9. AUTOCORRELATION

[1] Definition of Autocorrelation (AUTO)

1) Model: $y_t = x_{t\bullet}'\beta + \varepsilon_t$.

- We say that AUTO exists if $\text{cov}(\varepsilon_t,\varepsilon_s) \neq 0$, $t \neq s$.

2) Assumptions:

- All of SIC except SIC.3 (the random sample assumption).
- Or, all of WIC except WIC.2 [$E(\varepsilon_t|\varepsilon_1,...,\varepsilon_{t-1},x_{1\bullet}, ... ,x_{t\bullet}) = 0$].

  Replace WIC.2 by $E(\varepsilon_t| x_{1\bullet}, ... ,x_{T\bullet}) = 0$ (strictly exogenous regressors).

Comments:

For OLS, the assumption of weakly exogenous regressors is enough. However, for GLS, we need the assumption of strictly exogenous regressors (that is, $E(\varepsilon_t \mid x_{T\bullet}, x_{T-1,\bullet}, ..., x_{1\bullet}) = 0$ for all t).

3) Example:

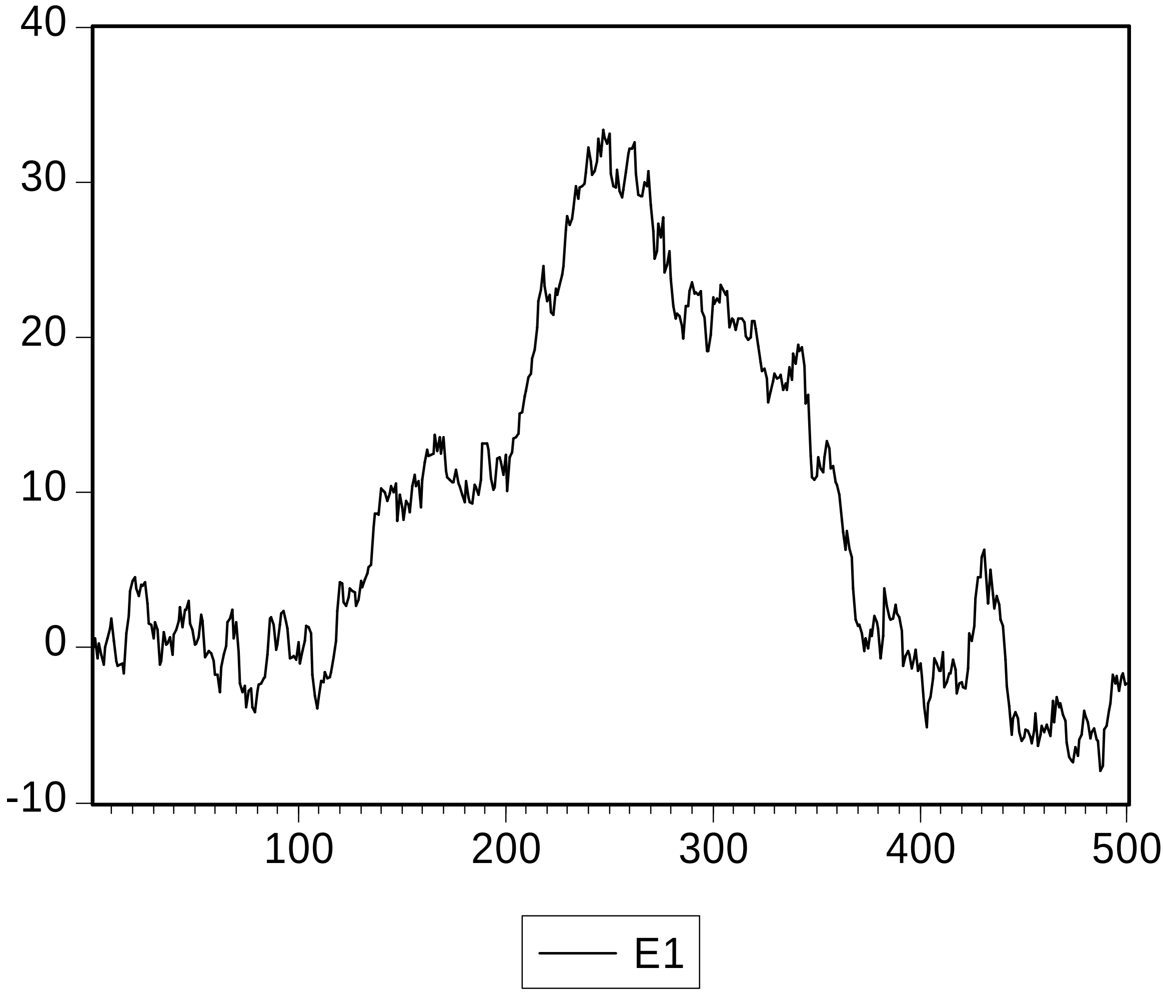
- AR(1): $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$, v_t iid $N(0,\sigma_v^2)$.

  $\rightarrow cov(\varepsilon_t,\varepsilon_{t-1}) = E(\varepsilon_t\varepsilon_{t-1}) = E[(\rho\varepsilon_{t-1}+v_t)\varepsilon_{t-1}] = \rho E(\varepsilon_{t-1}^2) \neq 0.$

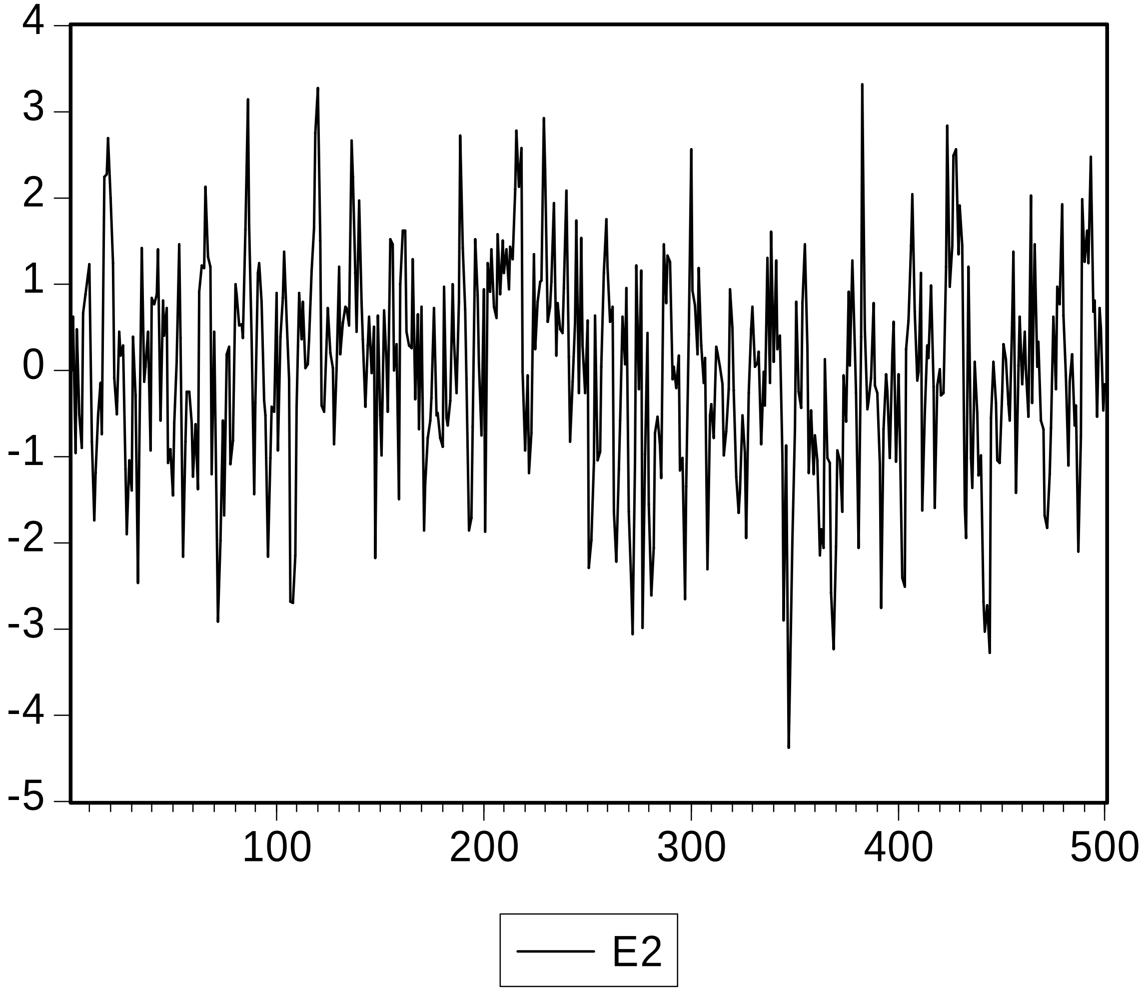
- MA(1): $\varepsilon_t = v_t + \psi v_{t-1}$, v_t iid $N(0,\sigma_v^2)$.

