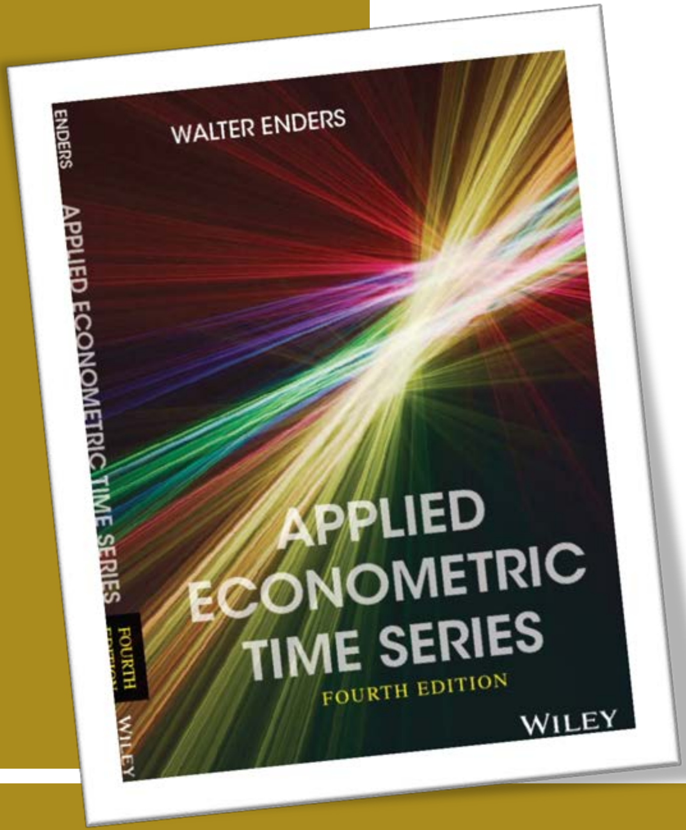


Applied Econometric Time Series 4th Edition

Chapter 2: STATIONARY TIME-SERIES MODELS

Walter Enders, University of Alabama

Section 1

STOCHASTIC DIFFERENCE EQUATION MODELS

Example of a time-series model

$$m_t = \rho(1.03)^t m_0^* + (1-\rho)m_{t-1} + \varepsilon_t \qquad (2.2)$$

1. Although the money supply is a continuous variable, (2.2) is a discrete difference equation. Since the forcing process $\{\varepsilon_t\}$ is stochastic, the money supply is stochastic; we can call (2.2) a linear stochastic difference equation.
2. If we knew the distribution of $\{\varepsilon_t\}$, we could calculate the distribution for each element in the $\{m_t\}$ sequence. Since (2.2) shows how the realizations of the $\{m_t\}$ sequence are linked across time, we would be able to calculate the various joint probabilities. Notice that the distribution of the money supply sequence is completely determined by the parameters of the difference equation (2.2) and the distribution of the $\{\varepsilon_t\}$ sequence.
3. Having observed the first t observations in the $\{m_t\}$ sequence, we can make forecasts of m_{t+1}, m_{t+2}, ….

White Noise

- $E(\varepsilon_t) = E(\varepsilon_{t-1}) = \ldots = 0$
- $E(\varepsilon_t)^2 = E(\varepsilon_{t-1})^2 = \ldots = \sigma^2$
  - [or $\mathrm{var}(\varepsilon_t) = \mathrm{var}(\varepsilon_{t-1}) = \ldots = \sigma^2$]
- $E(\varepsilon_t\, \varepsilon_{t-s}) = E(\varepsilon_{t-j}\, \varepsilon_{t-j-s}) = 0$ for all j and s
  - [or $\mathrm{cov}(\varepsilon_t, \varepsilon_{t-s}) = \mathrm{cov}(\varepsilon_{t-j}, \varepsilon_{t-j-s}) = 0$]

$$x_t = \sum_{i=0}^{q} \beta_i \varepsilon_{t-i}$$

A sequence formed in this manner is called a **moving average** of order q and is denoted by MA(q)

2. ARMA MODELS

In the *ARMA(p, q)* model

$$y_t = a_0 + a_1 y_{t-1} + \dots + a_p y_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}$$

where ε_t series are serially uncorrelated "shocks"

The particular solution is:

$$y_t = \left(a_0 + \sum_{i=0}^{q} \beta_i \varepsilon_{t-i} \right) \Big/ \left(1 - \sum_{i=1}^{p} a_i L^i \right)$$

Note that all roots must lie outside of the unit circle.
If this is the case, we have the MA Representation

$$y_t = c + \sum_{i=0}^{\infty} c_i \varepsilon_{t-i}$$

Section 3

Stationarity Restrictions for an AR(1) Process

STATIONARITY

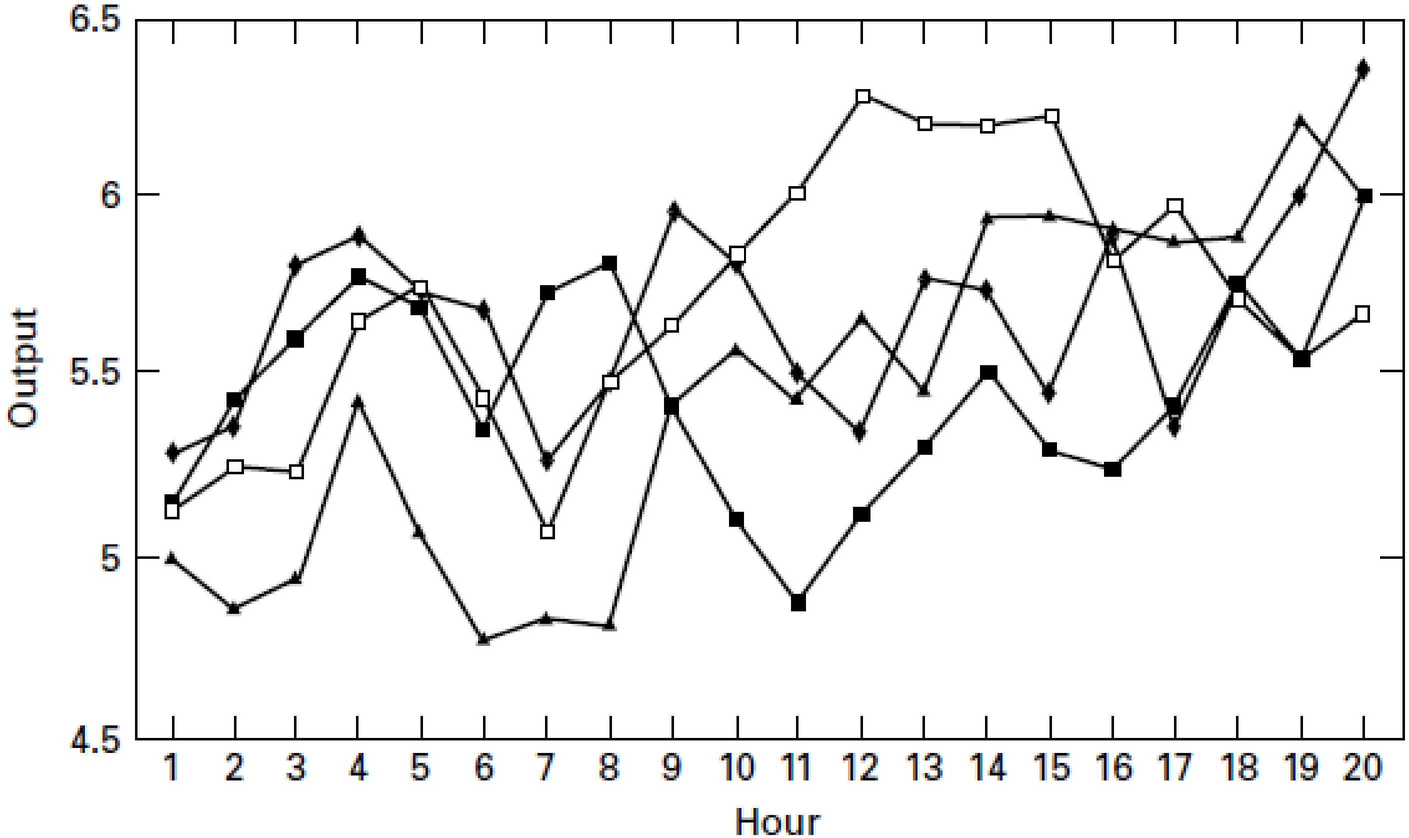


FIGURE 2.1 Hourly Output of Four Machines

Covariance Stationary Series

- Mean is time-invariant
- Variance is constant
- All covariances are constant
  - all autocorrelations are constant
- Example of a series that are not covariance stationary
  - $y_t = \alpha + \beta \, time$
  - $y_t = y_{t-1} + \varepsilon_t$ (Random Walk)

Formal Definition

A stochastic process having a finite mean and variance is **covariance stationary** if for all t and $t - s$,

1. $E(y_t) = E(y_{t-s}) = \mu$

2. $E[(y_t - \mu)^2] = E[(y_{t-s} - \mu)^2]$

   or $[\text{var}(y_t) = \text{var}(y_{t-s}) =]$

3. $E[(y_t - \mu)(y_{t-s} - \mu)] = E[(y_{t-j} - \mu)(y_{t-j-s} - \mu)] = \gamma_s$
   or $\text{cov}(y_t, y_{t-s}) = \text{cov}(y_{t-j}, y_{t-j-s}) = \gamma_s$

where μ, and γ_s are all constants.

