Практическое занятие № 6.

«Однофакторные стохастические модели динамических процессов»

Процедуры анализа временных рядов и построения простейших прогнозов включают в себя очень широкий спектр проблем, возникающих в процессе эконометрического исследования. В рамках однофакторных стохастических моделей предполагается, что рассматриваемый процесс состоит из компонентов анализируемых временных рядов (если будущие значения рассматриваемой переменной сколько-нибудь предсказуемы, то они являются функцией от прошлых значений этой переменной). Проведем анализ составляющих однофакторного стохастического процесса в контексте авторегрессионного анализа, процессов со скользящей средней и степени интегрирования. Эти три подпроцесса в зарубежной научной литературе объединены под названием авторегрессионные интегрированные модели со скользящей средней $(ARIMA)$ [1].

Перед применением $ARIMA$ необходимо определить разности уровней с целью получения стационарного ряда, то есть нужно знать порядок этих разностей. Таким образом, процесс $ARIMA$ обладает тремя параметрами: p - порядок авторегрессии, d - требуемый порядок предварительно определяемых разностей и q - порядок скользящей средней в модели.

Понятие интегрирования означает, какого порядка разности должны быть рассчитаны для того, чтобы получить стационарный временной ряд. Нахождение разностей - это всего лишь нахождение изменений значения переменной в последующие периоды, т.е. нахождение величины $\Delta Y_t = Y_t - Y_{t-1}$. Ряд значений ΔY_t - это ряд разностей.

Если во временном ряду должны быть рассчитаны первые разности, чтобы получить стационарный ряд, то первоначальный ряд называется интегрированным рядом первого порядка, или $I(1)$. Если же требуется рассчитать вторые разности для получения стационарного ряда, то это интегрированный ряд второго порядка, или $I(2)$. Если же в ряду вообще не требуется вычислять разности, то он называется интегрированным рядом нулевого порядка, или $I(0)$.

Если ряд $I(0)$, т.е. стационарен, то его дисперсия будет конечна. Изменения рассматриваемой переменной будут иметь только промежуточное влияние на временной ряд. Коэффициенты автокорреляции будут постепенно убывать таким образом, что их сумма станет конечной. И наоборот, если ряд $I(1)$, то изменения будут

[1] Уотшем Т.Дж., Паррамоу К. Количественные методы в финансах: Учебное пособие для вузов /Пер. с англ. под ред. М.Р.Ефимовой. – М.: Финансы, ЮНИТИ, 1999. – 527 с.

иметь постоянный эффект, дисперсия с течением времени будет возрастать до бесконечности .

Проведя расчет разностей, можно переходить к моделированию полученного стационарного ряда (ряда, обладающего постоянной средней и дисперсией, ковариация в котором зависит только от временного интервала между двумя отдельными наблюдениями) с помощью $ARMA$.

Начнем анализ $ARIMA$ с рассмотрения авторегрессионного процесса**.** Авторегрессионным называется процесс, при котором значение ряда находится в линейной зависимости от предыдущих значений. Например, если текущее наблюдаемое значение является функцией всего лишь одного значения, непосредственно предшествующего наблюдению, т.е. процесс зависит всего лишь от одного значения рассматриваемой переменной, то процесс называется авторегрессионным процессом первого порядка и обозначается $AR(1)$. Это можно обобщить следующим образом: если анализируемый динамический процесс зависит от значений, отстоящих от 1 до n временных лагов назад, то это авторегрессионный процесс порядка n, т.е. $AR(n)$. Например, процесс $AR(3)$ можно отобразить следующим образом:

$$\hat{Y}_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \alpha_3 Y_{t-3} + \varepsilon_t . \qquad (6.1)$$

Из формулы (6.1) следует, что текущее значение Y - это функция от трех наиболее недавних предыдущих значений; ε_t - остаток или ошибка (погрешность).

Модель скользящей средней (MA) [2]- это модель, где моделируемая величина задается линейной функцией от прошлых ошибок, то есть разностей между прошлыми смоделированными значениями и прошлыми фактическими наблюдениями:

$$\hat{Y}_t = B_0 + B_1 \varepsilon_{t-1} + B_2 \varepsilon_{t-2} + B_3 \varepsilon_{t-3} + \varepsilon_t , \qquad (6.2)$$

где $\varepsilon_t = Y_t - \hat{Y}_t$.