  $\rightarrow cov(\varepsilon_t,\varepsilon_{t-1}) = E(\varepsilon_t\varepsilon_{t-1}) = E[(v_t+\psi v_{t-1})(v_{t-1}+\psi v_{t-2})]$

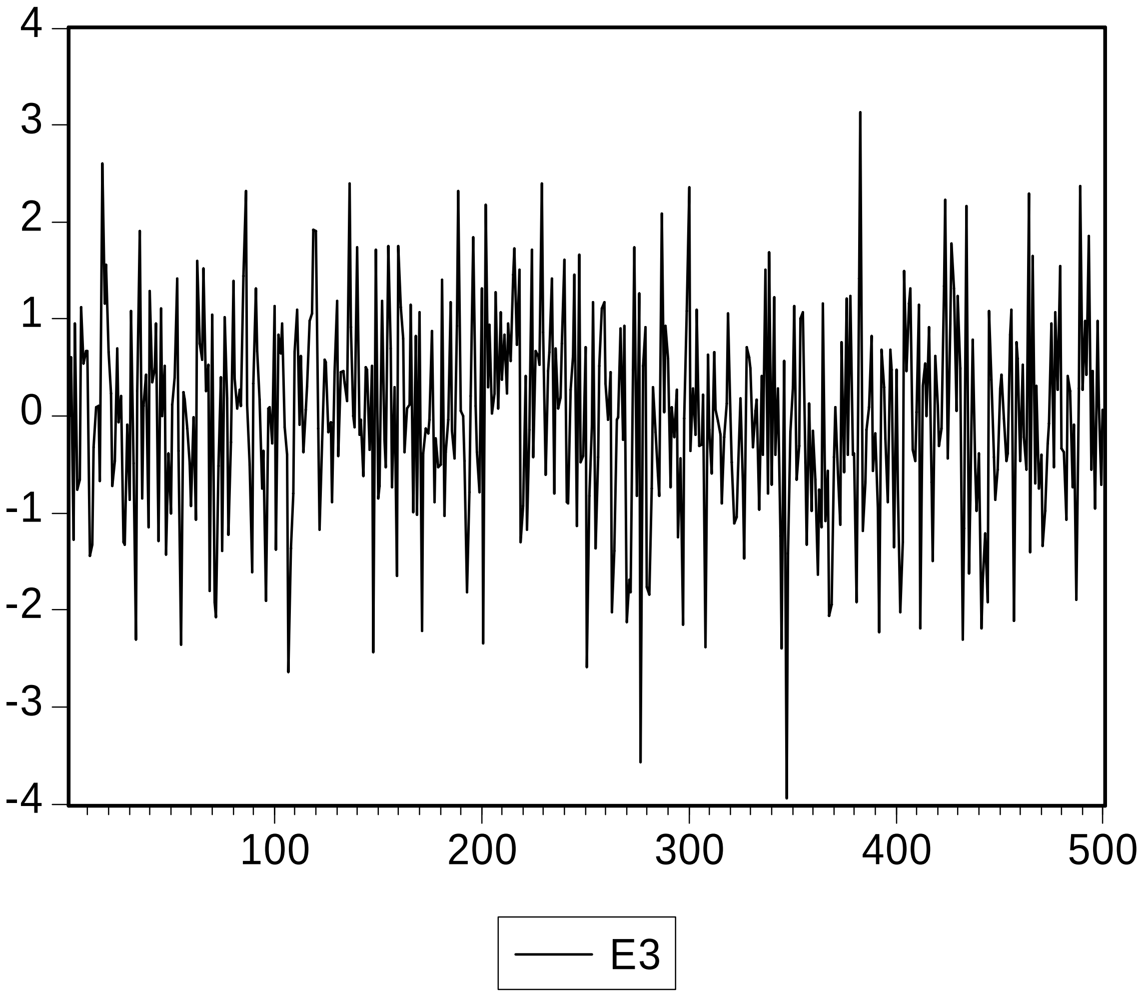
  $= \psi E(v_{t-1}^2) = \psi\sigma_v^2 \neq 0.$



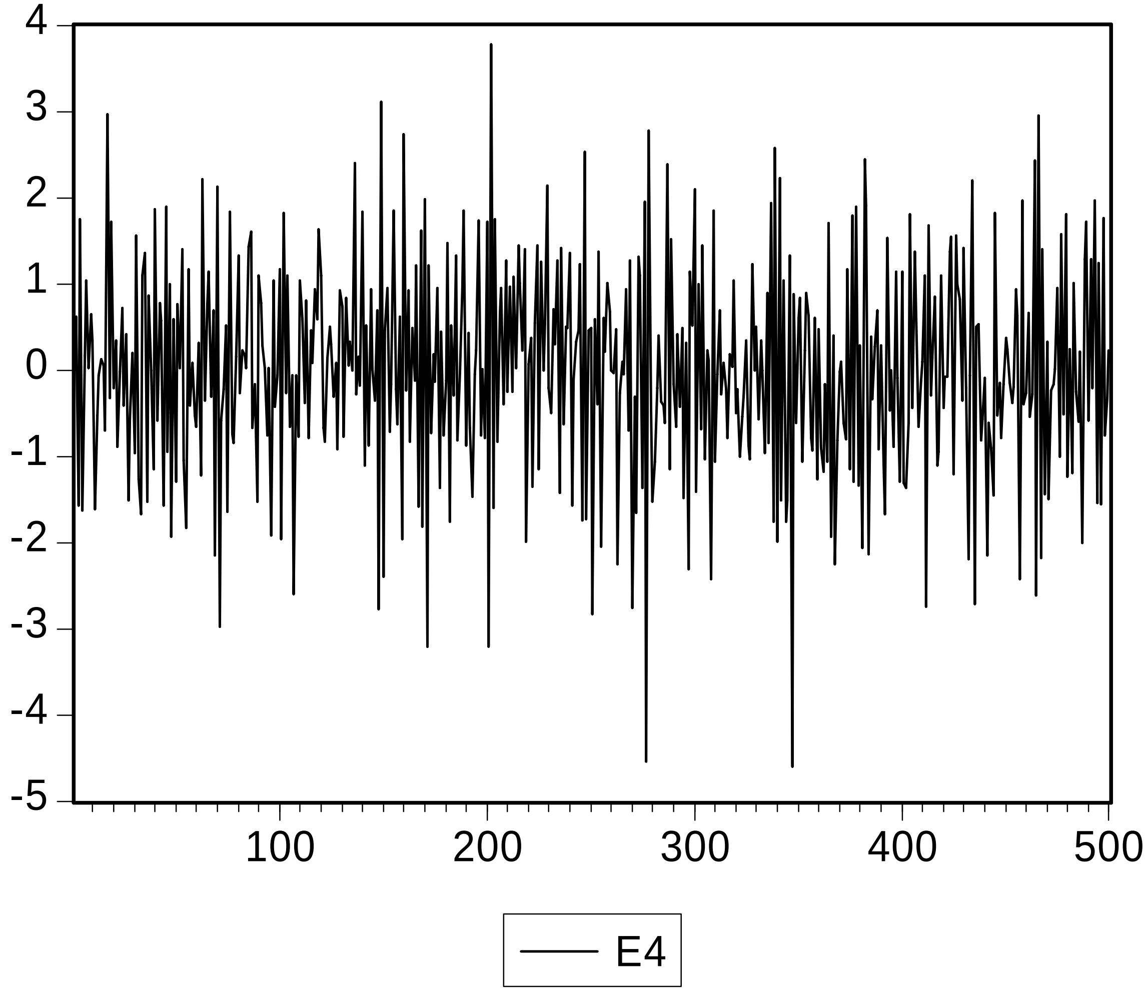
$\varepsilon_t = \varepsilon_{t-1} + v_t$, $\sigma_v^2 = 1$.



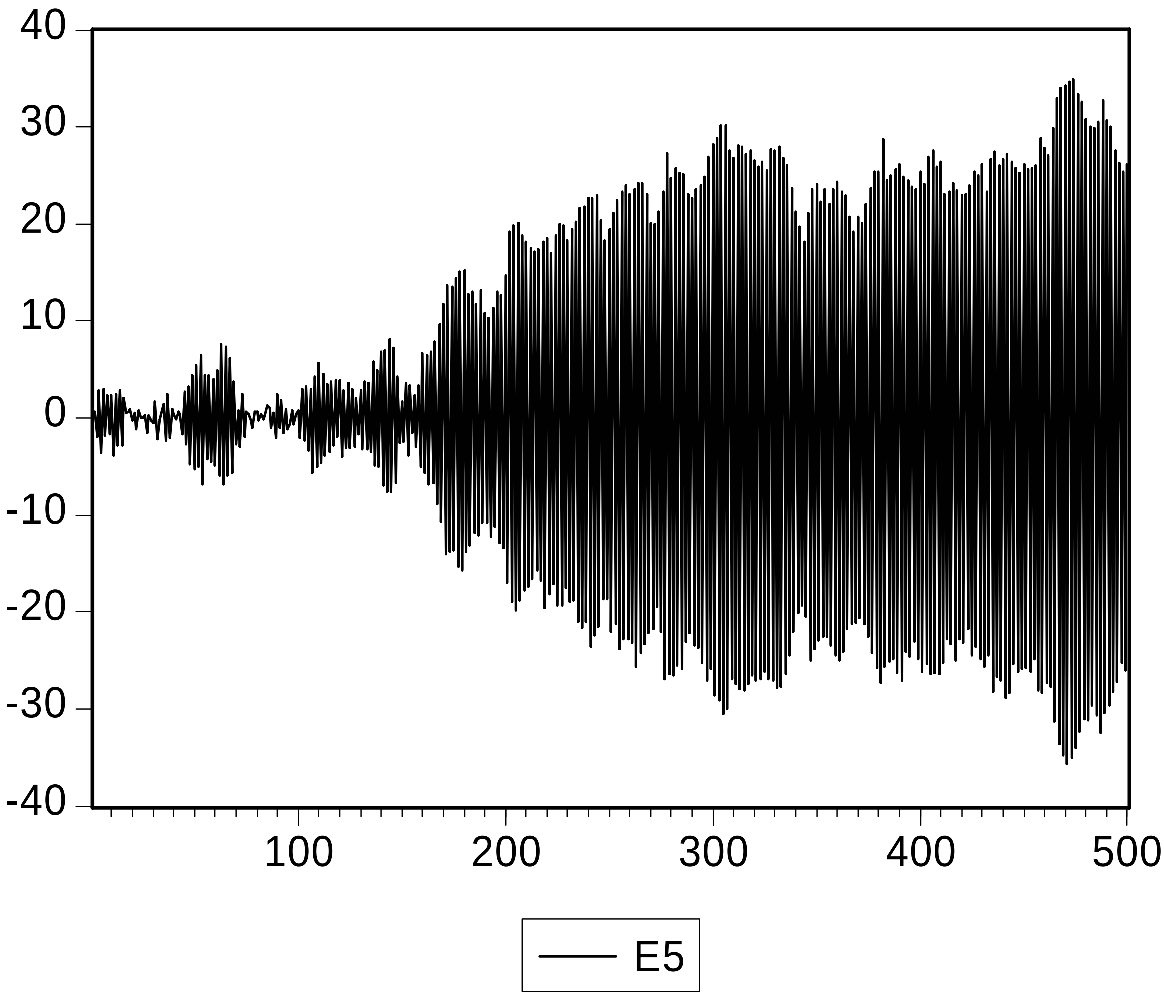
$\varepsilon_t = 0.5\varepsilon_{t-1} + \nu_t.$



