вида:

$$\hat{Y}_t = B_0 + AR_1 \cdot Y_{t-1} + A \cdot TREND + \sum_{i=1}^{11} B_i \cdot D_i$$



count @trend @expand(month, @drop(2)) ar(1) c

Dependent Variable: COUNT
Method: Least Squares
Date: 03/04/09 Time: 17:54
Sample (adjusted): 2002M02 2006M07
Included observations: 54 after adjustments
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
@TREND	0.385042	0.021295	18.08164	0.0000
C	80.79028	0.879158	91.89502	0.0000
MONTH=1	1.858072	0.590819	3.144909	0.0031
MONTH=3	10.64361	0.568190	18.73247	0.0000
MONTH=4	8.545016	0.706026	12.10298	0.0000
MONTH=5	10.84901	0.770549	14.07958	0.0000
MONTH=6	8.546261	0.803091	10.64171	0.0000
MONTH=7	10.80118	0.819154	13.18578	0.0000
MONTH=8	12.02436	0.859931	13.98293	0.0000
MONTH=9	10.36868	0.865633	11.97815	0.0000
MONTH=10	14.36096	0.852537	16.84497	0.0000
MONTH=11	10.35215	0.818711	12.64445	0.0000
MONTH=12	8.817765	0.747847	11.79087	0.0000
AR(1)	0.531354	0.115871	4.585745	0.0000

R-squared	0.983830	Mean dependent var	100.1296
Adjusted R-squared	0.978574	S.D. dependent var	7.533088
S.E. of regression	1.102653	Akaike info criterion	3.251729
Sum squared resid	48.63376	Schwarz criterion	3.767392
Log likelihood	-73.79669	Hannan-Quinn criter.	3.450600
F-statistic	187.2064	Durbin-Watson stat	1.502368
Prob(F-statistic)	0.000000		

Inverted AR Roots .53

Авторегрессионная составляющая в модели значима (Prob. < 0.05).

Проверив нормальность распределения остатков полученной модели, отсутствие остаточной автокорреляции на коррелограмме можно утверждать об успешном учете проблемы автокорреляции.
Хорошие показатели качества указывают на возможность прогноза по данной модели.

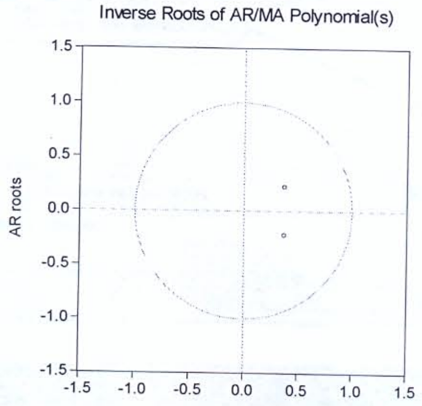
Общие замечания по учету автокорреляции

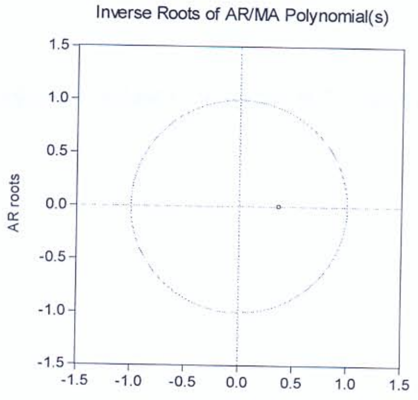
- Учет сезонной автокорреляции может осуществляться с использованием параметров SAR и SMA. Коэффициенты моделей с добавками AR(12) и SAR(12) будут различаться.
- Часто бывает сложно однозначно идентифицировать параметры модели ARIMA (наличие и порядок AR, MA, SAR и SMA компонент). Поэтому строят несколько моделей и выбирают лучшую согласно «хорошим» остаткам, наибольшему R-squared и наименьшему значению критерия Шварца.
- В EViews SAR компонента не может участвовать в модели без AR компоненты (если нет AR, то SAR превращается а AR). Аналогично для SMA компоненты.
- Коэффициенты модели при AR, MA, SAR и SMA компонентах не интерпретируются. Они являются лишь «добавками», которые улучшают построенную модель.
- Для получения адекватного прогноза корни AR и MA процессов не должны превосходить единицу. Если они близки к единице, то, вероятно, при приведении ряда к стационарному, была взята излишняя разность. (Процессы ARIMA определенным образом выражаются при помощи алгебраических уравнений, для которых находят корни, как вещественные, так и комплексные.)