Используемый термин «скользящая средняя» не следует путать со схожим термином, относящимся к технике сглаживания данных.

Таким образом, модель $ARMA(pq)$ имеет p временных лагов в авторегрессионном процессе и q интервалов в модели скользящей средней. Например, $ARMA(3,2)$ будет иметь следующий вид:

$$\hat{Y}_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \alpha_3 Y_{t-3} + B_1 \varepsilon_{t-1} + B_2 \varepsilon_{t-2} + B_3 \varepsilon_{t-3} + u_t , \qquad (6.3)$$

где u_t - остаточный член ошибки в данном уравнении.

[2] Используемый термин «скользящая средняя» не следует путать со схожим термином, относящимся к технике сглаживания данных.

Для определения степени автокорреляции временных рядов необходимо определить силу связи между текущими и прошлыми значениями рассматриваемой переменной. Одним из способов измерения этой связи являются коэффициенты автокорреляции (ACC), совокупность которых образует функции автокорреляции (ACF). Коэффициент автокорреляции измеряет связь между текущими и прошлыми наблюдениями временного ряда и рассчитывается следующим образом:

$$\rho_k = \frac{\sum_{t=1}^{n-k}\left(Y_t - \overline{Y}\right)\bullet\left(Y_{t+k} - \overline{Y}\right)}{\sum_{t=1}^{n}\left(Y_t - \overline{Y}\right)^2}, \qquad (6.4)$$

где k - количество лагов.

Таким образом, коэффициент автокорреляции первого порядка будет рассчитан с лагом в один период, коэффициент автокорреляции второго порядка будет учитывать степень связи между значениями, отстоящими на два временных периода, и т.д. Рассчитываются коэффициенты автокорреляции всех порядков и затем проводится статистическая проверка для определения, при каких лагах коэффициенты статистически значимы. Только лаги, являющиеся статистически значимыми, оставляются в модели. Иногда проверка значимости коэффициентов автокорреляции проводится при помощи критерия стандартной ошибки и Q-критерия Бокса-Пирса[3]. Два критерия предлагаются потому, что существуют два подхода к проверке наличия автокорреляции. При первом подходе, то есть при использовании критерия стандартной ошибки, проверяются коэффициенты автокорреляции каждого порядка отдельно, чтобы выявить, какие из них значимы. Второй подход использует Q-критерий Бокса-Пирса для того, чтобы проверить на значимость все множество коэффициентов как группу. В настоящее время, с развитием компьютерных программ и средств, проверку удобнее проводить по методу P-value (методом определения уровня вероятности, то есть получением p-значения, которое в случае верности нулевой гипотезы представляет собой вероятность получения величины стандартизированного критерия проверки, большего по абсолютному значению, чем рассчитанный критерий проверки).

Частный коэффициент автокорреляции (PAC), лежащий в основе частной функции автокорреляции (PAF), измеряет связь между текущим значением переменной X_t и последующими значениями этой переменной $X_{t-1}, X_{t-2},..., X_{t-k}$, когда влияние всех промежуточных временных лагов устранено. Таким образом, частный коэффициент автокорреляции первого порядка будет равен коэффициенту автокорреляции первого порядка, так как нет промежуточных лагов. Но частные коэффициенты второго и следующих порядков будут уже отличаться друг от друга.

Частный коэффициент автокорреляции используется для определения степени автокорреляции внутри временного ряда. Например, ряд, обозначенный $AR(m)$, показывает, что последний статистически значимый частный коэффициент автокорреляции рассчитан с лагом m. Таким образом, в ряде $AR(2)$ текущее значение переменной обладает значимой корреляцией только со значениями, отстоящими на 1 и 2 временных лага назад. В ряде $AR(4)$ значимыми будут частные коэффициенты автокорреляции с лагами от одного до четырех периодов, но коэффициенты с более высокими лагами не будут значимо отличаться от нуля.

В динамическом процессе $AR(m)$ частные коэффициенты автокорреляции значимо отличаются от нуля для временных лагов от 1 до m и затем резко падают до нуля для интервалов $m+1$ и больше.