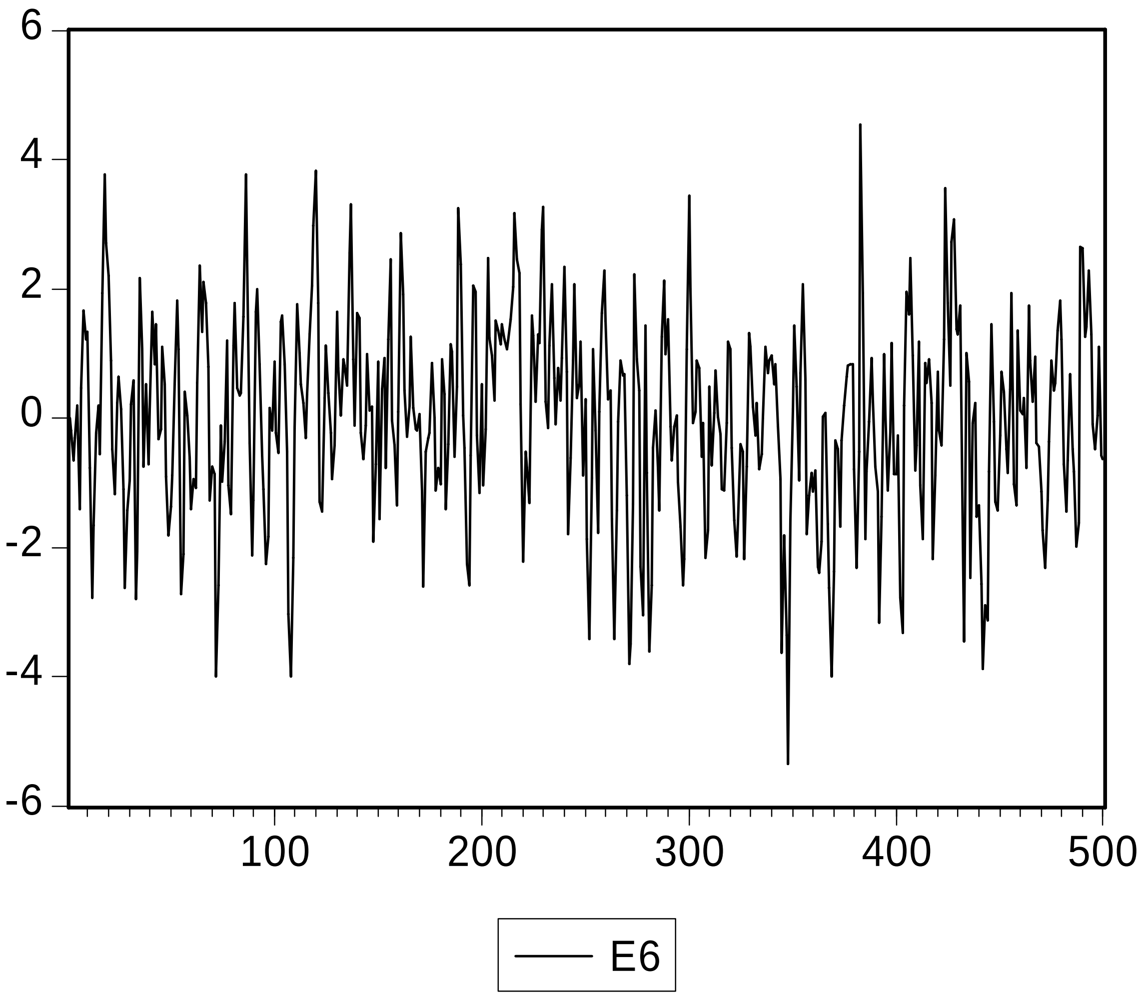
$\varepsilon_t = \nu_t$



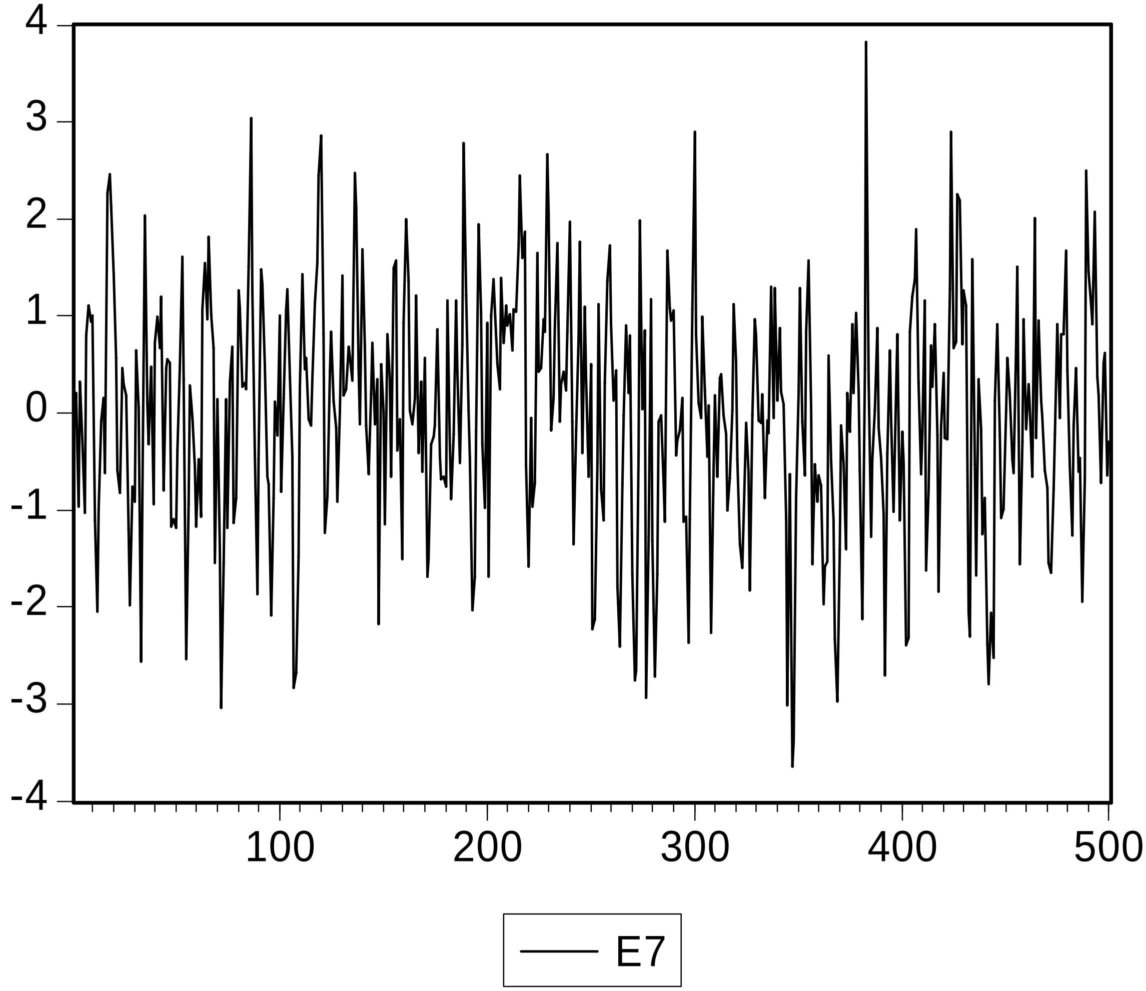
$$\varepsilon_t = -0.5\varepsilon_{t-1} + \nu_t$$