Inverted AR Roots	.38-.22i	.38+.22i

Inverted AR Roots	.37	.37

View – ARMA Structure…, Roots, Display Graph





Суть стационарности и проверка на стационарность

На практике **стационарность** временного ряда означает отсутствие:

- Тренда
- Сезонности
- Резких изменений дисперсии (т.е. стационарный ряд имеет постоянную дисперсию, является гомоскедастичным)

Замечания

- Наличие или отсутствие стационарности ряда влияет на выбор метода моделирования
- В случае использования в ряде методов моделирования нестационарных рядов, можно получить «ложную регрессию», в которой кажущееся влияние факторов на отклик обусловлено, например, общей тенденцией к росту значений во времени

Графические методы проверки на стационарность

- Анализ графика и коррелограммы исходного ряда
- Анализ графика и коррелограммы ряда, очищенного от тренда, путем взятия первой (второй) разности: `genr new_1=d(new,1,0)`
- Анализ графика и коррелограммы ряда, очищенного от тренда и сезонности, путем взятия первой (второй) разности и первой (второй) сезонной разности: `genr new_2=d(new,1,12)`
- Анализ графика и коррелограммы ряда, очищенного от тренда, сезонности и неоднородности дисперсии путем взятия первой (второй) разности, первой (второй) сезонной разности и логарифмирования: `genr new_3=dlog(new,1,12)`

Аналитические методы проверки на стационарность (тесты единичного корня)

- Тесты Augment Dickey-Fuller (ADF), Phillips-Perron (PP), Kwiatkowski-Phillips-Schmidt-Shin (KPSS) и др.

Все реализованные в EViews тесты единичного корня основаны на различных математических идеях и имеют разную мощность. Для оценки стационарности ряда рекомендуется использовать несколько тестов, указанных в схеме.

ВАЖНО! В EViews аналитические методы обнаружения стационарности применимы лишь к тем рядам, которые очищены от сезонности, т.е.:

- Либо к данным, полученным после взятия сезонной разности (обычно 1-ой или 2-ой)
- Либо к данным, полученным в результате сезонной декомпозиции
- Либо к остаткам, полученным от регрессионной модели, в которой откликом является исходный временной ряд, а факторами – DUMMY для сезонности

Идея теста ADF

Пусть временной ряд имеет линейный тренд и автокорреляцию порядка *p*, т.е. описывается моделью:

$$\hat{y}_t = a + b_0 \cdot t + b_1 \cdot y_{t-1} + \dots b_{p-1} \cdot y_{t-(p-1)} + b_p \cdot y_{t-p}.$$

Ряд будет нестационарным, если коэффициент автокорреляции 1-го порядка $b_1 = 1$.
Для удобства дальнейших вычислений переходят к разностному уравнению:

$$\Delta y_t = b_0 + b_1 \cdot t + \varphi \cdot y_{t-1} + \sum_{i=1}^{p-1} \gamma_i \Delta y_{t-i}.$$

Тест ADF проверяет гипотезу о нестационарности ряда $(\varphi = 0)$ на основе построения регрессионного уравнения на основе LS и использования модифицированного T-критерия Стьюдента.

- Если $t < t_{critical}$, то гипотеза отвергается, ряд первых разностей признается стационарным, а исходный ряд является интегрируемым 1-го порядка.
- Если $t > t_{critical}$, то гипотеза не отвергается.

Проблемы при использовании теста ADF

- Принятие решения о включении в тестируемую модель константы, константы и тренда или их отсутствии.
  Подсказкой здесь может служить анализ графика исходного ряда (ряда первых или вторых разностей).
- Выбор включаемого числа лагов.
  Может осуществляться автоматически (на основе критерия Шварца), либо экспертно, указанием конкретного числа лагов.

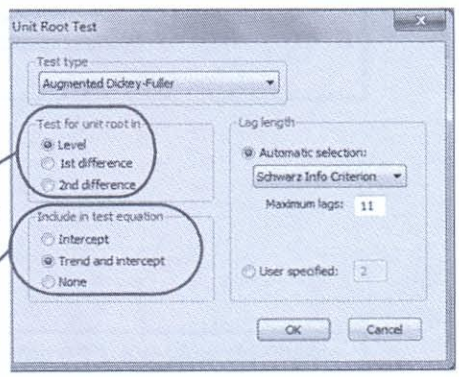
Замечания

- Остатки модели, анализируемой в тесте ADF, должны быть нормально распределены, не иметь автокорреляции и иметь постоянную дисперсию (быть гомоскедастичными).
- Остатки сохраняются в рабочем файле EViews в объекте **resid**.

Реализация в EViews

Исходный ряд, очищенный от сезонности, имеет тренд. Будем тестировать его на стационарность.