4. STATIONARITY RESTRICTIONS FOR AN ARMA(*p*, *q*) MODEL

- **Stationarity Restrictions for the Autoregressive Coefficients**

Stationarity of an AR(1) Process

$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$ with an initial condition

$$y_t = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0 + \sum_{i=0}^{t-1} a_1^i \varepsilon_{t-i}$$

Only if t is large is this stationary:

$$\lim y_t = \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i}$$

$$E[(y_t - \mu)(y_{t-s} - \mu)] = E\{[\varepsilon_t + a_1\varepsilon_{t-1} + (a_1)^2\varepsilon_{t-2} + \ldots][\varepsilon_{t-s} + a_1\varepsilon_{t-s-1} + (a_1)^2\varepsilon_{t-s-2} + \ldots]\}$$

$$= \sigma^2(a_1)^s[1 + (a_1)^2 + (a_1)^4 + \ldots]$$

$$= \sigma^2(a_1)^s/[1 - (a_1)^2]$$

Restrictions for the AR Coefficients

Let $y_t = a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \varepsilon_t$

so that $y_t = a_0 / \left[1 - \sum_{i=1}^{p} a_i\right] + \sum_{i=0}^{\infty} c_i \varepsilon_{t-i}$

We know that the sequence $\{c_i\}$ will eventually solve the difference equation

$$c_i - a_1 c_{i-1} - a_2 c_{i-2} - \ldots - a_p c_{i-p} = 0 \qquad (2.21)$$

If the characteristic roots of (2.21) are all inside the unit circle, the $\{c_i\}$ sequence will be convergent.

The stability conditions can be stated succinctly:

1. The homogeneous solution must be zero. Either the sequence must have started infinitely far in the past or the process must always be in equilibrium (so that the arbitrary constant is zero).
2. The characteristic root a_1 must be less than unity in absolute value.

A Pure MA Process $x_t = \sum_{i=0}^{\infty} \beta_i \varepsilon_{t-i}$

1. Take the expected value of x_t

$$
\begin{aligned}
E(x_t) &= E(\varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \;\;) \\
&= E\varepsilon_t + \beta_1 E\varepsilon_{t-1} + \beta_2 E\varepsilon_{t-2} + \;\; = 0 \\
E(x_{t-s}) &= E(\varepsilon_{t-s} + \beta_1 \varepsilon_{t-s-1} + \beta_2 \varepsilon_{t-s-2} + \;\;) = 0
\end{aligned}
$$

Hence, all elements in the $\{x_t\}$ sequence have the same finite mean ($\mu = 0$).

2. Form var(x_t) as

$$
\begin{aligned}
\text{var}(x_t) &= E[(\varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \;\;)^2] \\
&= \sigma^2[1 + (\beta_1)^2 + (\beta_2)^2 + \;\;]
\end{aligned}
$$

As long as $\Sigma(\beta_i)^2$ is finite, it follows that var(x_t) is finite.

$$
\begin{aligned}
\text{var}(x_{t-s}) &= E[(\varepsilon_{t-s} + \beta_1 \varepsilon_{t-s-1} + \beta_2 \varepsilon_{t-s-2} + \;\;)^2] \\
&= \sigma^2[1 + (\beta_1)^2 + (\beta_2)^2 + \;\;]
\end{aligned}
$$

Thus, var(x_t) = var(x_{t-s}) for all t and $t-s$.

Are all autocovariances finite and time independent?

$$
\begin{aligned}
E[x_t x_{t-s}] &= E[(\varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \;\;)(\varepsilon_{t-s} + \beta_1 \varepsilon_{t-s-1} + \beta_2 \varepsilon_{t-s-2} + \;\;)] \\
&= \sigma^2(\beta_s + \beta_1 \beta_{s+1} + \beta_2 \beta_{s+2} + \;\;)
\end{aligned}
$$

Restricting the sum $\beta_s + \beta_1 \beta_{s+1} + \beta_2 \beta_{s+2} + \;\;$ to be finite means that $E(x_t x_{t-s})$ is finite.

The Autocorrelation Function of an AR(2) Process

The Autocorrelation Function of an MA(1) Process

The Autocorrelation Function of an ARMA(1, 1) Process

5. THE AUTOCORRELATION FUNCTION

The Autocorrelation Function of an MA(1) Process

Consider $y_t = \varepsilon_t + \beta\varepsilon_{t-1}$. Again, multiply y_t by each y_{t-s} and take expectations

$$\gamma_0 = \text{var}(y_t) = Ey_ty_t = E[(\varepsilon_t + \beta\varepsilon_{t-1})(\varepsilon_t + \beta\varepsilon_{t-1})] = (1 + \beta^2)\sigma^2$$

$$\gamma_1 = \text{cov}(y_ty_{t-1}) = Ey_ty_{t-1} = E[(\varepsilon_t + \beta\varepsilon_{t-1})(\varepsilon_{t-1} + \beta\varepsilon_{t-2})] = \beta\sigma^2$$

and

$$\gamma_s = Ey_ty_{t-s} = E[(\varepsilon_t + \beta\varepsilon_{t-1})(\varepsilon_{t-s} + \beta\varepsilon_{t-s-1})] = 0 \quad \text{for all } s > 1$$

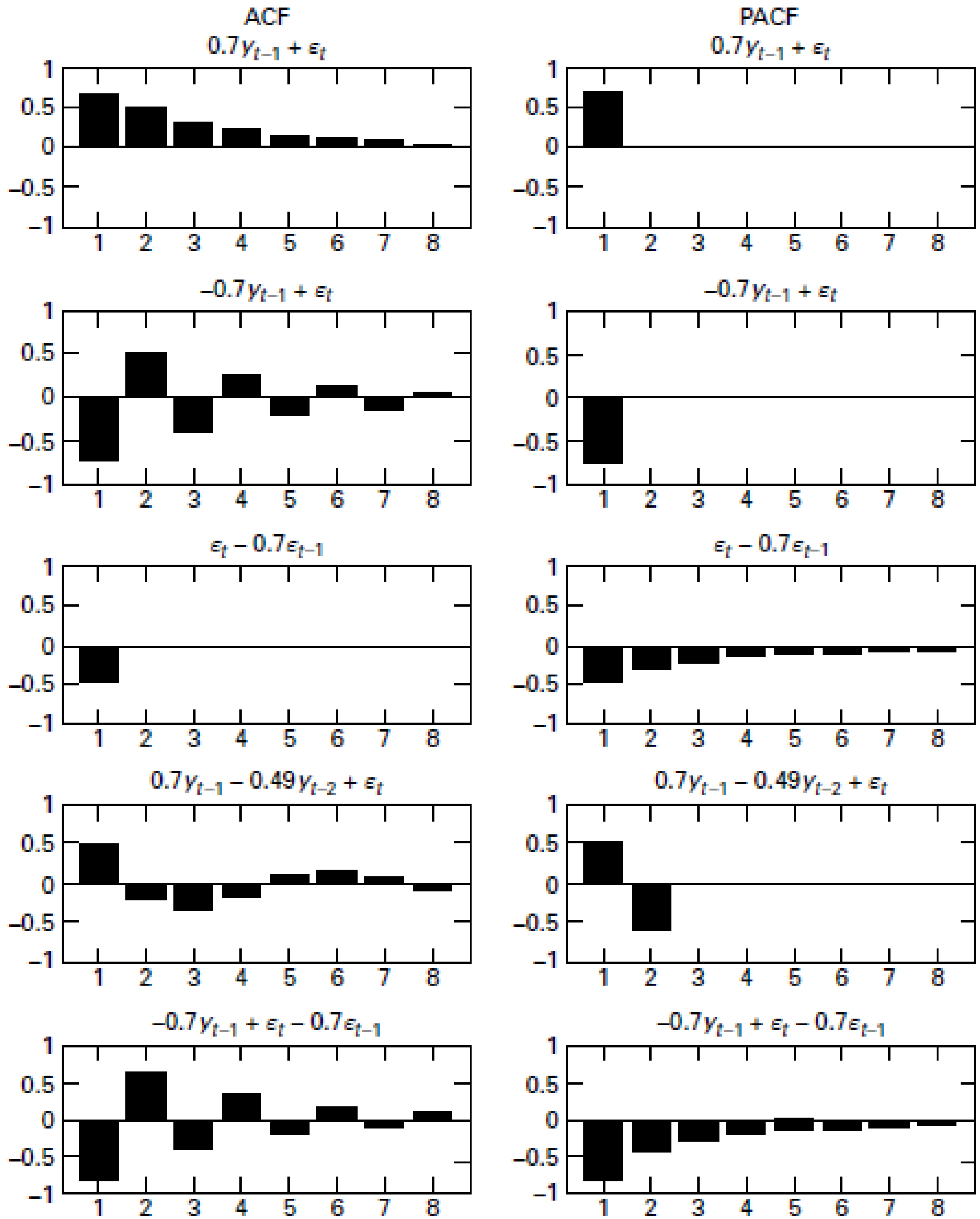


FIGURE 2.2 Theoretical ACF and PACF Patterns

The ACF of an ARMA(1, 1) Process:

Let $y_t = a_1 y_{t-1} + \varepsilon_t + \beta_1 \varepsilon_{t-1}$.

$$Ey_t y_t = a_1 E y_{t-1} y_t + E\varepsilon_t y_t + \beta_1 E\varepsilon_{t-1} y_t$$

$$\Rightarrow \gamma_0 = a_1\gamma_1 + \sigma^2 + \beta_1(a_1+\beta_1)\sigma^2$$

$$Ey_t y_{t-1} = a_1 E y_{t-1} y_{t-1} + E\varepsilon_t y_{t-1} + \beta_1 E\varepsilon_{t-1} y_{t-1}$$

$$\Rightarrow \gamma_1 = a_1\gamma_0 + \beta_1\sigma^2$$

$$Ey_t y_{t-2} = a_1 E y_{t-1} y_{t-2} + E\varepsilon_t y_{t-2} + \beta_1 E\varepsilon_{t-1} y_{t-2}$$

$$\Rightarrow \gamma_2 = a_1\gamma_1$$

.

$$Ey_t y_{t-s} = a_1 E y_{t-1} y_{t-s} + E\varepsilon_t y_{t-s} + \beta_1 E\varepsilon_{t-1} y_{t-s}$$

$$\Rightarrow \gamma_s = a_1\gamma_{s-1}$$

ACF of an AR(2) Process

$$Ey_ty_t = a_1Ey_{t-1}y_t + a_2Ey_{t-2}y_t + E\varepsilon_ty_t$$
$$Ey_ty_{t-1} = a_1Ey_{t-1}y_{t-1} + a_2Ey_{t-2}y_{t-1} + E\varepsilon_ty_{t-1}$$
$$Ey_ty_{t-2} = a_1Ey_{t-1}y_{t-2} + a_2Ey_{t-2}y_{t-2} + E\varepsilon_ty_{t-2}$$

.