Зная поведение коэффициента автокорреляции и частного коэффициента автокорреляции, можно попытаться определить, содержит ли ряд элемент скользящей средней. Если ряд скорее AR чем MA, то автокорреляция не будет показывать порядок MA-процесса. Хотя, если значение частных коэффициентов автокорреляции падает по экспоненте, а не опускается резко до нуля, то можно предположить, что ряд содержит процесс скользящей средней, а не AR.

Для проверки автокорреляции в рядах, где присутствуют элементы и авторегрессии и скользящей средней, используется критерий Люнга-Бокса (LB) (Ljung-Box)[4]. Критерий LB рассчитывается следующим образом:

$$LB = n(n+2)\sum_{k=1}^{m}\left(\frac{1}{n-k}\right)\bullet r_k^2 \approx \chi^2_{m-p-q}, \qquad (6.5)$$

где m - максимальное число временных лагов, рассматриваемых в модели;
p - порядок авторегрессии;
q - порядок процесса скользящей средней.

Как уже отмечалось выше, интеграция означает, в какой степени ряд должен быть преобразован с помощью разностей различного порядка, чтобы стать стационарным. Это очень важно, так как многие методы анализа временных рядов подразумевают, что анализируемый ряд в действительности является стационарным. Провер-

[3] Box, G.E.P. and Pierce, D.A. «Distribution of Residual Autocorrelations in Autoregressive Moving Average Time Series Models». – Journal of the American Statistical Association, 1970, № 65. – P. 1509-1526.

[4] Ljung, G.M. and Box, G.E.P. «On a Measure of Lack Of Fit in Time Series Models». – Biometrika, 1978, № 66. – P. 67-72.

Этот текст доступен на постоянно обновляющемся сайте http://www.molchanov.narod.ru/econometrics.html

ка стационарности производится при помощи теста единичного корня (unit root). Если говорить более строго, то проверка стационарности производится на основе анализа корней характеристического уравнения (единичный корень соответствует границе области стационарности). Если данные показывают единичный корень, то ряд является $I(1)$.

Подход к проверке стационарности и степени интеграции называют критерием Дики-Фуллера (DF)[5]. С помощью этого критерия проверяется, имеет ли коэффициент α в уравнении (1) значение, равное единице или меньше единицы:

$$Y_t = \alpha_1 Y_{t-1} + \varepsilon_t. \qquad (6.6)$$

Если α равно единице, то данные имеют единичный корень и степень интегрирования равна 1, т.е. ряд является $I(1)$. Если же α меньше единицы и больше нуля, то ряд стационарен, т.е. $I(0)$. Обычно α бывает не больше 1, поскольку это подразумевает взрывные ряды. Такие ряды маловероятны, поскольку давление экономической среды не позволяет переменной принимать бесконечно большие значения. Существуют некоторые теоретические проблемы с уравнением (6.6), поскольку возможность нестационарности нарушает допущения регрессии МНК, которая подразумевает постоянную дисперсию остатков.

Тогда требуется уравнение, выражающее изменения Y_t следующим образом:

$$\Delta Y_t = \beta Y_{t-1} + e_t, \qquad (6.7)$$

где $\beta = (\alpha_t - 1)$.

Если β равно нулю, то говорят, что ряд Y обладает единичным корнем и является $I(1)$, и ряд ΔY будет стационарным. Если же β меньше нуля, то есть α меньше единицы, то сам ряд Y является стационарным, $I(0)$.

Проблемы при проверке стационарности, когда существует автокорреляция остатков, решаются применением расширенного критерия Дики-Фуллера (DF). При использовании этого метода прошлые значения независимой переменной включаются в уравнение регрессии с лагом, достаточным для того, чтобы избавиться от автокорреляции остатков. Это уравнение может иметь вид:

$$\Delta Y_t = \alpha_0 + \beta_1 Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \ldots + \gamma_n \Delta Y_{t-n} + e_t. \qquad (6.8)$$