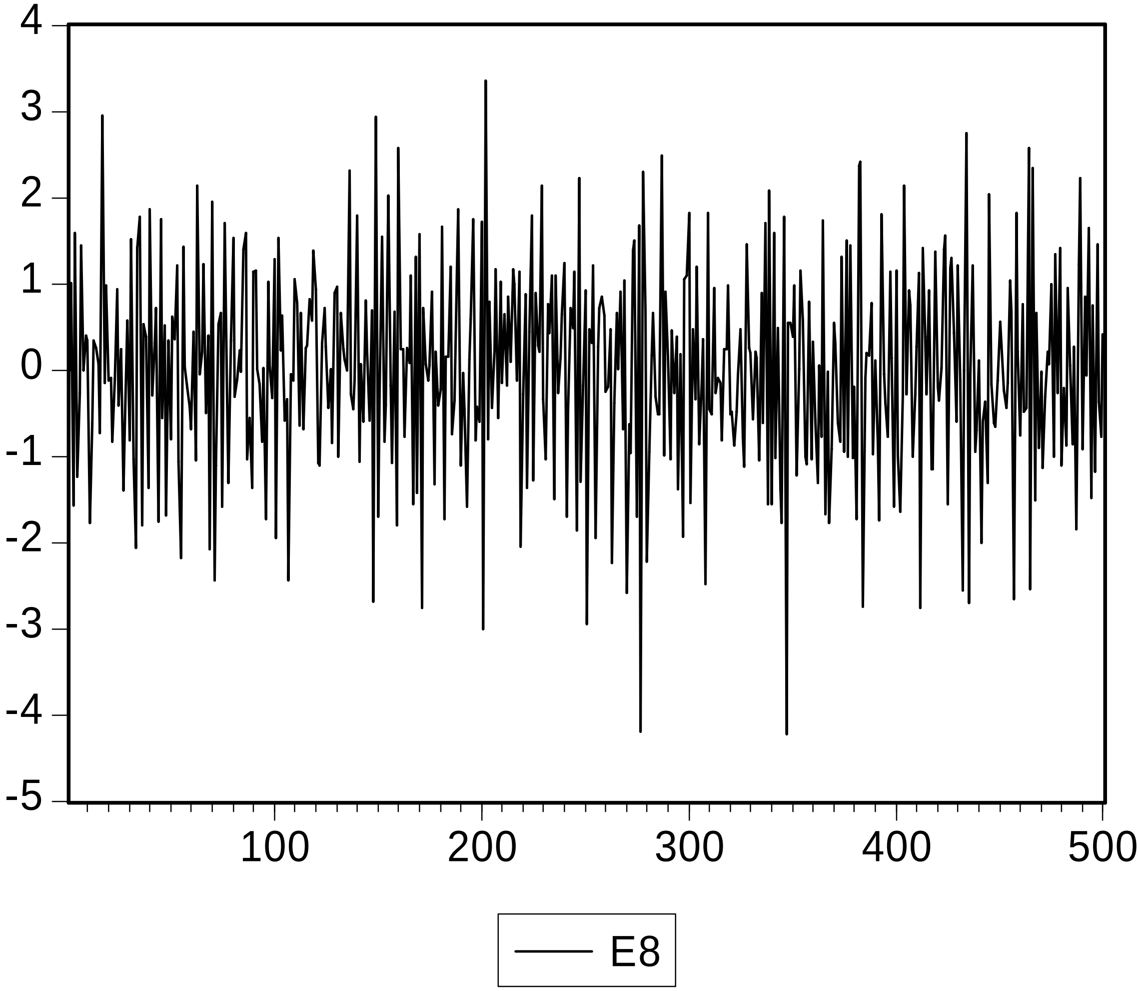
$$\varepsilon_t = -\varepsilon_t + \nu_t$$



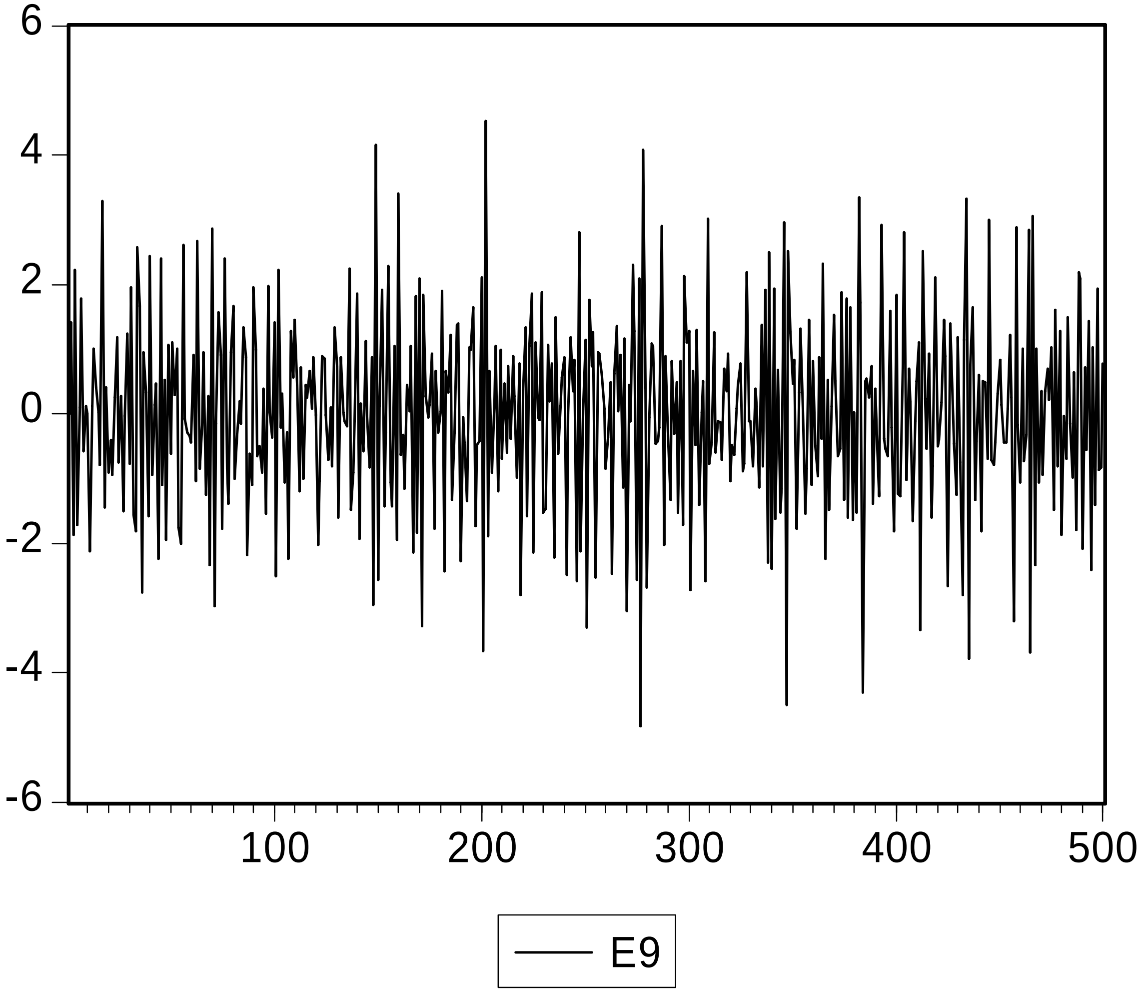
$\varepsilon_t = v_t + v_{t-1}$



$\varepsilon_t = \nu_t + 0.5\nu_{t-1}$



$\varepsilon_t = \nu_t - 0.5\nu_{t-1}$.



$\varepsilon_t = \nu_t - \nu_{t-1}$.

4) Typical in time-series data:

   i) Presence of unobservable variables.

   ii) Seasonally adjusted data.

Empirical Example: Expectations-Augmented Phillips curve:

- Data: (Table F5.1 – from Greene)
  - From 1950I to 2000II.
  - For data descriptions, see p. 948.
  - INFL: Annualized inflation rate (%).
  - UNEMP: Unemployment rate (%).

- Expectation-Augmented Phillips curve:

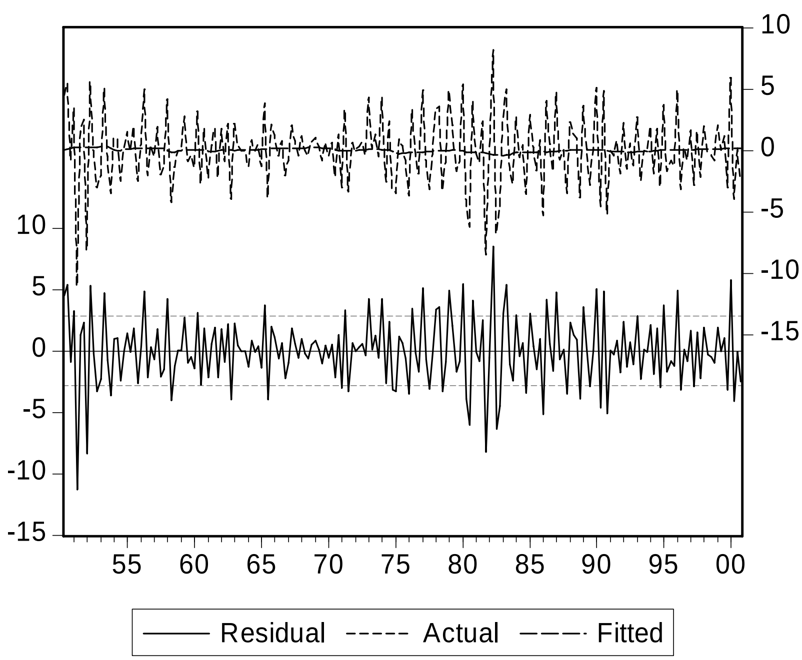
$$INFL_t - INFL_{t-1} = \beta_1 + \beta_2 UNEMP_t + \varepsilon_t,$$

where $\beta_1 = -\beta_2 u^*$ and u^* is the natural unemployment rate.

Dependent Variable: INFL-INFL(-1)
Method: Least Squares
Date: 04/22/03 Time: 17:03
Sample(adjusted): 1950:2 2000:4
Included observations: 203 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.517770	0.743156	0.696717	0.4868
UNEMP	-0.090768	0.126270	-0.718842	0.4731

R-squared	0.002564	Mean dependent var	0.003028
Adjusted R-squared	-0.002398	S.D. dependent var	2.829120
S.E. of regression	2.832510	Akaike info criterion	4.930007
Sum squared resid	1612.646	Schwarz criterion	4.962649
Log likelihood	-498.3957	F-statistic	0.516733
Durbin-Watson stat	2.792357	Prob(F-statistic)	0.473073



[2] Stationarity Assumption on ε_t

(1) Stationarity and Ergodicity

- Let $\{s_t\}$ be a process of interest.
- Notation:
  - $E(s_t) \equiv \mu_t$; $var(s_t) \equiv \gamma_{0t}$; $cov(s_t,s_{t-j}) \equiv \gamma_{jt}$ [Autocovariance]

- Definition of Covariance (weak) Stationarity:
  - $\{s_t\}$ is called covariance-stationary (CS) iff

    $\mu_t = \mu$ (finite) for all t and $\gamma_{jt} = \gamma_j$ (finite) for all t.