.

$$Ey_ty_{t-s} = a_1Ey_{t-1}y_{t-s} + a_2Ey_{t-2}y_{t-s} + E\varepsilon_ty_{t-s}$$

So that

$$\gamma_0 = a_1\gamma_1 + a_2\gamma_2 + \sigma^2$$
$$\gamma_1 = a_1\gamma_0 + a_2\gamma_1 \quad \rightarrow \quad \rho_1 = a_1/(1 - a_2)$$
$$\gamma_s = a_1\gamma_{s-1} + a_2\gamma_{s-2} \quad \rightarrow \quad \rho_i = a_1\rho_{i-1} + a_2\rho_{i-2}$$

6. THE PARTIAL AUTOCORRELATION FUNCTION

PACF of an AR Process

$$y_t = a + a_1 y_{t\text{-}1} + \varepsilon_t$$

$$y_t = a + a_1 y_{t\text{-}1} + a_2 y_{t\text{-}2} + \varepsilon_t$$

$$y_t = a + a_1 y_{t\text{-}1} + a_2 y_{t\text{-}2} + a_3 y_{t\text{-}3} + \varepsilon_t$$

…

The successive estimates of the a_i are the partial autocorrelations

PACF of a MA(1)

$y_t = \varepsilon_t + \beta_1 \varepsilon_{t\text{-}1}$

$but\ \varepsilon_{t\text{-}1} = y_{t\text{-}1} - \beta_1 \varepsilon_{t\text{-}2}$

$y_t = \varepsilon_t + \beta_1 [\, y_{t\text{-}1} - \beta_1 \varepsilon_{t\text{-}2} \,]$

$= \varepsilon_t + \beta_1 y_{t\text{-}1} - (\beta_1)^2 \varepsilon_{t\text{-}2}$

$y_t = \varepsilon_t + \beta_1 y_{t\text{-}1} - (\beta_1)^2 [y_{t\text{-}2} - \beta_1 \varepsilon_{t\text{-}3} \,] \ldots$

It looks like an MA(∞)

or Using Lag Operators

$$y_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} = (1 + \beta_1 L)\varepsilon_t$$

$$y_t/(1 + \beta_1 L) = \varepsilon_t$$

Recall $y_t/(1 - a_1 L) = y_t + a_1 y_{t-1} + a_1{}^2 y_{t-2} + a_1{}^3 y_{t-3} + \ldots$

so that $-\beta_1$ plays the role of a_1

$$y_t/(1 + \beta_1 L)\varepsilon_t = y_t/[1 - (-\beta_1)L]\varepsilon_t =$$

$$y_t - \beta_1 y_{t-1} + \beta_1{}^2 y_{t-2} - a_1{}^3 y_{t-3} + \ldots = \varepsilon_t$$

or

$$y_t = \beta_1 y_{t-1} - \beta_1{}^2 y_{t-2} + \beta_1{}^3 y_{t-3} + \ldots = \varepsilon_t$$

Summary: Autocorrelations and Partial Autocorrelations

ACF

- AR(1)
  - geometric decay
- MA(q)
  - cuts off at lag q

PACF

- AR(p)
  - Cuts off at lag p
- MA(1)
  - Geometric Decay

For stationary processes, the key points to note are the following:

- 1. The ACF of an ARMA(p,q) process will begin to decay after lag q. After lag q, the coefficients of the ACF (i.e., the r_i) will satisfy the difference equation ($\rho_i = a_1\rho_{i-1} + a_2\rho_{i-2} + \;\; + a_p\rho_{i-p}$).

- 2. The PACF of an ARMA(p,q) process will begin to decay after lag p. After lag p, the coefficients of the PACF (i.e., the f_{ss}) will mimic the ACF coefficients from the model $y_t/(1 + \beta_1 L + \beta_2 L^2 + \;\; + \beta_q L^q)$.

TABLE 2.1: Properties of the ACF and PACF

Process	ACF	PACF
White Noise	All $\rho_s = 0$ $(s \neq 0)$	All $\phi_{ss} = 0$
AR(1): $a_1 > 0$	Direct exponential decay: $\rho_s = a_1^s$	$\phi_{11} = \rho_1$; $\phi_{ss} = 0$ for $s \geq 2$
AR(1): $a_1 < 0$	Oscillating decay: $\rho_s = a_1^s$	$\phi_{11} = \rho_1$; $\phi_{ss} = 0$ for $s \geq 2$
AR(p)	Decays toward zero. Coefficients may oscillate.	Spikes through lag p. All $\phi_{ss} = 0$ for $s > p$.
MA(1): $\beta > 0$	Positive spike at lag 1. $\rho_s = 0$ for $s \geq 2$	Oscillating decay: $\phi_{11} > 0$.
MA(1): $\beta < 0$	Negative spike at lag 1. $\rho_s = 0$ for $s \geq 2$	Geometric decay: $\phi_{11} < 0$.
ARMA(1, 1) $a_1 > 0$	Geometric decay beginning after lag1. Sign ρ_1 = sign(a_1+β)	Oscillating decay after lag 1. $\phi_{11} = \rho_1$
ARMA(1, 1) $a_1 < 0$	Oscillating decay beginning after lag 1. Sign ρ_1 = sign(a_1+β)	Geometric decay beginning after lag 1. $\phi_{11} = \rho_1$ and sign(ϕ_{ss}) = sign(ϕ_{11}).
ARMA(p, q)	Decay (either direct or oscillatory) beginning after lag q.	Decay (either direct or oscillatory) beginning after lag p.

Testing the significance of ρ_i

- Under the null ρ_i = 0, the sample distribution of is:
  - approximately **normal** (but bounded at -1.0 and +1.0) when T is **large**
  - distributed as a student**s-*t*** when T is **small**.

- The standard formula for computing the appropriate ***t* value** to test significance of a correlation coefficient is:

$$t = \hat{\rho}_i \sqrt{\frac{T-2}{1-\hat{\rho}_i^2}}$$

with df = $T - 2$

- SD(ρ) = $[\,(1 - \rho^2) / (T - 2)\,]^{1/2}$

- In reasonably large samples, the test for the null that $\rho_i = 0$ is simplified to $T^{1/2}$. Alternatively, the standard deviation of the correlation coefficient is $(1/T)^{0.5}$.

Significance Levels

- A single autocorrelation
  - st.dev(ρ) = [(1 – ρ^2) / (T – 2)] ^{½}
    - For small ρ and large T, st.dev(ρ) is approx. $(1/T)^{1/2}$
  - If the autocorrelation exceeds | $2/T^{1/2}$ | we can reject the null that r = 0.
- A group of k autocorrelations:

$$Q = T(T+2)\sum_{i=1}^{k} \rho_i /(T-k)$$

Is a Chi-square with degrees of freedom = k

Model Selection Criteria

Estimation of an AR(1) Model

Estimation of an ARMA(1, 1) Model

Estimation of an AR(2) Model

7. SAMPLE AUTOCORRELATIONS OF STATIONARY SERIES

Sample Autocorrelations

$$r_s = \frac{\sum_{t=s+1}^{T}(y_t - \bar{y})(y_{t-s} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$$

Form the sample autocorrelations

$$Q = T(T+2)\sum_{k=1}^{s} r_k^2/(T-k)$$

Test groups of correlations

If the sample value of Q exceeds the critical value of χ^2 with s derees of freedom, then *at least* one value of r_k is statistically different from zero at the specified significance level.

The Box–Pierce and Ljung–Box Q-statistics also serve as a check to see if the *residuals* from an estimated ARMA(p,q) model behave as a white-noise process. However, the degrees of freedom are reduced by the number of estimated parameters

Model Selection

- AIC = T ln(sum of squared residuals) + 2n
- SBC = T ln(sum of squared residuals) + n ln(T)

where n = number of parameters estimated (p + q + possible constant term) T = number of usable observations.

ALTERNATIVE

- $AIC^* = -2\ln(L)/T + 2n/T$
- $SBC^* = -2\ln(L)/T + n\ln(T)/T$

- *where* n and T are as defined above, and L =maximized value of the log of the likelihood function.
- For a normal distribution, $-2\ln(L) = T\ln(2\pi) + T\ln(\sigma^2) + (1/\sigma^2)$ (sum of squared residuals)