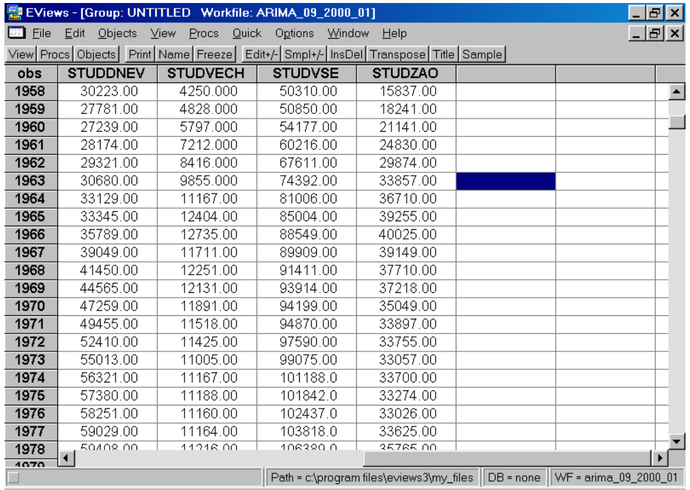
Все вышеизложенное проиллюстрируем примером конкретного расчета.

Пример 6. Имеются данные о численности студентов в вузах Ростовской области (1958-1998 гг.) (файл **example_06.xls**.). В созданном файле изначально присутствуют четыре переменные: STUDVSE – общая численность студентов (чел.);

[5] Dickey, D.A. and Fuller, W.A. «Distribution of the Estimators for Autoregressive Time Series with a Unit Root». – Journal of the American Statistical Association, 1979, № 74. – P. 427-431.

Этот текст доступен на постоянно обновляющемся сайте http://www.molchanov.narod.ru/econometrics.html

STUDDNEV – количество студентов на дневной форме обучения (чел.); STUDVECH – количество студентов на вечерней форме обучения (чел.); STUDZAO – количество студентов на заочной форме обучения (чел.). Фрагмент окна с данными представим на рис. 73.



EViews - [Group: UNTITLED Workfile: ARIMA_09_2000_01]

obs	STUDDNEV	STUDVECH	STUDVSE	STUDZAO
1958	30223.00	4250.000	50310.00	15837.00
1959	27781.00	4828.000	50850.00	18241.00
1960	27239.00	5797.000	54177.00	21141.00
1961	28174.00	7212.000	60216.00	24830.00
1962	29321.00	8416.000	67611.00	29874.00
1963	30680.00	9855.000	74392.00	33857.00
1964	33129.00	11167.00	81006.00	36710.00
1965	33345.00	12404.00	85004.00	39255.00
1966	35789.00	12735.00	88549.00	40025.00
1967	39049.00	11711.00	89909.00	39149.00
1968	41450.00	12251.00	91411.00	37710.00
1969	44565.00	12131.00	93914.00	37218.00
1970	47259.00	11891.00	94199.00	35049.00
1971	49455.00	11518.00	94870.00	33897.00
1972	52410.00	11425.00	97590.00	33755.00
1973	55013.00	11005.00	99075.00	33057.00
1974	56321.00	11167.00	101188.0	33700.00
1975	57380.00	11188.00	101842.0	33274.00
1976	58251.00	11160.00	102437.0	33026.00
1977	59029.00	11164.00	103818.0	33625.00
1978	59408.00	11216.00	106389.0	35765.00

Path = c:\program files\eviews3\my_files | DB = none | WF = arima_09_2000_01

Рис. 73. Исходные данные

Учитывая иллюстрационный характер проводимых вычислений, дальнейшие расчеты будем производить с использованием данных о численности студентов по дневной форме обучения (STUDDNEV). Представим исходные цифры в виде линейного графика (рис. 74).

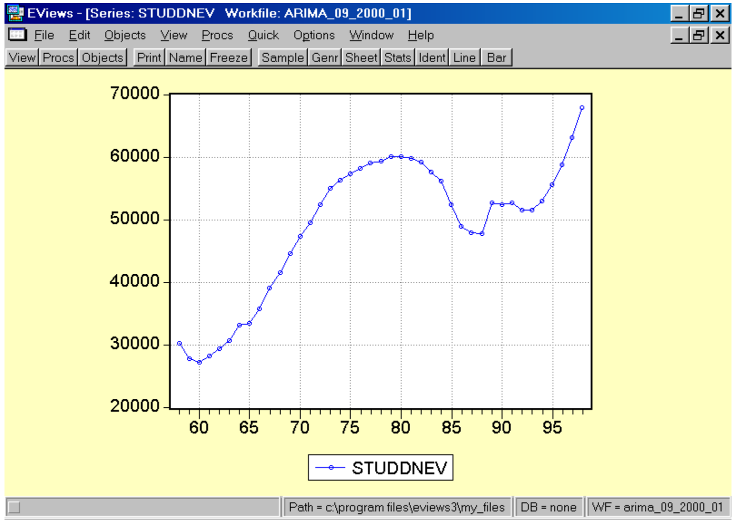


Рис. 74. Динамика численности студентов дневной формы обучения

Для приведения ряда к стационарному найдем первые разности, введя переменную DSTUDDNEV. Визуализируем полученные данные (рис. 75)