    [e.g., $cov(s_5,s_3) = cov(s_4,s_2) = \gamma_2$]

- Rough Definition of Ergodicity: $\gamma_{jt} \to 0$ as $j \to \infty$.

- Ergodicity is needed for CLT and LLN.

(2) Stationarity Assumption on ε_t in $y_t = x_{t\bullet}'\beta + \varepsilon_t$:

- $E(\varepsilon_t) = 0$, for all t.
- $var(\varepsilon_t) = \sigma_\varepsilon^2$, for all t.
- $cov(\varepsilon_t,\varepsilon_{t-j}) = \gamma_{tj} = \gamma_j$, for all t.
- $\gamma_j \to 0$ as $j \to \infty$.

[3] Properties of OLS

(1) Autocorrelation:

$\rho_j = \gamma_j/\gamma_0 = \gamma_j/\text{var}(\varepsilon_t)$.

(2) Autocorrelation Matrix under the stationarity assumption:

- $$R = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \rho_3 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \rho_2 & \cdots & \rho_{T-2} \\ \rho_2 & \rho_1 & 1 & \rho_1 & \cdots & \rho_{T-3} \\ \rho_3 & \rho_2 & \rho_1 & 1 & \cdots & \rho_{T-4} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \rho_{T-4} & \cdots & 1 \end{pmatrix}.$$
- Let $\sigma^2 = \text{var}(\varepsilon_t) \equiv \gamma_0$.
- $\text{Cov}(\varepsilon) = \sigma^2\Omega = \sigma^2 R$.

Theorem:

The OLS estimator $\hat{\beta}$ is unbiased (and consistent). But it is inefficient.

Theorem:

$$Cov(\hat{\beta}) = (X'X)^{-1}X'\sigma^2\Omega X(X'X)^{-1}.$$

Implications:

- The t and F-tests based on $s^2(X'X)^{-1}$ are all irrelevant.
- Need to estimate $\Delta = X'(\sigma^2\Omega)X$.
  - If $\hat{\Delta}$ is available, a consistent estimate of $Cov(\hat{\beta})$ is

    $$(X'X)^{-1}\hat{\Delta}(X'X)^{-1}.$$
- Using this estimate, we can compute t and Wald statistics.

(3) Nonparametric Estimation of $X'(\sigma^2\Omega)X$.

- Definitions:
  - Let e_t = OLS residuals; and $g_t = e_t x_{t\bullet}$.
  - Define: $S_0 = \frac{1}{T}\Sigma_t g_t g_t'$;

    $$S_j = \frac{1}{T}\Sigma_{t=j+1}^{T} g_t g_{t-j}';$$

    $$\hat{\Delta} = T\left(S_0 + \Sigma_{i=1}^{q} k\left(\frac{i}{q+1}\right)(S_j + S_j')\right),$$

    where q is called bandwidth and k(•) is a kernel function.
  - Can also control for heteroskedasticity.

- Choice of q?
  - If we believe that $\rho_j = 0$ for $j > \tau$ [it happens if ε_t follows MA(τ)], choose $q = \tau$ or more.
  - If don't know? Choose q such that $q \to \infty$ and $q/T \to 0$, as $T \to \infty$.

- Newey-West Method (1987, ECON)
  - Use Bartlett's kernel, $k(z) = 1 - z$.
  - Choose $q = T^{1/5}$ or $q = T^{2/9}$ (same for other methods) so that $q/T^{1/4} \to 0$.

- Gallant Method (1987, Nonlinear Statistic Models)
  - Use Pazen kernel,

    $k(z) = 1 - 6z^2 + 6z^3$ for $0 \le z \le 1/2$;

    $= 2(1-z)^3$, $1/2 \le z \le 1$ and $= 0$, otherwise.

- Andrews (1991, ECON)
  - Quadratic kernel:

$$k(z) = \frac{3}{(6\pi z/5)^2}\left[\frac{\sin(6\pi z/5)}{6\pi z/5} - \cos(6\pi z/5)\right].$$

  - $\hat{\Delta} = \frac{T^2}{T-p}\left[S_0 + \Sigma_{i=1}^{T-1} k\left(\frac{i}{q+1}\right)(S_i + S_i')\right]$.

[4] Detection of First-Order AUTO

[Testing H_o: $\rho_1 = 0$ against H_a: $\rho_1 \neq 0$ assuming $\rho_2 = ... = 0$]

(1) Notation:

- Sample Autocovariance: $r_j = \frac{1}{T-j}\Sigma_{t=j+1}^{T} e_t e_{t-j}$.
- Sample Autocorrelation: $a_j = \frac{r_j}{r_0}$ $(-1 \leq a_j \leq 1)$.

Note: Under the CS and Ergodicity assumptions,

$r_j \to_p \gamma_j$ as $T \to \infty$;

$a_j \to_p \rho_j$ as $T \to \infty$.

(2) Durbin-Watson Test (Assuming regressors are strictly exogenous)

- D-W test statistic: DW = $\frac{\Sigma_{t=2}^{T}(e_t - e_{t-1})^2}{\Sigma_{t=1}^{T} e_t^2}$.
- Intuition:
    - DW = $2(1-a_1) - \frac{e_1^2 + e_T^2}{\Sigma_{t=1}^{T} e_t^2} \approx 2(1-a_1)$.

      → If H_o is correct, DW ≈ 2.

      → For formal test with DW, see book.

(3) LM-type Test (Regressors need not be strictly exogenous)

STEP 1: Regress e_t on $x_{t\bullet}$ and e_{t-1}, and get R^2.

STEP 2: t-test for Ho: the coefficient on $e_{t-1} = 0$. Or

$T \times R^2 \rightarrow_p \chi^2(1)$.

(4) Note on these tests.

- Suppose $\rho_j \neq 0$, $j = 2, 3, ...$ Then, these tests may fail to detect autocorrelation.
- Accepting Ho: $\rho_1 = 0$ does not mean that $\rho_j = 0$, $j = 2, 3, ...$

[5] Detection of Higher-Order AUTO

[H_o: $\rho_1 = \rho_2 = ... = \rho_m = 0$]

(1) LM-type test.

STEP 1: Regress e_t on $x_{t\bullet}$ and $e_{t-1}, ... , e_{t-m}$.

STEP 2: $LM = TR^2 \to_d \chi^2(m)$.

(2) Box-Pierce Test.

- $BP = T\Sigma_{j=1}^{m} a_j^2 \to_d \chi^2(m)$.

(3) Ljung-Box Test

- $LB = T(T+2)\Sigma_{j=1}^{m} \frac{a_j^2}{T-j} \to_d \chi^2(m)$.

[EMPIRICAL EXAMPLE]

Data: ustc96.wf1 (Quarterly data)

- FEDFRQ = Fed Fund Rate (%)
- GDP
- INF = annualized quarterly inflation rate (%)
- M1Q = M1
- M2Q = M2
- TBILLQ = treasury bill rate (%)
- GGDP = 400*ln(GDP/GDP(-1)):
  Annualized growth rate of GDP (%).
- GM1 = 400*LOG(M1Q/M1Q(-1)):
  Annualized growth rate of M1 (%)

Research project:

How Fed reacts to the economy?

Model to estimate:

$(fedfrq\text{-}tbillq) = \beta_1 + \beta_2 ggdp_{t-4} + \beta_3 inf_{t-1} + \varepsilon_t.$

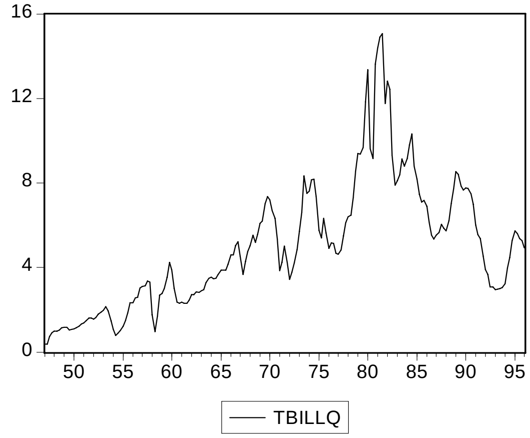
- Why $(fedfrq_t\text{-}tbillq_t)$ instead of fedfrq?

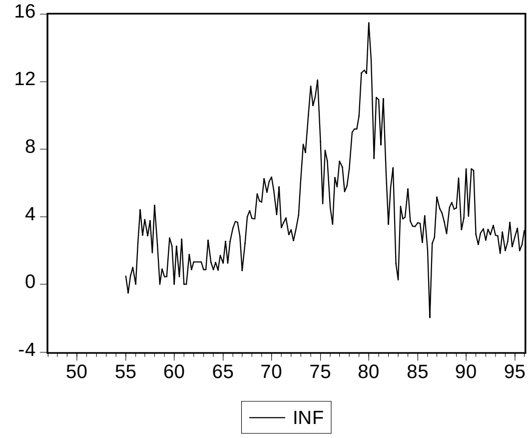
  The variable "fedfrq" appears to be nonstationary.