Figure 2.3: ACF and PACF for two simulated processes

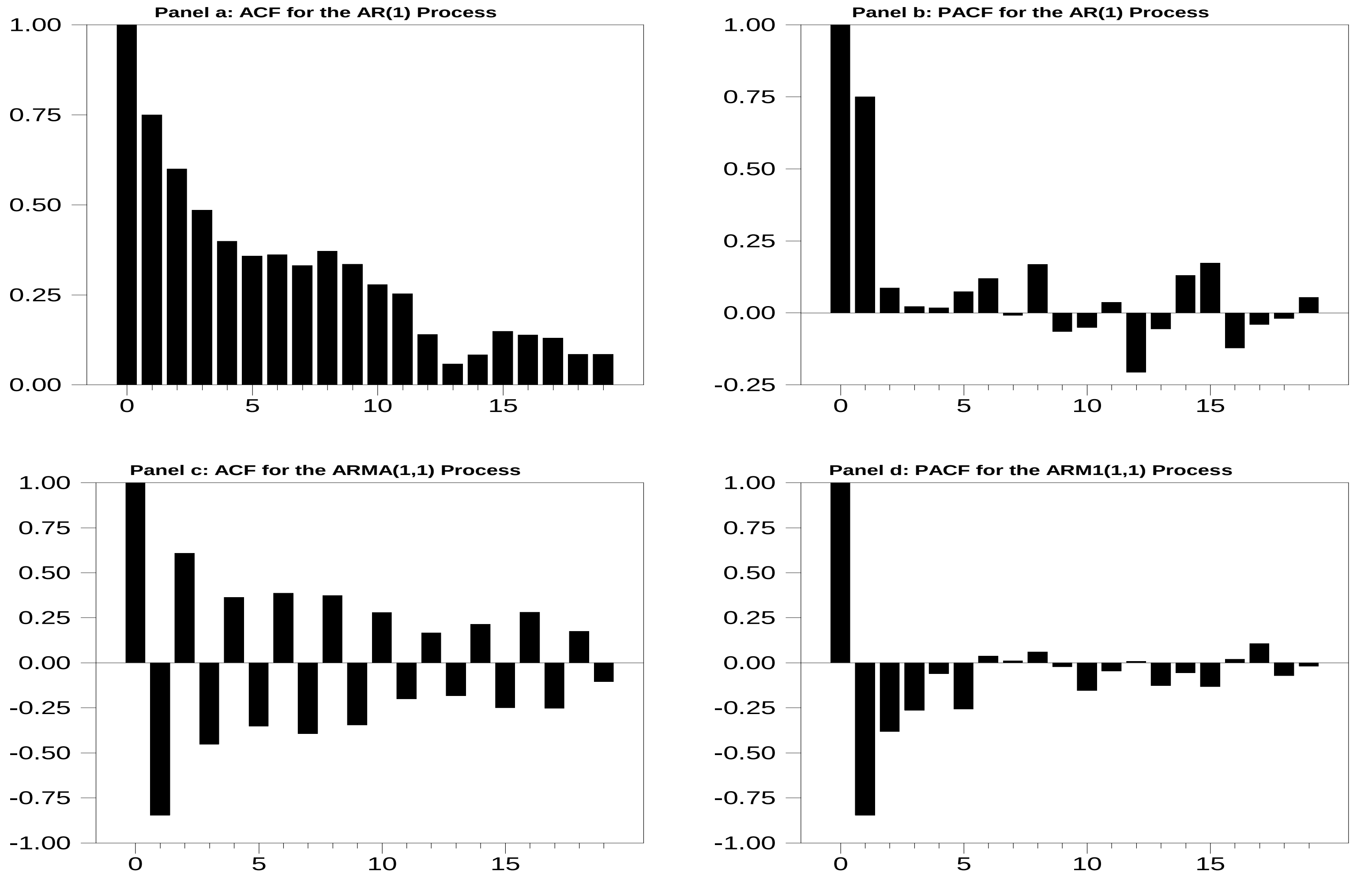


Table 2.2: Estimates of an AR(1) Model

	Model 1 $y_t = a_1 y_{t-1} + e_t$	Model 2 $y_t = a_1 y_{t-1} + e_t + b_{12} e_{t-12}$
Degrees of Freedom	98	97
Sum of Squared Residuals	85.10	85.07
Estimated a_1 (standard error)	0.7904 (0.0624)	0.7938 (0.0643)
Estimated b (standard error)		-0.0250 (0.1141)
AIC / SBC	AIC = 441.9 ; SBC = 444.5	AIC = 443.9 ; SBC = 449.1
Ljung-Box Q-statistics for the residuals (significance level in parentheses)	Q(8) = 6.43 (0.490) Q(16) = 15.86 (0.391) Q(24) = 21.74 (0.536)	Q(8) = 6.48 (0.485) Q(16) = 15.75 (0.400) Q(24) = 21.56. (0.547)

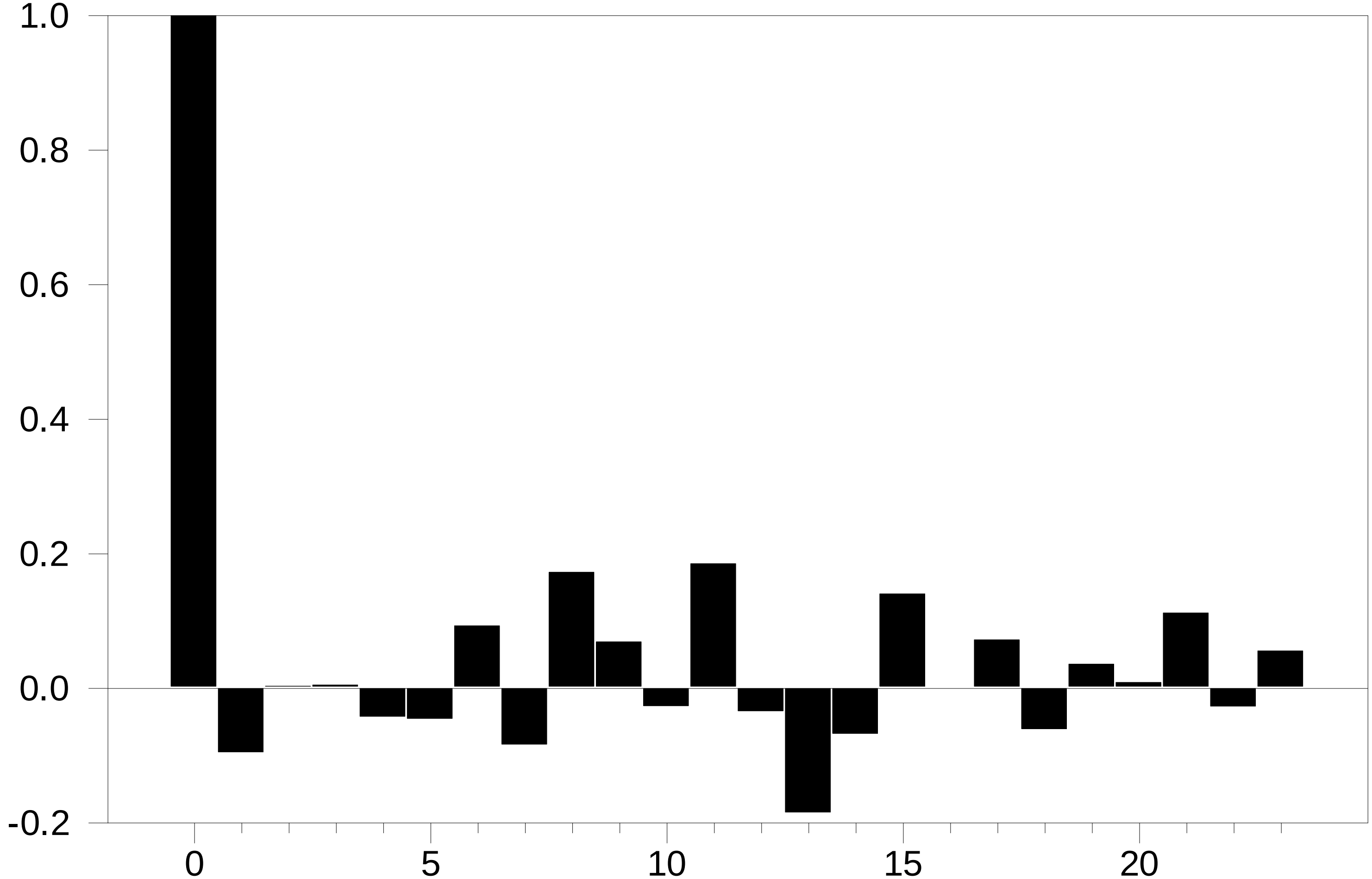
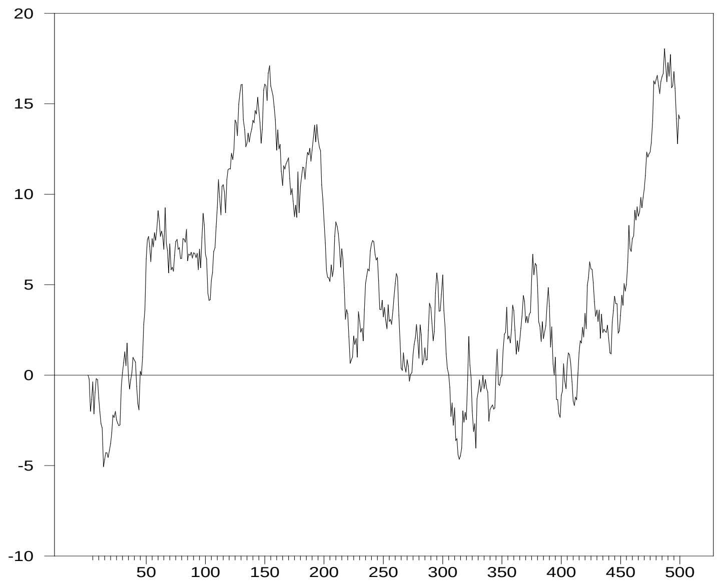


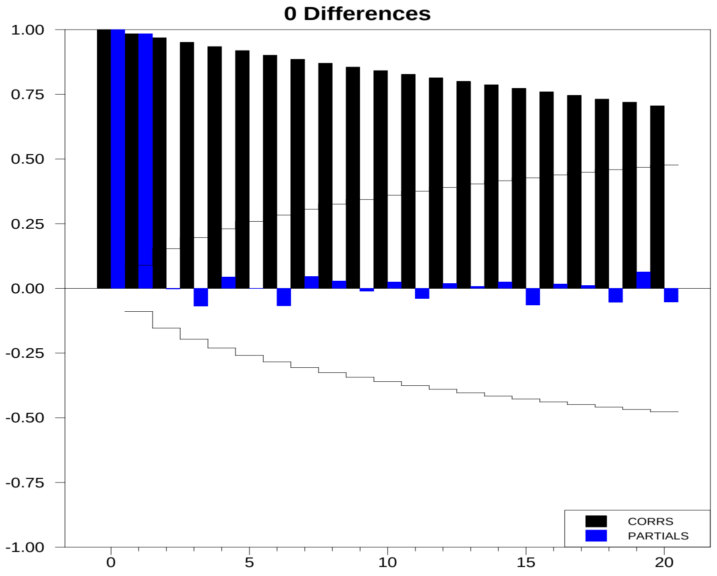
Figure 2.4: ACF of the Residuals from Model 1