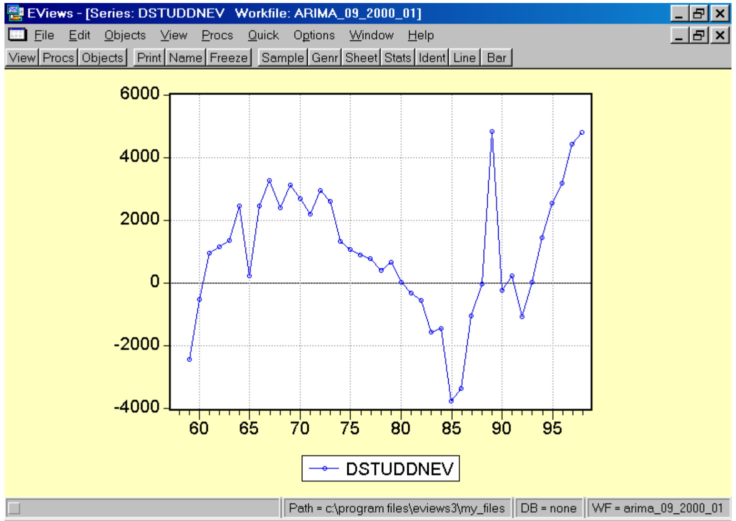


Рис. 75. Первые разности переменной STUDDNEV

Стационарность данной переменной проверим с помощью критерия DF (рис. 76).

ADF Test Statistic	-1.756750	1% Critical Value*	-3.6117
		5% Critical Value	-2.9399
		10% Critical Value	-2.6080

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DSTUDDNEV)
Method: Least Squares
Date: 09/17/00 Time: 14:13
Sample(adjusted): 1961 1998
Included observations: 38 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DSTUDDNEV(-1)	-0.246303	0.140204	-1.756750	0.0877
D(DSTUDDNEV(-1))	-0.103576	0.170774	-0.606508	0.5481
C	388.4561	269.8087	1.439746	0.1588

R-squared	0.131069	Mean dependent var	140.5526
Adjusted R-squared	0.081416	S.D. dependent var	1551.744
S.E. of regression	1487.234	Akaike info criterion	17.52288
Sum squared resid	77415306	Schwarz criterion	17.65216
Log likelihood	-329.9347	F-statistic	2.639700
Durbin-Watson stat	1.945933	Prob F-statistic)	0.085554

Рис. 76. Проверка переменной DSTUDDNEV с помощью критерия DF

Как видно из рис. 75 и рис. 76, переменная DSTUDDNEV не является стационарной.

В файле данных создадим новую серию D2STUDDNEV, как первые разности переменной DSTUDDNEV (см. рис. 77).

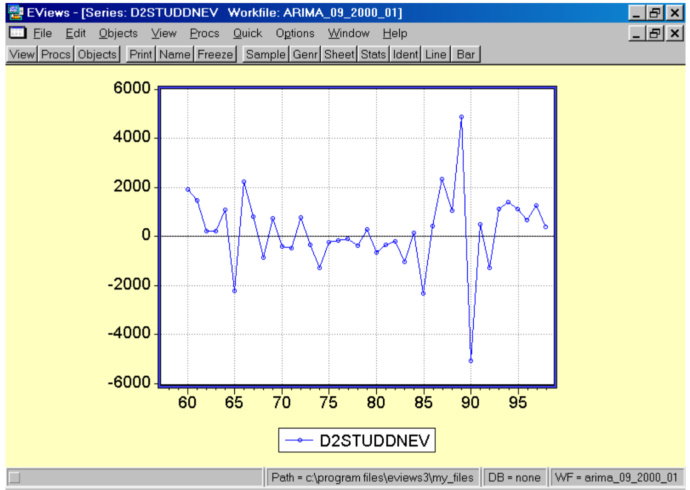


Рис. 77. Первые разности переменной DSTUDDNEV

Даже на глаз видно, что новый ряд является стационарным. Подтвердим наше предположение с помощью расчета критерия DF (рис. 78).