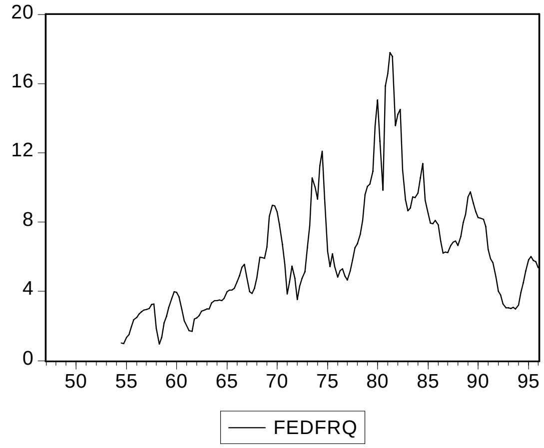
  But (fedfrq-tbillq) is stationary.

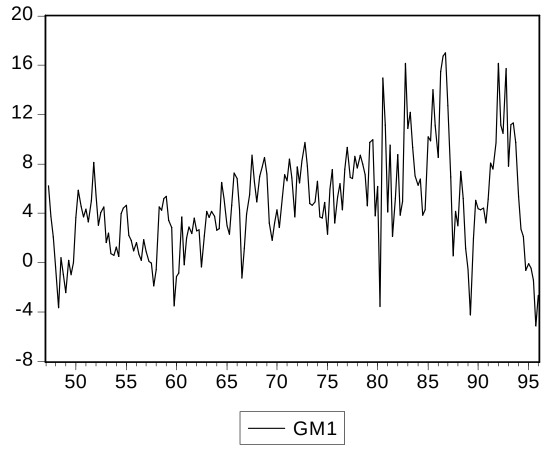
- Why not $ggdp_{t-1}$, $ggdp_{t-2}$, ... and inf_{t-2}, inf_{t-3}... ?

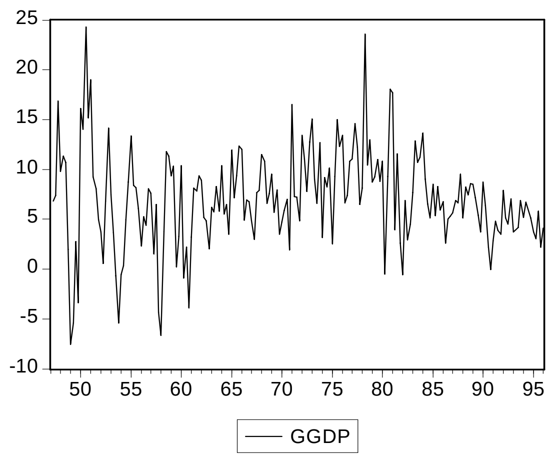
  They appear to be insignificant.


16
12
8
4
0
50 55 60 65 70 75 80 85 90 95
TBILLQ


16
12
8
4
0
-4
50 55 60 65 70 75 80 85 90 95
INF


20
16
12
8
4
0
50 55 60 65 70 75 80 85 90 95
FEDFRQ


20
16
12
8
4
0
-4
-8
50 55 60 65 70 75 80 85 90 95
GM1


25
20
15
10
5
0
-5
-10
50 55 60 65 70 75 80 85 90 95
GGDP

OLS Results:

Dependent Variable: DEVFEDTBILL
Method: Least Squares
Sample(adjusted): 1955:2 1996:1
Included observations: 164 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.166704	0.097205	-1.714962	0.0883
INF(-1)	0.161085	0.015494	10.39627	0.0000
GGDP(-4)	0.004583	0.011710	0.391376	0.6960

R-squared	0.443511	Mean dependent var	0.553354
Adjusted R-squared	0.436598	S.D. dependent var	0.775250
S.E. of regression	0.581904	Akaike info criterion	1.773099
Sum squared resid	54.51650	Schwarz criterion	1.829804
Log likelihood	-142.3942	F-statistic	64.15699
Durbin-Watson stat	0.836136	Prob(F-statistic)	0.000000

- DW = 0.836 → Implies AUTO.

view/residual tests/serial correlation LM test

lags to include: 4

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	22.08502	Probability	0.000000
Obs*R-squared	59.05181	Probability	0.000000

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 04/16/02 Time: 15:42
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.089264	0.080027	1.115428	0.2664
INF(-1)	-0.013771	0.012714	-1.083200	0.2804
GGDP(-4)	-0.004093	0.009658	-0.423734	0.6723
RESID(-1)	0.547954	0.080008	6.848755	0.0000
RESID(-2)	-0.001062	0.089982	-0.011801	0.9906
RESID(-3)	0.104931	0.089976	1.166214	0.2453
RESID(-4)	0.029028	0.080553	0.360359	0.7191

R-squared	0.360072	Mean dependent var	-4.56E-16
Adjusted R-squared	0.335616	S.D. dependent var	0.578323
S.E. of regression	0.471389	Akaike info criterion	1.375480
Sum squared resid	34.88664	Schwarz criterion	1.507792
Log likelihood	-105.7894	F-statistic	14.72335
Durbin-Watson stat	1.945094	Prob(F-statistic)	0.000000

OLS with NEWEY-WEST

option/heteroskedasticity/Newey-West

Dependent Variable: DEVFEDTBILL
Method: Least Squares
Sample(adjusted): 1955:2 1996:1
Included observations: 164 after adjusting endpoints
Newey-West HAC Standard Errors & Covariance (lag truncation=4)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.166704	0.104925	-1.588788	0.1141
INF(-1)	0.161085	0.026485	6.082115	0.0000
GGDP(-4)	0.004583	0.009885	0.463627	0.6435

R-squared	0.443511	Mean dependent var	0.553354
Adjusted R-squared	0.436598	S.D. dependent var	0.775250
S.E. of regression	0.581904	Akaike info criterion	1.773099
Sum squared resid	54.51650	Schwarz criterion	1.829804
Log likelihood	-142.3942	F-statistic	64.15699
Durbin-Watson stat	0.836136	Prob(F-statistic)	0.000000

[6] AR(1)

(1) Assumptions:

- $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$, v_t iid $N(0,\sigma_v^2)$, $|\rho| < 1$.
- $\{\varepsilon_t\}$ starts at the infinite past $(-\infty)$

(2) Derivation of Ω and V.

- $\varepsilon_t = \rho(\rho\varepsilon_{t-2}+v_{t-1}) + v_t = \rho^2\varepsilon_{t-2} + v_t + \rho v_{t-1}$

  $= \rho^2(\rho\varepsilon_{t-3}+v_{t-2}) + v_t + \rho v_{t-1} = \rho^3\varepsilon_{t-3} + v_t + \rho v_{t-1} + \rho^2 v_{t-2}$

  $= \rho^\infty\varepsilon_{t-\infty} + \Sigma_{j=0}^{\infty}\rho^j v_{t-j} = \Sigma_{j=0}^{\infty}\rho^j v_{t-j}$.

- $\gamma_0 = \text{var}(\varepsilon_t) = \text{var}\left(\Sigma_{j=0}^{\infty}\rho^j v_{t-j}\right) = \Sigma_{j=0}^{\infty}\rho^{2j}\,\text{var}(v_{t-j}) = \dfrac{\sigma_v^{\ 2}}{1-\rho^2}$.

- $\gamma_1 = E(\varepsilon_t\varepsilon_{t-1}) = E[(\rho\varepsilon_{t-1}+v_t)\varepsilon_{t-1}] = \rho E(\varepsilon_{t-1}^{\ 2}) = \rho\text{var}(\varepsilon_{t-1}) = \dfrac{\rho\sigma_v^{\ 2}}{1-\rho^2}$.

- $\gamma_2 = E(\varepsilon_t\varepsilon_{t-2}) = E[(\rho^2\varepsilon_{t-2}+\rho v_{t-1}+v_t)\varepsilon_{t-2}] = \rho^2 E(\varepsilon_{t-2}^{\ 2}) = \dfrac{\rho^2\sigma_v^{\ 2}}{1-\rho^2}$.

  $\vdots$

- In general, $\gamma_j = \rho^j\text{var}(\varepsilon_t)$: $\rho_j = \dfrac{\gamma_j}{\gamma_0} = \rho^j$.

- $\text{Cov}(\varepsilon) = \sigma^2\Omega = \sigma_v^2 \frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \ldots & \rho^{T-1} \\ \rho & 1 & \rho & \rho^2 & \ldots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \rho & \ldots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \rho^{T-4} & \ldots & 1 \end{pmatrix}$.

- Set $\Omega^* = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \ldots & \rho^{T-1} \\ \rho & 1 & \rho & \rho^2 & \ldots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \rho & \ldots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \rho^{T-4} & \ldots & 1 \end{pmatrix}$.

- $(\Omega^*)^{-1} = \begin{pmatrix} 1 & -\rho & 0 & 0 & \ldots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & \ldots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & 0 & \ldots & -\rho & 1 \end{pmatrix}$.

- $V = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & 0 & \ldots & 0 & 0 \\ -\rho & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & -\rho & 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1 & 0 \\ 0 & 0 & 0 & 0 & \ldots & -\rho & 1 \end{pmatrix} \rightarrow V'V = (\Omega^*)^{-1}$.

(3) GLS when ρ is known.

1) Prais-Winston GLS:

$$Vy = \begin{pmatrix} \sqrt{1-\rho^2}\,y_1 \\ y_2 - \rho y_1 \\ y_3 - \rho y_2 \\ \vdots \\ y_T - \rho y_{T-1} \end{pmatrix}; \; VX = \begin{pmatrix} \sqrt{1-\rho^2}\,x_{11} & \cdots \\ x_{21} - \rho x_{11} & \cdots \\ x_{31} - \rho x_{21} & \cdots \\ \vdots & \\ x_{T1} - \rho x_{T-1,1} & \cdots \end{pmatrix}.$$

→ OLS on $Vy = VX\beta + V\varepsilon$.