Table 2.3: Estimates of an ARMA(1,1) Model

	Estimates	Q-Statistics	AIC / SBC
Model 1	a_1: -0.835 (.053)	Q(8) = 26.19 (.000) Q(24) = 41.10 (.001)	AIC = 496.5 SBC = 499.0
Model 2	a_1: -0.679 (.076) b_1: -0.676 (.081)	Q(8) = 3.86 (.695) Q(24) = 14.23 (.892)	AIC = 471.0 SBC = 476.2
Model 3	a_1: -1.16 (.093) a_2: -0.378 (.092)	Q(8) = 11.44 (.057) Q(24) = 22.59 (.424)	AIC = 482.8 SBC = 487.9

ACF of Nonstationary Series





Parsimony

Stationarity and Invertibility

Goodness of Fit

Post-Estimation Evaluation

8. BOX–JENKINS MODEL SELECTION

Box Jenkins Model Selection

- Parsimony
  - Extra AR coefficients reduce degrees of freedom by 2
  - Similar processes can be approximated by very different models
  - Common Factor Problem
    - $y_t = \varepsilon_t$ and $y_t = 0.5\, y_{t-1} + \varepsilon_t - 0.5\varepsilon_{t-1}$
  - Hence: All t-stats should exceed 2.0

  - Model should have a good fit as measured by AIC or BIC (SBC)

Box-Jenkins II

- Stationarity and Invertibility
  - t-stats, ACF, Q-stats, … all assume that the process is stationary
  - Be suspicious of implied roots near the unit circle
  - Invertibility implies the model has a finite AR representation.
    - No unit root in MA part of the model
- Diagnostic Checking
  - Plot residuals—look for outliers, periods of poor fit
  - Residuals should be serially uncorrrelated
    - Examine ACF and PACF of residuals
  - Overfit the model
  - Divide sample into subperiods
  - $F = (ssr - ssr_1 - ssr_2)/(p+q+1) / (ssr_1 + ssr_2)/(T\text{-}2p\text{-}2q\text{-}2)$

Residuals Plot

What can we learn by plotting the residuals?
What if there is a systematic pattern in the residuals?

Requirements for Box-Jenkins

- Successful in practice, especially short term forecasts
- Good forecasts generally require at least 50 observations
  - more with seasonality
- Most useful for short-term forecasts
- You need to ‘detrend’ the data.
- Disadvantages
  - Need to rely on individual judgment
    - However, very different models can provide nearly identical forecasts

Higher-Order Models

Forecast Evaluation

The Granger–Newbold Test

The Diebold–Mariano Test

9. PROPERTIES OF FORECASTS

Forecasting with ARMA Models

The MA(1) Model

$$y_t = \beta_0 + \beta_1\varepsilon_{t-1} + \varepsilon_t$$

Updating 1 period:

$$y_{t+1} = \beta_0 + \beta_1\varepsilon_t + \varepsilon_{t+1}$$

Hence, the optimal 1-step ahead forecast is:

$$E_t y_{t+1} = \beta_0 + \beta_1\varepsilon_t$$

Note: $E_t y_{t+j}$ is a short-hand way to write the conditional expectation of y_{t+j}

The 2-step ahead forecast is:

$$E_t y_{t+2} = E_t[\ \beta_0 + \beta_1\varepsilon_{t+1} + \varepsilon_{t+2}\ ] = \beta_0$$

Similarly, the n-step ahead forecasts are all β_0

Forecast errors

The 1-step ahead forecast error is:

$y_{t+1} - E_t y_{t+1} = \beta_0 + \beta_1\varepsilon_t + \varepsilon_{t+1} - \beta_0 - \beta_1\varepsilon_t = \varepsilon_{t+1}$

Hence, the 1-step ahead forecast error is the "unforecastable" portion of y_{t+1}

The 2-step ahead forecast error is:

$y_{t+2} - E_t y_{t+2} = \beta_0 + \beta_1\varepsilon_{t+1} + \varepsilon_{t+2} - \beta_0 = \beta_1\varepsilon_{t+1} + \varepsilon_{t+2}$

Forecast error variance

The variance of the 1-step ahead forecast error is: $var(\varepsilon_{t+1}) = \sigma^2$

The variance of the 2-step ahead forecast error is: $var(\beta_1\varepsilon_{t+1} + \varepsilon_{t+2}) = (1 + \beta_1^2)\sigma^2$

Confidence intervals

The 95% confidence interval for the 1-step ahead forecast is:

$\beta_0 + \beta_1\varepsilon_t \pm 1.96\sigma$

The 95% confidence interval for the 2-step ahead forecast is:

$\beta_0 \pm 1.96(1 + \beta_1^2)^{1/2}\sigma$

In the general case of an MA(q), the confidence intervals increase up to lag q.

The AR(1) Model: $y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$.

Updating 1 period, $y_{t+1} = a_0 + a_1 y_t + \varepsilon_{t+1}$, so that

$$E_t y_{t+1} = a_0 + a_1 y_t \qquad [\,*\,]$$

The 2-step ahead forecast is:

$$E_t y_{t+2} = a_0 + a_1 E_t y_{t+1}$$

and using [*]

$$E_t y_{t+2} = a_0 + a_0 a_1 + a_1^2 y_t$$

It should not take too much effort to convince yourself that:

$$E_t y_{t+3} = a_0 + a_0 a_1 + a_0 a_1^2 + a_1^3 y_t$$

and in general:

$$E_t y_{t+j} = a_0[\,1 + a_1 + a_1^2 + \ldots + a_1^{j-1}\,] + a_1^j y_t$$

If we take the limit of $E_t y_{t+j}$ we find that $E_t y_{t+j} = a_0/(1 - a_1)$. This result is really quite general; *for any stationary ARMA model, the conditional forecast of y_{t+j} converges to the unconditional mean.*

Forecast errors

The 1-step ahead forecast error is:

$$y_{t+1} - E_t y_{t+1} = a_0 + a_1 y_t + \varepsilon_{t+1} - a_0 - a_1 y_t = \varepsilon_{t+1}$$

The 2-step ahead forecast error is: $y_{t+2} - E_t y_{t+2}$. Since $y_{t+2} = a_0 + a_1 a_0 + a_1^2 y_t + \varepsilon_{t+2} + a_1 \varepsilon_{t+1}$ and $E_t y_{t+2} = a_0 + a_1 a_0 + a_1^2 y_t$, it follows that:

$$y_{t+2} - E_t y_{t+2} = \varepsilon_{t+2} + a_1 \varepsilon_{t+1}$$

Continuing is this fashion, the j-step ahead forecast error is :

$$y_{t+j} - E_t y_{t+j} = \varepsilon_{t+j} + a_1 \varepsilon_{t+j-1} + a_1^2 \varepsilon_{t+j-2} + a_1^3 \varepsilon_{t+j-3} + ... + a_1^{j-1} \varepsilon_{t+1}$$

Forecast error variance: The j-step ahead forecast error variance is:

$$\sigma^2[1 + a_1^2 + a_1^4 + a_1^6 + ... + a_1^{2(j-1)}]$$

The variance of the forecast error is an increasing function of j. As such, you can have more confidence in short-term forecasts than in long-term forecasts. In the limit the forecast error variance converges to $\sigma^2/(1-a_1^2)$; hence, the forecast error variance converges to the unconditional variance of the $\{y_t\}$ sequence.

Confidence intervals

- The 95% confidence interval for the 1-step ahead forecast is:

$$a_0 + a_1 y_t \pm 1.96\sigma$$

- Thus, the 95% confidence interval for the 2-step ahead forecast is:

$$a_0(1+a_1) + a_1^2 y_t \pm 1.96\sigma(1+a_1^2)^{1/2}.$$

Forecast Evaluation

- **Out-of-sample Forecasts**:

  1. Hold back a portion of the observations from the estimation process and estimate the alternative models over the shortened span of data.

  2. Use these estimates to forecast the observations of the holdback period.

  3. Compare the properties of the forecast errors from the two models.

- **Example**:

  1. If $\{y_t\}$ contains a total of 150 observations, use the first 100 observations to estimate an AR(1) and an MA(1) and use each to forecast the value of y_{101}. Construct the forecast error obtained from the AR(1) and from the MA(1).

  2. Reestimate an AR(1) and an MA(1) model using the first 101 observations and construct two more forecast errors.

  3. Continue this process so as to obtain two series of one-step ahead forecast errors, each containing 50 observations.

- A regression based method to assess the forecasts is to use the 50 forecasts from the AR(1) to estimate an equation of the form

$$y_{100+t} = a_0 + a_1 f_{1t} + v_{1t}$$

- If the forecasts are unbiased, an *F*-test should allow you to impose the restriction $a_0 = 0$ and $a_1 = 1$. Repeat the process with the forecasts from the MA(1). In particular, use the 50 forecasts from the MA(1) to estimate

$$y_{100+t} = b_0 + b_1 f_{2t} + v_{2t} \qquad t = 1, \ldots, 50$$

- If the significance levels from the two *F*-tests are similar, you might select the model with the smallest residual variance; that is, select the AR(1) if $\text{var}(v_{1t}) < \text{var}(v_{2t})$.