ADF Test Statistic	-4.359137	1% Critical Value*	-3.6171
		5% Critical Value	-2.9422
		10% Critical Value	-2.6092

*MacKinnon critical values for rejection of hypothesis of a unit root.

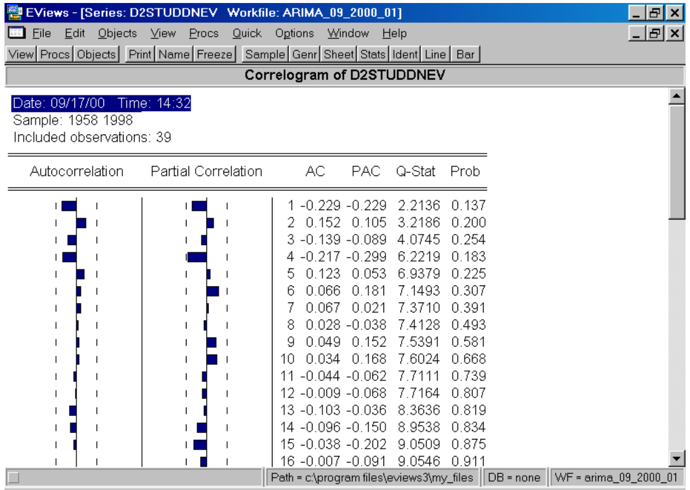
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(D2STUDDNEV)
Method: Least Squares
Date: 09/17/00 Time: 14:19
Sample(adjusted): 1962 1998
Included observations: 37 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D2STUDDNEV(-1)	-1.143815	0.262395	-4.359137	0.0001
D(D2STUDDNEV(-1))	-0.096798	0.165913	-0.583428	0.5635
C	122.0629	255.3559	0.478011	0.6357

R-squared	0.640568	Mean dependent var	-29.89189
Adjusted R-squared	0.619425	S.D. dependent var	2490.451
S.E. of regression	1536.378	Akaike info criterion	17.58985
Sum squared resid	80255511	Schwarz criterion	17.72046
Log likelihood	-322.4122	F-statistic	30.29691
Durbin-Watson stat	1.989844	Prob(F-statistic)	0.000000

Рис. 78. Проверка переменной D2STUDDNEV с помощью критерия DF

Теперь есть все основания отвергнуть нулевую гипотезу о наличии единичного корня (AFF t-статистика больше критических значений критерия на всех трех уровнях значимости). Другими словами, ряд D2STUDDNEV является стационарным.


EViews - [Series: D2STUDDNEV Workfile: ARIMA_09_2000_01]
File Edit Objects View Procs Quick Options Window Help
View Procs Objects Print Name Freeze Sample Genr Sheet Stats Ident Line Bar
Correlogram of D2STUDDNEV

Date: 09/17/00 Time: 14:32
Sample: 1958 1998
Included observations: 39

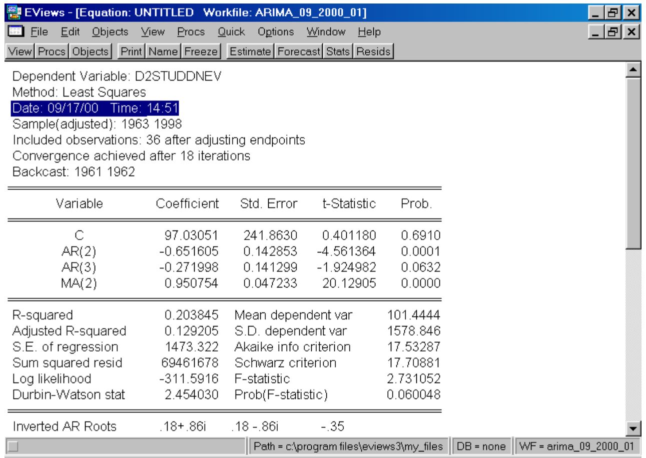
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.229	-0.229	2.2136	0.137
		2	0.152	0.105	3.2186	0.200
		3	-0.139	-0.089	4.0745	0.254
		4	-0.217	-0.299	6.2219	0.183
		5	0.123	0.053	6.9379	0.225
		6	0.066	0.181	7.1493	0.307
		7	0.067	0.021	7.3710	0.391
		8	0.028	-0.038	7.4128	0.493
		9	0.049	0.152	7.5391	0.581
		10	0.034	0.168	7.6024	0.668
		11	-0.044	-0.062	7.7111	0.739
		12	-0.009	-0.068	7.7164	0.807
		13	-0.103	-0.036	8.3636	0.819
		14	-0.096	-0.150	8.9538	0.834
		15	-0.038	-0.202	9.0509	0.875
		16	-0.007	-0.091	9.0546	0.911