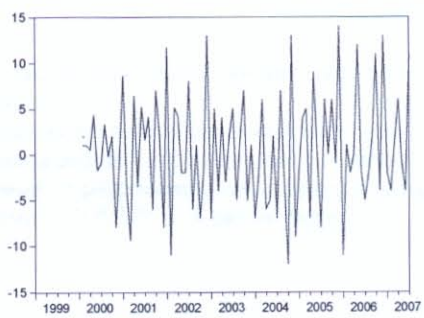
View – Unit Root Test...



Тестируется исходный ряд (Level)

Исходный ряд имеет тренд и константу

Если исходный ряд будет нестационарным, то будем тестировать ряд первых разностей, который не имеет ни тренда, ни константы.



Реализация в Eviews (продолжение)

Series: S_MEAT Workfile: ПРОДУКТЫ_ВЫПУСК::Untitled\

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph

Augmented Dickey-Fuller Unit Root Test on S_MEAT

Null Hypothesis: S_MEAT has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 2 (Automatic - based on SIC, maxlag=11)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.748194	0.9658
Test critical values:	1% level	-4.065702	
	5% level	-3.461686	
	10% level	-3.157121	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(S_MEAT)
Method: Least Squares
Date: 06/08/12 Time: 11:20
Sample (adjusted): 2000M04 2007M07
Included observations: 88 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
S_MEAT(-1)	-0.057429	0.076757	-0.748194	0.4565
D(S_MEAT(-1))	-0.583950	0.118990	-4.907565	0.0000
D(S_MEAT(-2))	-0.282203	0.110491	-2.554082	0.0125
C	-0.545355	1.442450	-0.378076	0.7063
@TREND(1999M01)	0.035831	0.025966	1.379915	0.1713

R-squared	0.304282	Mean dependent var	0.456818
Adjusted R-squared	0.270753	S.D. dependent var	6.220725
S.E. of regression	5.312248	Akaike info criterion	6.233048
Sum squared resid	2342.259	Schwarz criterion	6.373805
Log likelihood	-269.2541	Hannan-Quinn criter.	6.289755
F-statistic	9.075284	Durbin-Watson stat	1.960800
Prob(F-statistic)	0.000004		

Проверяемая гипотеза: «Исходный ряд нестационарен».

$t = -0.748 > t_{critical} = -3.157,$
значит гипотеза о нестационарности принимается. Это означает, что исходный ряд интегрируем более высокого порядка, либо неинтегрируем вообще.

Повторный тест ADF без включения тренда и константы для первых разностей (**Test for unit root in – 1st difference, Include in test equation – None**).

Series: D_S_MEAT Workfile: ПРОДУКТЫ_ВЫПУСК::Untitled\

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph

Augmented Dickey-Fuller Unit Root Test on D(D_S_MEAT)

Null Hypothesis: D(D_S_MEAT) has a unit root
Exogenous: None
Lag Length: 5 (Automatic - based on SIC, maxlag=11)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-8.706291	0.0000
Test critical values:	1% level	-2.593121	
	5% level	-1.944762	
	10% level	-1.614204	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(D_S_MEAT,2)
Method: Least Squares
Date: 06/08/12 Time: 11:31
Sample (adjusted): 2000M09 2007M07
Included observations: 83 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(D_S_MEAT(-1))	-7.791083	0.894880	-8.706291	0.0000
D(D_S_MEAT(-1),2)	5.276709	0.821293	6.424884	0.0000
D(D_S_MEAT(-2),2)	3.672508	0.675750	5.434710	0.0000
D(D_S_MEAT(-3),2)	2.314197	0.486494	4.756885	0.0000
D(D_S_MEAT(-4),2)	1.154896	0.283922	4.067646	0.0001
D(D_S_MEAT(-5),2)	0.350199	0.110303	3.174882	0.0022

R-squared	0.922760	Mean dependent var	0.128916
Adjusted R-squared	0.917744	S.D. dependent var	19.65644
S.E. of regression	5.637512	Akaike info criterion	6.366306
Sum squared resid	2447.179	Schwarz criterion	6.541162
Log likelihood	-258.2017	Hannan-Quinn criter.	6.436553
Durbin-Watson stat	2.040480		

Проверяемая гипотеза: «Ряд первых разностей нестационарен».

$t = -8.706 < t_{critical} = -1.945,$
значит гипотеза о нестационарности отклоняется, т.е. ряд первых разностей стационарен, а исходный ряд является интегрируемым 1-го порядка.

Таким образом:

- Если для прогнозирования будет применен метод ARIMA, то надо использовать порядок интегрирования $d = 1$.
- При использовании в дальнейшем методов, требующих стационарности, для приведения этого нестационарного ряда к стационарному виду, достаточно взять первую разность и использовать ее при моделировании.