- Instead of using a regression-based approach, many researchers would select the model with the smallest mean square prediction error (MSPE). If there are *H* observations in the holdback periods, the MSPE for the AR(1) can be calculated as

$$MSPE = \frac{1}{H}\sum_{i=1}^{H} e_{1i}^2$$

The Diebold–Mariano Test

Let the loss from a forecast error in period i be denoted by $g(e_i)$. In the typical case of mean-squared errors, the loss is e_t^2

We can write the differential loss in period i from using model 1 versus model 2 as $d_i = g(e_{1i}) - g(e_{2i})$. The mean loss can be obtained as

$$\bar{d} = \frac{1}{H}\sum_{i=1}^{H}\left[g(e_{1i}) - g(e_{2i})\right]$$

If the $\{d_i\}$ series is serially uncorrelated with a sample variance of γ_0, the estimate of var($\bar{d}$) is simply $\gamma_0/(H-1)$. The expression

$$\bar{d} / \sqrt{\gamma_0 /(H-1)}$$

has a t-distribution with $H - 1$ degrees of freedom

Out-of-Sample Forecasts

10. A MODEL OF THE INTEREST RATE SPREAD

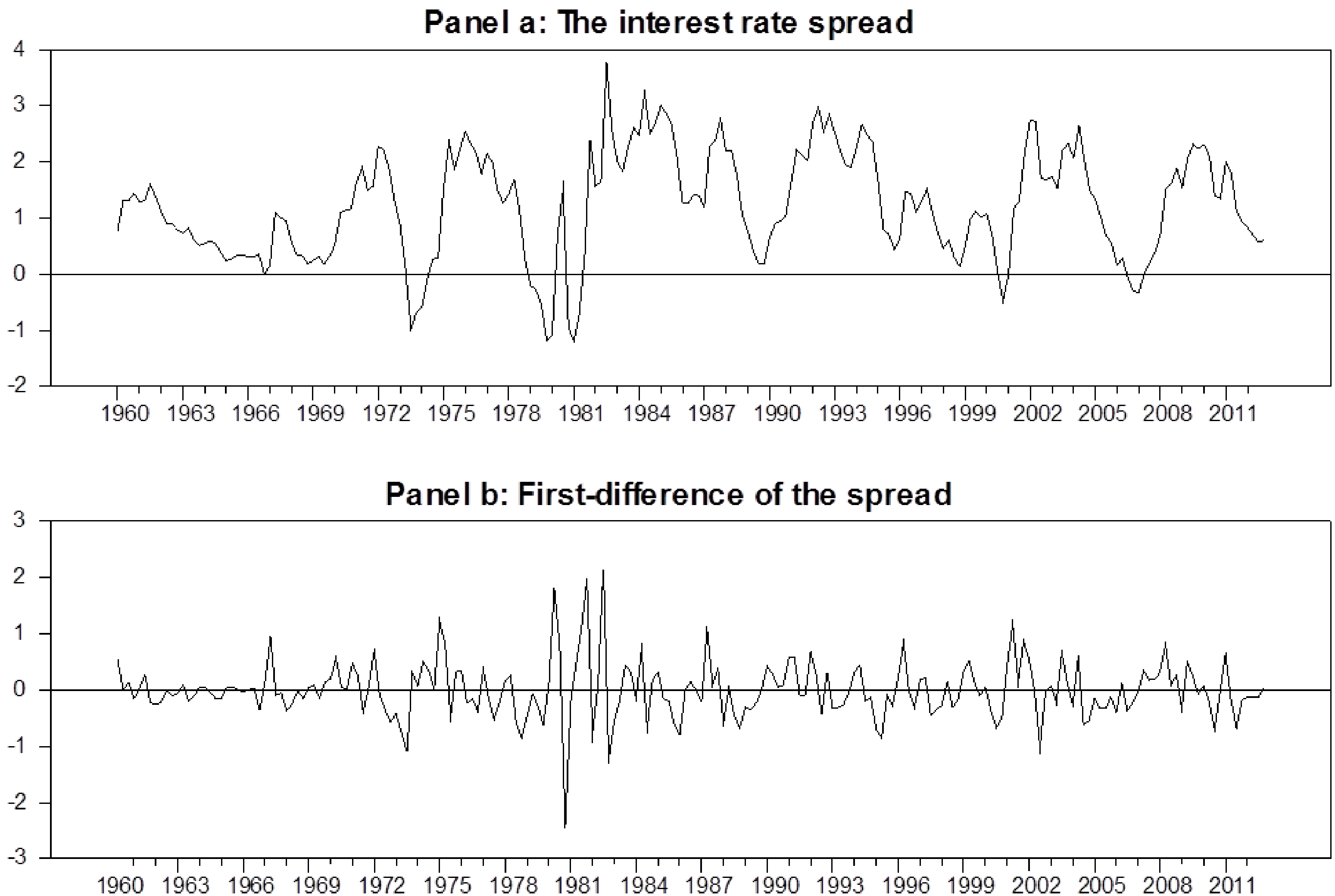


Figure 2.5: Time Path of the Interest Rate Spread

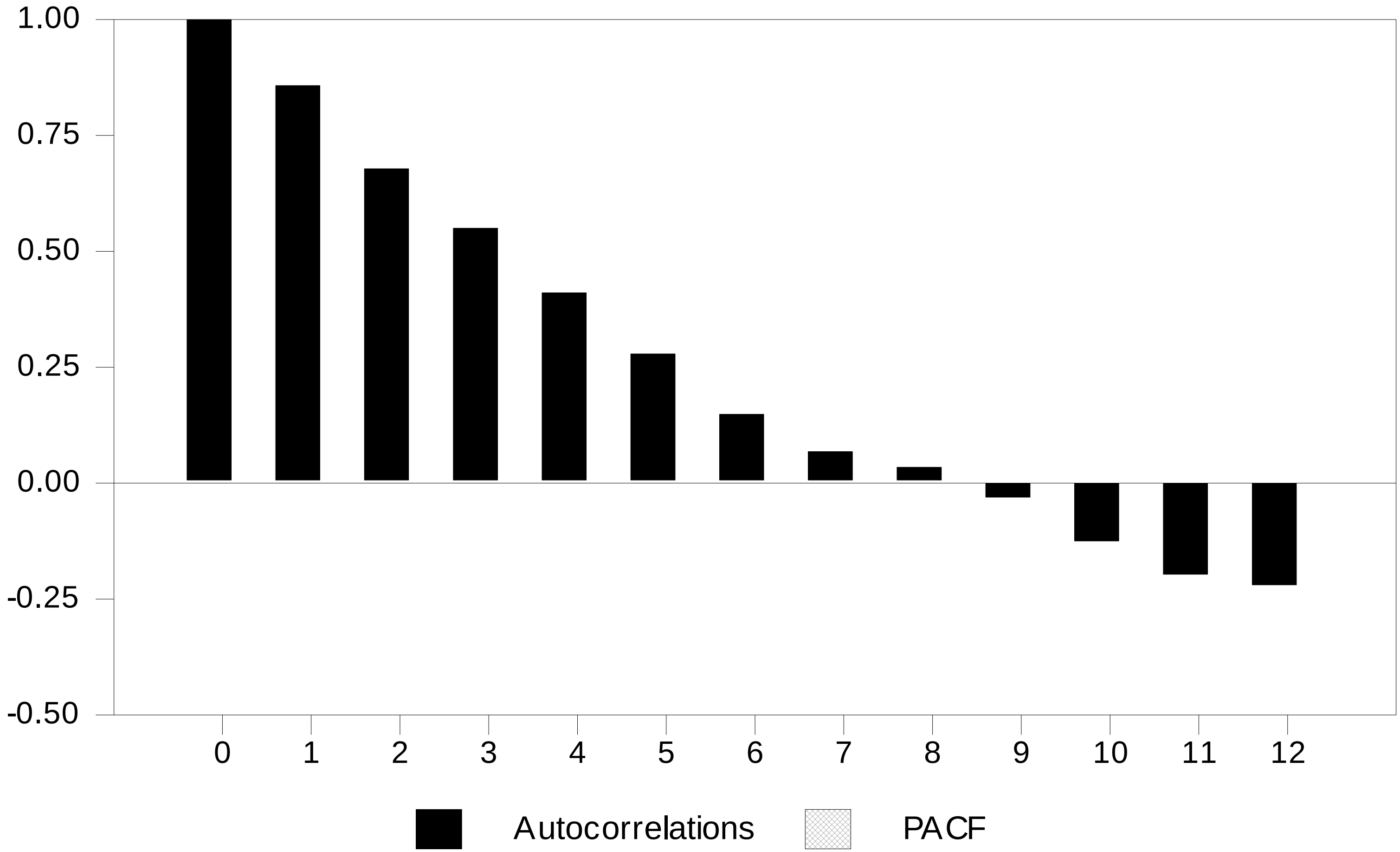


Figure 2.6: ACF and PACF of the Spread

Table 2.4: Estimates of the Interest Rate Spread

	AR(7)	AR(6)	AR(2)	p = 1, 2, and 7	ARMA(1, 1)	ARMA(2, 1)	p = 2 ma = (1, 7)
μ_y	1.20	1.20	1.19	1.19	1.19	1.19	1.20
	(6.57)	(7.55)	(6.02)	(6.80)	(6.16)	(5.56)	(5.74)
a_1	1.11	1.09	1.05	1.04	0.76	0.43	0.36
	(15.76)	(15.54)	(15.25)	(14.83)	(14.69)	(2.78)	(3.15)
a_2	-0.45	-0.43	-0.22	-0.20		0.31	0.38
	(-4.33)	(-4.11)	(-3.18)	(-2.80)		(2.19)	(3.52)
a_3	0.40	0.36					
	(3.68)	(3.39)					
a_4	-0.30	-0.25					
	(-2.70)	(-2.30)					
a_5	0.22	0.16					
	(2.02)	(1.53)					
a_6	-0.30	-0.15					
	(-2.86)	(-2.11)					
a_7	0.14			-0.03			
	(1.93)			(-0.77)			
β_1					0.38	0.69	0.77
					(5.23)	(5.65)	(9.62)
β_7							-0.14
							(-3.27)
SSR	43.86	44.68	48.02	47.87	46.93	45.76	43.72
AIC	791.10	792.92	799.67	801.06	794.96	791.81	784.46
SBC	817.68	816.18	809.63	814.35	804.93	805.10	801.07
Q(4)	0.18	0.29	8.99	8.56	6.63	1.18	0.76
Q(8)	5.69	10.93	21.74	22.39	18.48	12.27	2.60
Q(12)	13.67	16.75	29.37	29.16	24.38	19.14	11.13