2) Cochrane-Orcutt GLS

$$y^* = \begin{pmatrix} y_2 - \rho y_1 \\ y_3 - \rho y_2 \\ \vdots \\ y_T - \rho y_{T-1} \end{pmatrix}; \; X^* = \begin{pmatrix} x_{21} - \rho x_{11} & \ldots \\ x_{31} - \rho x_{21} & \ldots \\ \vdots & \\ x_{T1} - \rho x_{T-1,1} & \ldots \end{pmatrix}.$$

- Asymptotically, P-W GLS = C-O GLS.
- When T is small, CO GLS is not desirable.

(4) FGLS when ρ is unknown

1) Consistent estimator of ρ

- Cochrane-Orcutt: $\hat{\rho} = \frac{\Sigma_{t=2}^{T} e_t e_{t-1}}{\Sigma_{t=1}^{T} e_t^{\,2}}$.
- Theil: $\hat{\rho} = \frac{(\Sigma_{t=2}^{T} e_t e_{t-1})/(T-1)}{(\Sigma_{t=1}^{T} e_t^{\,2})/(T-k)}$ or $\hat{\rho} = 1 - \frac{DW}{2}$.

2) Two-Step FGLS of β

- Do GLS with $\hat{\rho}$.
- Asymptotically identical to MLE of β if regressors are strictly exogenous.
- Two-Step FGLS and GLS (or MLE) can have different distributions if regressors are only weakly exogenous.

3) Iterative FGLS of β.

STEP 1: Get Two-Step FGLS.

STEP 2: Using T-S FGLS, compute residuals.

STEP 3: Using these residuals, estimate ρ again.

STEP 4: Using this estimate ρ, compute FGLS.

STEP 5: Keep doing this until the estimates of β do not change.

Comments:

- Asymptotically identical to MLE of β and two-step GLS if regressors are strictly exogenous.
- Iterative FGLS and GLS (or MLE) can have different distributions if regressors are only weakly exogenous.

(5) Nonlinear Least Square Approach (Estimate β and ρ jointly)

- Durbin Equation: $y_t = \rho y_{t-1} + x_{t\bullet}'\beta + x_{t-1,\bullet}'(-\rho\beta) + v_t$.

$$(y_t = x_{t\bullet}'\beta + \varepsilon_t) - (\rho y_{t-1} = x_{t-1,\bullet}'(\rho\beta) + \rho\varepsilon_{t-1})$$

→ $y_t = \rho y_{t-1} + x_{t\bullet}'\beta + x_{t-1,\bullet}'(-\rho\beta) + v_t$. [Durbin Equation]

→ Regress y_t on y_{t-1}, $x_{t\bullet}$ and $x_{t-1,\bullet}$.

→ The estimated coefficient on y_{t-1} is a consistent estimate of ρ.

- Estimate this equation by NLLS
- Both estimates of β and ρ are asymptotically identical to MLE if the v_t are normal.
- Okay with weakly exogenous $x_{t\bullet}$.

(6) MLE (Estimate β, ρ and σ_v^2 jointly): See Greene.

- $y \sim N(X\beta, \sigma^2\Omega)$:

$$\rightarrow f(y) = \frac{1}{(2\pi)^{T/2}\left|\sigma^2\Omega\right|^{T/2}}\exp\left(-\frac{1}{2\sigma^2}(y - X\beta)'\Omega^{-1}(y - X\beta)\right),$$

where $\sigma^2 = \dfrac{\sigma_v^2}{1-\rho^2}$.

- $l_T(\theta) = \ln f(y) = -(T/2)\ln(2\pi) - (T/2)\ln(\sigma_v^2) + (1/2)\ln(1-\rho^2)$

  $- [1/(2\sigma_v^2)](Vy-VX\beta)'(Vy-VX\beta)$

  $= -(T/2)\ln(2\pi) - (T/2)\ln(\sigma_v^2) + (1/2)\ln(1-\rho^2)$

  $- [(1-\rho^2)/(2\sigma_v^2)](y_1-x_{1\bullet}'\beta)^2$

  $-[1/(2\sigma_v^2)]\Sigma_{t=2}^{T}(y_t - \rho y_{t-1} - x_{t\bullet}'\beta + x_{t-1,\bullet}'(\rho\beta))^2$,

  where $\theta = (\beta',\rho,\sigma_v^2)'$.

- If T is large: $(1/2)\ln(1-\rho^2) - [(1-\rho^2)/(2\sigma_v^2)](y_1-x_{1\bullet}'\beta)^2$ is negligible.

- Full MLE ≈ partial MLE (called MLE conditional on y_1) on

  $-(T/2)\ln(2\pi) - (T/2)\ln(\sigma_v^2)$

  $- [1/(2\sigma_v^2)]\ \Sigma_{t=2}^{T}(y_t - \rho y_{t-1} - x_{t\bullet}'\beta + x_{t-1,\bullet}'(\rho\beta))^2$.

  → Conditional MLE = NLLS.

  → In EViews, y c x1 x2 x3 ar(1).

[EMPIRICAL EXAMPLE]

Data: ustc96.wf1

- FEDFRQ = Fed Fund Rate (%)
- GDP
- INF = annualized quarterly inflation rate (%)
- M1Q = M1
- M2Q = M2
- TBILLQ = treasury bill rate (%)
- GGDP = 400*ln(GDP/GDP(-1)):
  Annualized growth rate of GDP (%).
- GM1 = 400*LOG(M1Q/M1Q(-1)):
  Annualized growth rate of M1 (%)

Model to estimate:

$$(fedfrq\text{-}tbillq) = \beta_1 + \beta_2 ggdp_{t-4} + \beta_3 inf_{t-1} + \varepsilon_t.$$

From OLS (p. AUTO_24), $\hat{\rho}$ = 1- DW/2 = 1 – 0.836/2 = 0.582.

Cochrane-Orcutt GLS:

codevfedtbill = devfedtbill – 0.582*devfedtbill(-1);

coinf = inf(-1) – 0.582*inf(-2);

coggdp = ggdp(-4) – 0.582*ggdp(-5).

Dependent Variable: CODEVFEDTBILL
Method: Least Squares
Sample(adjusted): 1955:3 1996:1
Included observations: 163 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.120528	0.046543	2.589568	0.0105
COINF	0.062266	0.019788	3.146685	0.0020
COGGDP	-0.013761	0.008305	-1.657006	0.0995

R-squared	0.071071	Mean dependent var	0.172108
Adjusted R-squared	0.059460	S.D. dependent var	0.452537
S.E. of regression	0.438877	Akaike info criterion	1.209040
Sum squared resid	30.81813	Schwarz criterion	1.265980
Log likelihood	-95.53674	F-statistic	6.120716
Durbin-Watson stat	1.889811	Prob(F-statistic)	0.002745

Residual Tests:

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1.455219	Probability	0.218506
Obs*R-squared	5.863290	Probability	0.209593

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.015854	0.048248	0.328589	0.7429
COINF	-0.012542	0.021580	-0.581194	0.5619
COGGDP	0.000382	0.008388	0.045568	0.9637
RESID(-1)	0.060842	0.085325	0.713054	0.4769
RESID(-2)	0.067904	0.080385	0.844726	0.3996
RESID(-3)	0.024977	0.079489	0.314223	0.7538
RESID(-4)	0.160788	0.079539	2.021496	0.0449

R-squared	0.035971	Mean dependent var	1.20E-17
Adjusted R-squared	-0.001107	S.D. dependent var	0.436160
Log likelihood	-92.55107	F-statistic	0.970146
Durbin-Watson stat	1.980720	Prob(F-statistic)	0.447394

NLLS Estimation: Joint estimation of ρ and β

Dependent Variable: DEVFEDTBILL
Method: Least Squares
Sample(adjusted): 1955:3 1996:1
Included observations: 163 after adjusting endpoints
Convergence achieved after 10 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.529433	0.204273	2.591796	0.0104
INF(-1)	0.036529	0.021294	1.715487	0.0882
GGDP(-4)	-0.016524	0.007907	-2.089836	0.0382
AR(1)	0.801587	0.049480	16.20038	0.0000

R-squared	0.690945	Mean dependent var	0.556646
Adjusted R-squared	0.685114	S.D. dependent var	0.776488
S.E. of regression	0.435724	Akaike info criterion	1.200618
Sum squared resid	30.18701	Schwarz criterion	1.276538
Log likelihood	-93.85038	F-statistic	118.4908
Durbin-Watson stat	2.037432	Prob(F-statistic)	0.000000

Inverted AR Roots	.80

[7] AR(p)

(1) Assumptions:

- $\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + ... + \rho_p\varepsilon_{t-p} + v_t$, v_t iid $N(0,\sigma_v^2)$.

  The solutions of $1 - \rho_1 z - ... - \rho_p z^p = 0$ are all outside of (-1,1).

  → Gaurantees CS and ergodicity.

- $\{\varepsilon_t\}$ starts at the infinite past ($-\infty$)

(2) Estimation

- θ's can be consistently estimated by regressing e_t on $e_{t-1}, ... , e_{t-p}$.

  → The estimated coefficient on e_{t-j} is a consistent estimate of ρ_j.

- Use Cochrane-Ocutt type FGLS:

  $\rightarrow y_t^* = y_t - \hat{\rho}_1 y_{t-1} - \hat{\rho}_2 y_{t-2} - ... - \hat{\rho}_p y_{t-p}$;

  $x_{t\bullet}^* = x_{t\bullet} - \hat{\rho}_1 x_{t-1,\bullet} - \hat{\rho}_2 x_{t-2,\bullet} - ... - \hat{\rho}_p x_{t-p,\bullet}$.