Path = c:\program files\eviews3\my_files DB = none WF = arima_09_2000_01

Рис. 79. Коррелограмма переменной D2STUDDNEV

Перед определением порядка авторегрессионного процесса полезным будет рассмотрение коррелограммы (рис. 79). Можно предположить наличие авторегрессии максимального порядка 3.

Проведем оценивание авторегрессионной модели скользящей средней $(ARMA)$. Для этого в командном окне пакета необходимо набрать: ***LS D2STUDDNEV C AR(2) AR(3) MA(2)***. То есть, мы подразумеваем модель вида $ARMA(3,2)$ с исключенными периодами ***ar(1)*** и ***ma(1)***. Итоговый расчет представлен на рис. 80.


EViews - [Equation: UNTITLED Workfile: ARIMA_09_2000_01]
File Edit Objects View Procs Quick Options Window Help
View Procs Objects Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: D2STUDDNEV
Method: Least Squares
Date: 09/17/00 Time: 14:51
Sample(adjusted): 1963 1998
Included observations: 36 after adjusting endpoints
Convergence achieved after 18 iterations
Backcast: 1961 1962

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	97.03051	241.8630	0.401180	0.6910
AR(2)	-0.651605	0.142853	-4.561364	0.0001
AR(3)	-0.271998	0.141299	-1.924982	0.0632
MA(2)	0.950754	0.047233	20.12905	0.0000

R-squared	0.203845	Mean dependent var	101.4444
Adjusted R-squared	0.129205	S.D. dependent var	1578.846
S.E. of regression	1473.322	Akaike info criterion	17.53287
Sum squared resid	69461678	Schwarz criterion	17.70881
Log likelihood	-311.5916	F-statistic	2.731052
Durbin-Watson stat	2.454030	Prob(F-statistic)	0.060048

Inverted AR Roots .18+.86i .18 -.86i -.35

Path = c:\program files\eviews3\my_files DB = none WF = arima_09_2000_01

Рис. 80. Результаты расчетов по $ARMA(3,2)$.

Этот текст доступен на постоянно обновляющемся сайте http://www.molchanov.narod.ru/econometrics.html

Получены довольно значимые наблюдаемые статистики t-критериев, которые могут быть приняты на уровне значимости 6,5%.

График остатков такой модели представлен на рис. 81.

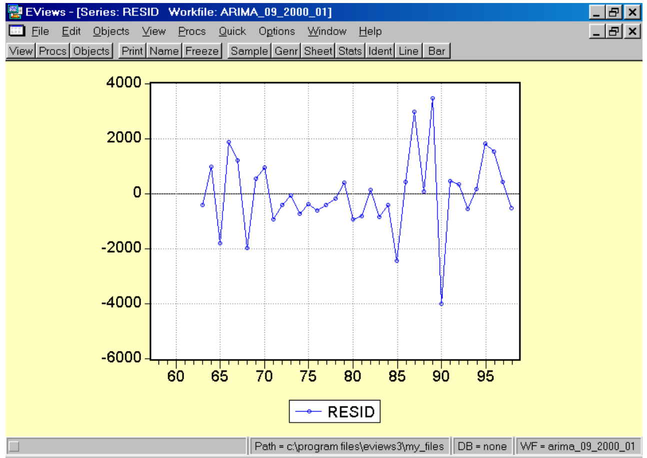


Рис. 81. Остатки уравнения $ARMA(3,2)$.

Анализ рис. 81 позволяет сделать предположение о его стационарности.

СОДЕРЖАНИЕ