Models of Seasonal Data

Seasonal Differencing

11. SEASONALITY

Seasonality in the Box-Jenkins framework

- Seasonal AR coefficients
  - $y_t = a_1 y_{t-1} + a_{12} y_{t-12} + a_{13} y_{t-13}$
  - $y_t = a_1 y_{t-1} + a_{12} y_{t-12} + a_1 a_{12} y_{t-13}$
  - $(1 - a_1 L)(1 - a_{12} L^{12}) y_t$

- Seasonal MA Coefficients

- Seasonal differencing:
  - $\Delta y_t = y_t - y_{t-1}$ versus $\Delta^{12} y_t = y_t - y_{t-12}$
    - NOTE: You do not difference 12 times
  - In RATS you can use: dif(sdiffs=1) y / sdy

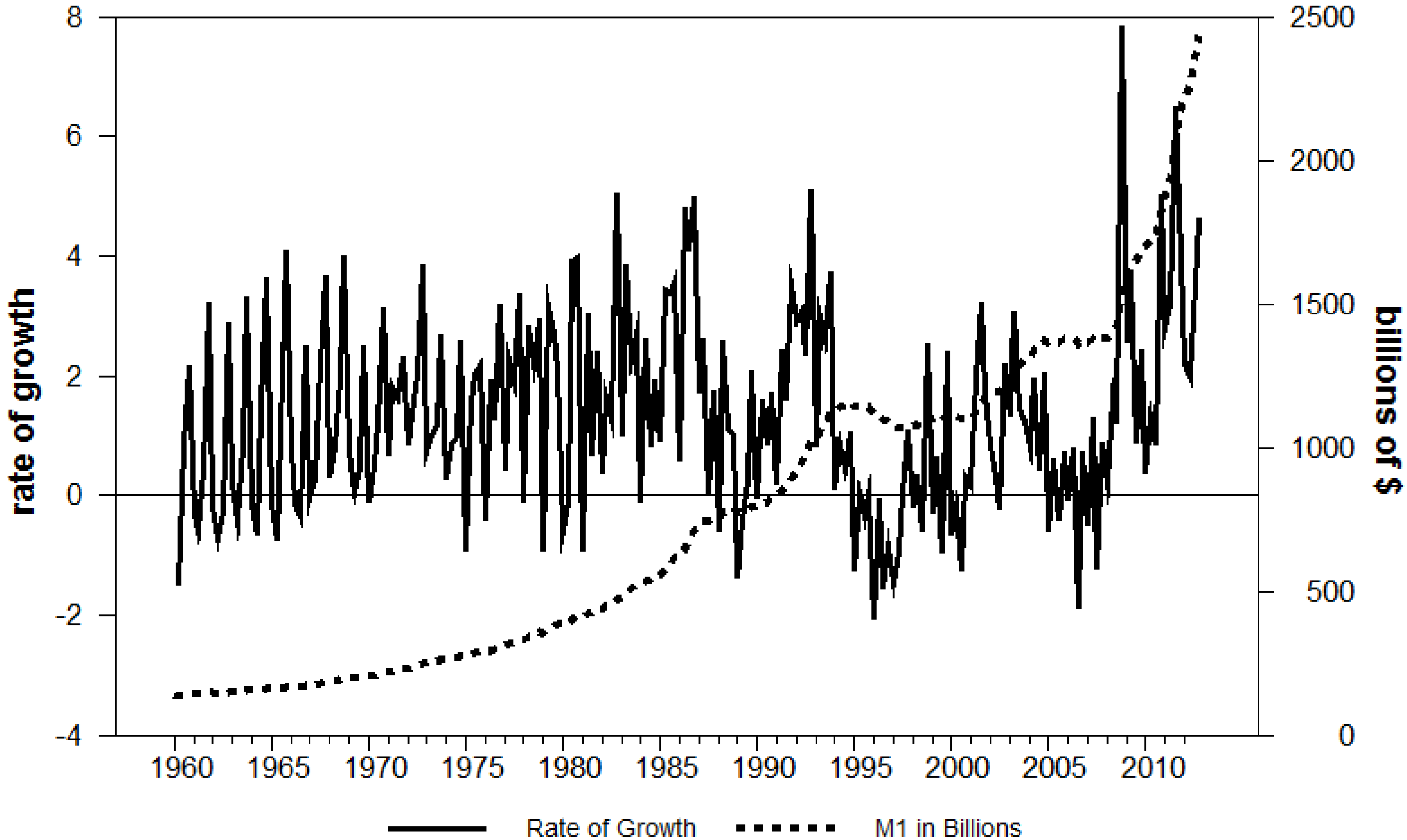


Figure 2.7: The Level and Growth Rate of M1

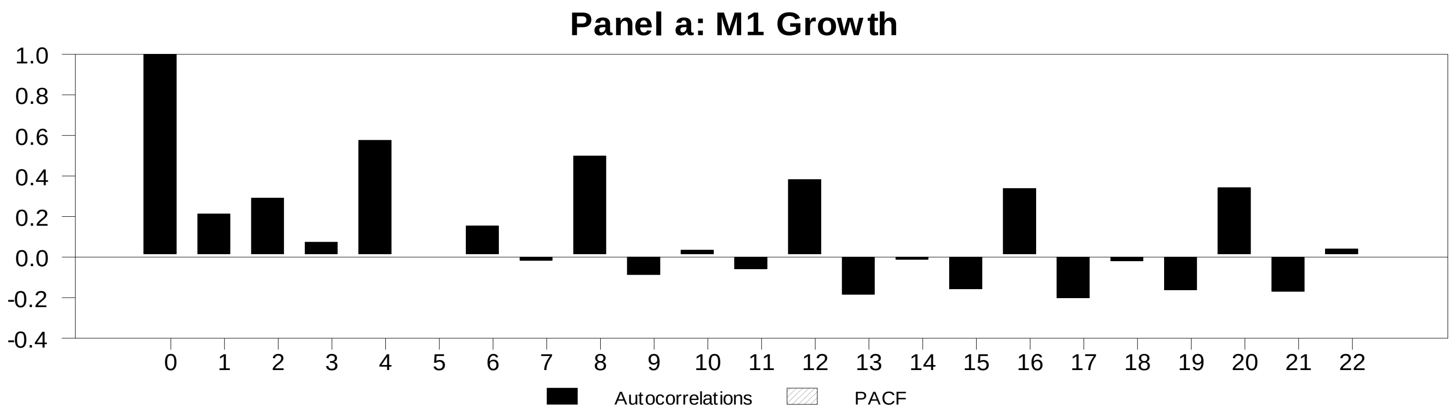


Figure 2.8: ACF and PACF

Three Models of Money growth

Model 1: AR(1) with Seasonal MA

$$m_t = a_0 + a_1 m_{t-1} + \varepsilon_t + \beta_4 \varepsilon_{t-4}$$

Model 2: Multiplicative Autoregressive

$$m_t = a_0 + (1 + a_1 L)(1 + a_4 L^4) m_{t-1} + \varepsilon_t$$

Model 3: Multiplicative Moving Average

$$m_t = a_0 + (1 + \beta_1 L)(1 + \beta_4 L^4) \varepsilon_t$$

Table 2.5 Three Models of Money Growth

	Model 1	Model 2	Model 3
a_1	0.541 (8.59)	0.496 (7.66)	
a_4		–0.476 (–7.28)	
β_1			0.453 (6.84)
β_4	–0.759 (–15.11)		–0.751 (–14.87)
SSR	0.0177	0.0214	0.0193
AIC SBC	–735.9 –726.2	–701.3 –691.7	–720.1 –710.4
Q(4) Q(8) Q(12)	1.39 (0.845) 6.34 (0.609) 14.34 (0.279)	3.97 (0.410) 24.21 (0.002) 32.75 (0.001)	22.19 (0.000) 30.41 (0.000) 42.55 (0.000)