- Use NLLS:

$$(y_t = x_{t\bullet}'\beta + \varepsilon_t) - (\rho_1 y_{t-1} = x_{t-1,\bullet}'(\rho_1\beta) + \rho_1\varepsilon_{t-1})$$

$$- \ldots - (\rho_p y_{t-p} = x_{t-p,\bullet}'(\rho_p\beta) + \rho_p\varepsilon_{t-p})$$

$\rightarrow y_t - \rho_1 y_{t-1} - \ldots - \rho_p y_{t-p} = x_{t\bullet}'\beta + x_{t-1,\bullet}'(-\rho_1\beta) + \ldots + x_{t-p,\bullet}'(-\rho_p\beta)$.

$\rightarrow y_t = \rho_1 y_{t-1} + \ldots + \rho_p y_{t-p} + x_{t\bullet}'\beta + x_{t-1,\bullet}'(-\rho_1\beta) + \ldots + x_{t-p,\bullet}'(-\rho_p\beta)$.

$\rightarrow$ Do NLLS on this equation = MLE conditional on $y_1, \ldots, y_p$.

$\approx$ Full MLE if T is large.

$\rightarrow$ In EViews, y c x1 x2 x3 ar(1) ar(2) ... ar(p).

[8] MOVING AVERAGE: MA(q)

(1) Assumption

1) $\varepsilon_t = v_t + \psi_1 v_{t-1} + ... + \psi_q v_{t-q}$, v_t iid $N(0,\sigma_v^2)$.

2) The ψ_j are finite: Guarantees CS.

(2) Estimation

1) FGLS is possible for MA(1), but is messy. (See Fomby, et al)

2) Use conditional MLE (See Hamilton, Ch. 5.)

→ This is a kind of NLLS.

→ In EViews, y c x1 x2 x3 ma(1) ma(2) ... ma(q)

[EMPIRICAL EXAMPLE]

Data: ustc96.wf1

- FEDFRQ = Fed Fund Rate (%)
- GDP
- INF = annualized quarterly inflation rate (%)
- M1Q = M1
- M2Q = M2
- TBILLQ = treasury bill rate (%)
- GGDP = 400*ln(GDP/GDP(-1)):
  Annualized growth rate of GDP (%).
- GM1 = 400*LOG(M1Q/M1Q(-1)):
  Annualized growth rate of M1 (%)
- DEVFEDTBILL = fedfrq – tbillq.

Model to estimate:

$$(\text{fedfrq-tbillq}) = ggdp_t = \beta_1 + \beta_2 ggdp_{t-4} + \beta_3 inf_{t-1} + \varepsilon_t.$$

NLLS with the MA(1) error assumption

Dependent Variable: DEVFEDTBILL
Method: Least Squares
Date: 04/16/02 Time: 16:32
Sample(adjusted): 1955:2 1996:1
Included observations: 164 after adjusting endpoints
Convergence achieved after 21 iterations
Backcast: 1955:1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003068	0.109018	0.028145	0.9776
INF(-1)	0.135388	0.017354	7.801677	0.0000
GGDP(-4)	-0.003800	0.009799	-0.387766	0.6987
MA(1)	0.500564	0.070720	7.078139	0.0000

R-squared	0.598690	Mean dependent var	0.553354
Adjusted R-squared	0.591165	S.D. dependent var	0.775250
S.E. of regression	0.495697	Akaike info criterion	1.458382
Sum squared resid	39.31442	Schwarz criterion	1.533989
Log likelihood	-115.5874	F-statistic	79.56466
Durbin-Watson stat	1.646195	Prob(F-statistic)	0.000000

Inverted MA Roots	-.50

view/residual tests/serial correlation LM test

lags to include: 4

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	8.456939	Probability	0.000003
Obs*R-squared	29.22516	Probability	0.000007

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.080678	0.101892	0.791802	0.4297
INF(-1)	-0.024547	0.016595	-1.479138	0.1411
GGDP(-4)	0.003187	0.009113	0.349778	0.7270
MA(1)	-0.966492	0.937799	-1.030596	0.3043
RESID(-1)	1.092440	0.950547	1.149275	0.2522
RESID(-2)	-0.172172	0.478126	-0.360097	0.7193
RESID(-3)	0.307428	0.264810	1.160938	0.2474
RESID(-4)	-0.002848	0.144443	-0.019715	0.9843

R-squared	0.178202	Mean dependent var	0.000221
Adjusted R-squared	0.141327	S.D. dependent var	0.491114
S.E. of regression	0.455089	Akaike info criterion	1.310902
Sum squared resid	32.30849	Schwarz criterion	1.462115
Log likelihood	-99.49395	F-statistic	4.832531
Durbin-Watson stat	1.946651	Prob(F-statistic)	0.000060

[9] ARMA(p,q)

(1) Assumption

1) $\varepsilon_t = \rho_1\varepsilon_{t-1} + ... + \rho_p\varepsilon_{t-p} + v_t + \psi_1 v_{t-1} + ... + \psi_q v_{t-q}$, v_t iid $N(0,\sigma_v^2)$.

2) The ψ_j are finite and all solutions of $1- \rho_1 z - \rho_2 z^2 - ... - \rho_p z^p = 0$ are all outside of (-1,1): Guarantees CS and ergodicity.

(2) Estimation

- Use conditional MLE (See Hamilton, Ch. 5.)

  $\rightarrow$ This is a kind of NLLS.

  $\rightarrow$ In EViews, y c x1 x2 x3 ar(1) ... ar(p) ma(1) ... ma(q).

Dependent Variable: DEVFEDTBILL
Method: Least Squares
Sample(adjusted): 1955:3 1996:1
Included observations: 163 after adjusting endpoints
Convergence achieved after 11 iterations
Backcast: 1955:2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.523522	0.208742	2.507982	0.0132
INF(-1)	0.037847	0.021901	1.728134	0.0859
GGDP(-4)	-0.016419	0.008010	-2.049704	0.0420
AR(1)	0.810490	0.058291	13.90420	0.0000
MA(1)	-0.029295	0.101800	-0.287767	0.7739

R-squared	0.691116	Mean dependent var	0.556646
Adjusted R-squared	0.683296	S.D. dependent var	0.776488
S.E. of regression	0.436980	Akaike info criterion	1.212337
Sum squared resid	30.17036	Schwarz criterion	1.307237
Log likelihood	-93.80543	F-statistic	88.37967
Durbin-Watson stat	2.000338	Prob(F-statistic)	0.000000

Inverted AR Roots	.81
Inverted MA Roots	.03

[10] Forecasting

(1) AR(1) Model:

$$y_t = x_{t\bullet}'\beta + \varepsilon_t ; \varepsilon_t = \rho\varepsilon_{t-1} + v_t .$$

$$\hat{E}(y_{T+1} \mid \Omega_T) = \hat{x}_{T+1,\bullet}'\hat{\beta} + \hat{E}(\varepsilon_{T+1} \mid \Omega_T) = \hat{x}_{T+1,\bullet}'\hat{\beta} + \widehat{\rho\varepsilon_T}$$

$$= \hat{x}_{T+1,\bullet}'\hat{\beta} + \hat{\rho}(y_T - x_{T\bullet}'\hat{\beta}).$$

$$\hat{E}(y_{T+2} \mid \Omega_T) = \hat{x}_{T+2,\bullet}'\hat{\beta} + \hat{E}(\varepsilon_{T+2} \mid \Omega_T) = \hat{x}_{T+2,\bullet}'\hat{\beta} + \widehat{\rho\varepsilon_{T+1}}$$

$$= \hat{x}_{T+2,\bullet}'\hat{\beta} + \widehat{\rho^2\varepsilon_T} = \hat{x}_{T\bullet}'\hat{\beta} + \hat{\rho}^2(y_T - x_{T\bullet}'\hat{\beta}).$$

$$\hat{E}(y_{T+n} \mid \Omega_T) = \hat{x}_{T+n,\bullet}\hat{\beta} + \hat{\rho}^n(y_T - x_{T\bullet}'\hat{\beta}).$$

(2) MA(1) Model:

$$y_t = x_{t\bullet}'\beta + \varepsilon_t ; \varepsilon_t = \psi v_{t-1} + v_t .$$

- Estimating v_T:

  Set $\hat{v}_0 = 0$.

  $\hat{v}_1 = \hat{\varepsilon}_1 - \hat{\psi}\hat{v}_o = (y_1 - x_{1\bullet}'\hat{\beta})$.

  $\hat{v}_2 = \hat{\varepsilon}_2 - \hat{\psi}\hat{v}_1 = (y_2 - x_{2\bullet}'\hat{\beta}) + \hat{\psi}v_0$.

  :

  $\hat{v}_T = (y_T - x_{T\bullet}'\hat{\beta}) - \hat{\psi}\hat{v}_{T-1}$.

- Forecasting:

$$\hat{E}(y_{T+1} \mid \Omega_T) = \hat{x}_{T+1,\bullet}\hat{\beta} + \hat{E}(\varepsilon_{T+1} \mid \Omega_T) = \hat{x}_{T+1,\bullet}'\hat{\beta} + \widehat{v_{T+1} + \psi v_T}$$

$$= \hat{x}_{T+1,\bullet}'\hat{\beta} + \hat{\psi}\hat{v}_T.$$

$$\hat{E}(y_{T+2} \mid \Omega_T) = \hat{x}_{T+2,\bullet}\hat{\beta} + \hat{E}(\varepsilon_{T+2} \mid \Omega_T) = \hat{x}_{T+2,\bullet}'\hat{\beta} + \widehat{v_{T+2} + \psi v_{T+1}}$$

$$= \hat{x}_{T+2,\bullet}'\hat{\beta}.$$

:

$$\hat{E}(y_{T+n} \mid \Omega_T) = \hat{x}_{T+n,\bullet}'\hat{\beta}.$$