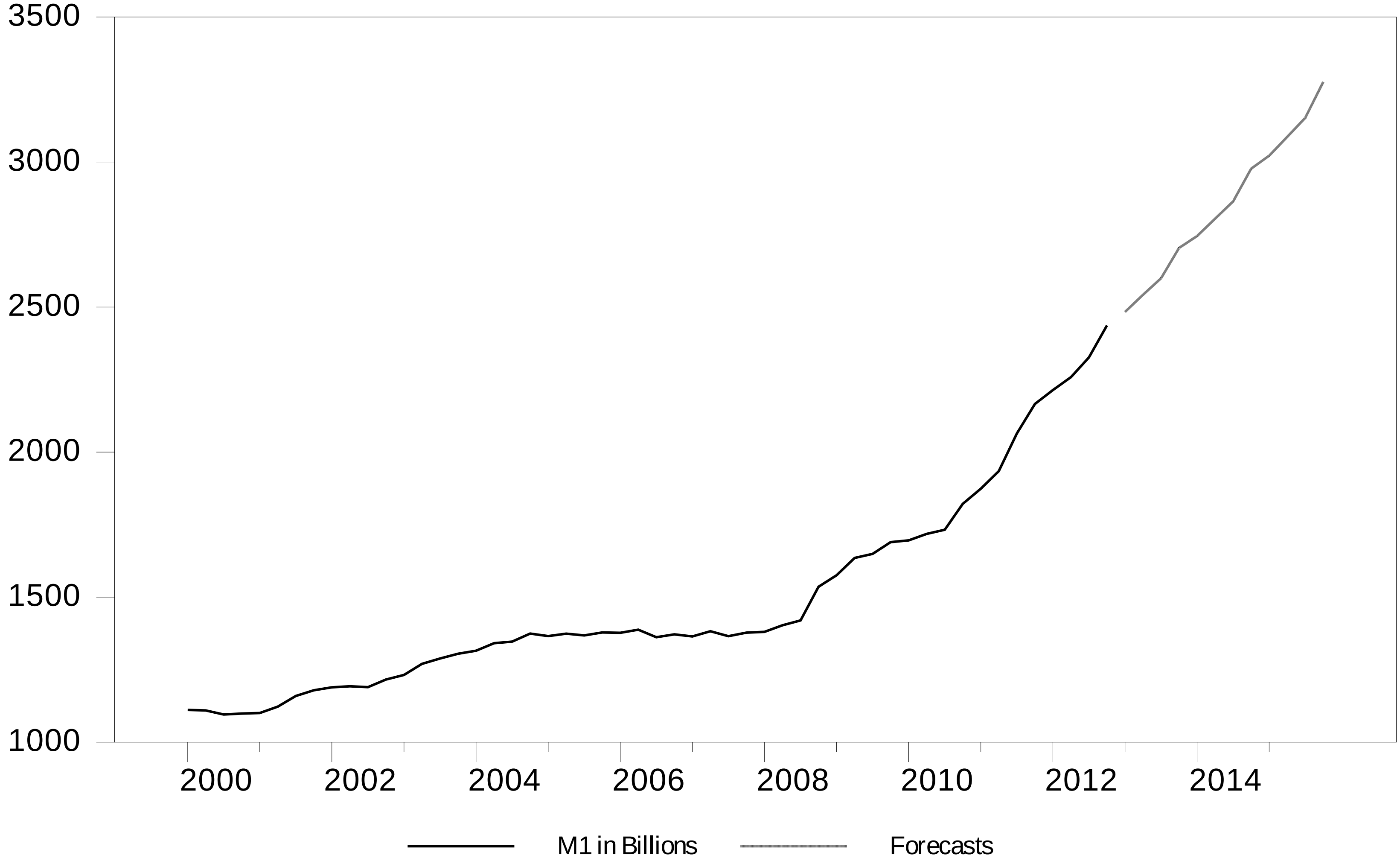


Figure 2.9: Forecasts of M1

Testing for Structural Change

Endogenous Breaks

Parameter Instability

An Example of a Break

12. PARAMETER INSTABILITY AND STRUCTURAL CHANGE

Parameter Instability and the CUSUMs

- Brown, Durbin and Evans (1975) calculate whether the cumulated sum of the forecast errors is statistically different from zero. Define:

$$CUSUM_N = \sum_{i=n}^{N} e_i(1)/\sigma_e \qquad N = n, \ldots, T-1$$

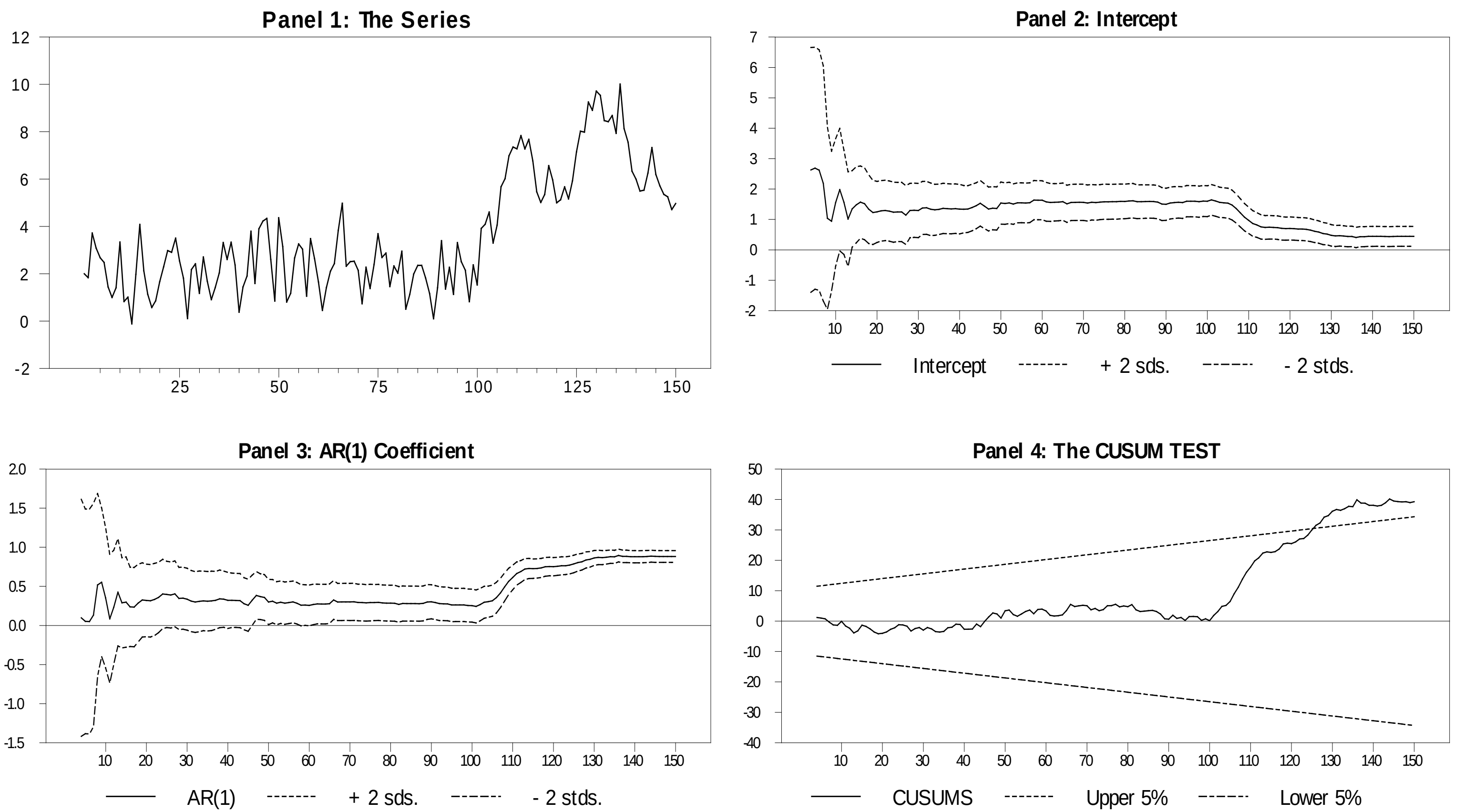
n = date of the first forecast error you constructed, σ_e is the estimated standard deviation of the forecast errors.

Example: With 150 total observations ($T = 150$), if you start the procedure using the first 10 observations ($n = 10$), 140 forecast errors ($T - n$) can be created. Note that σ_e is created using all $T - n$ forecast errors.

To create $CUSUM_{10}$, use the first ten observations to create $e_{10}(1)/\sigma_e$. Now let $N = 11$ and create $CUSUM_{11}$ as $[e_{10}(1)+e_{11}(1)]/\sigma_e$. Similarly, $CUSUM_{T\text{-}1} = [e_{10}(1)+\ldots+e_{T\text{-}1}(1)]/\sigma_e$.

If you use the 5% significance level, the plot value of each value of $CUSUM_N$ should be within a band of approximately $\pm\, 0.948\, [\, (T-n)^{0.5} + 2(N-n)\, (T-n)^{-0.5}\,]$.

Figure 2.10: Recursive Estimation of the Model



Section 13

COMBINING FORECASTS

13 Combining Forecasts

Consider the composite forecast f_{ct} constructed as weighted average of the individual forecasts

$$f_{ct} = w_1 f_{1t} + w_2 f_{2t} + \quad + w_n f_{nt} \qquad (2.71)$$

and $\sum w_i = 1$

If the forecasts are unbiased (so that $E_{t-1} f_{it} = y_t$), it follows that the composite forecast is also unbiased:

$$E_{t-1} f_{ct} = w_1 E_{t\text{-}1} f_{1t} + w_2 E_{t-1} f_{2t} + \quad + w_n E_{t-1} f_{nt}$$
$$= w_1 y_t + w_2 y_t + \quad + w_n y_t = y_t$$

A Simple Example

To keep the notation simple, let $n = 2$.

Subtract y_t from each side of (2.71) to obtain

$$f_{ct} - y_t = w_1(f_{1t} - y_t) + (1 - w_1)(f_{2t} - y_t)$$

Now let e_{1t} and e_{2t} denote the series containing the one-step-ahead forecast errors from models 1 and 2 (i.e., $e_{it} = y_t - f_{it}$) and let e_{ct} be the composite forecast error. As such, we can write

$$e_{ct} = w_1 e_{1t} + (1 - w_1)e_{2t}$$

The variance of the composite forecast error is

$$\text{var}(e_{ct}) = w_1{}^2\text{var}(e_{1t}) + (1 - w_1)^2\text{var}(e_{2t}) + 2w_1(1 - w_1)\text{cov}(e_{1t}e_{2t}) \quad (2.72)$$

Suppose that the forecast error variances are the same size and that $\text{cov}(e_{1t}e_{2t}) =0$. If you take a simple average by setting $w_1 = 0.5$, (2.72) indicates that the variance of the composite forecast is 25% of the variances of either forecast: $\text{var}(e_{ct}) = 0.25\text{var}(e_{1t}) = 0.25\text{var}(e_{2t})$.

Optimal Weights

$$\text{var}(e_{ct}) = (w_1)^2\text{var}(e_{1t}) + (1 - w_1)^2\text{var}(e_{2t}) + 2w_1(1 - w_1)\text{cov}(e_{1t}e_{2t})$$

Select the weight w_1 so as to minimize $\text{var}(e_{ct})$:

$$\frac{\delta\,\text{var}(e_{ct})}{\delta w_1} = 2w_1\,\text{var}(e_{1t}) - 2(1 - w_1)\,\text{var}(e_{2t}) + 2(1 - 2w_1)\,\text{cov}(e_{1t}e_{2t})$$

Bates and Granger (1969), recommend constructing the weights excluding the covariance terms.

$$w_1^* = \frac{\text{var}(e_{2t})}{\text{var}(e_{1t}) + \text{var}(e_{2t})} = \frac{\text{var}(e_{1t})^{-1}}{\text{var}(e_{1t})^{-1} + \text{var}(e_{2t})^{-1}}$$

In the n-variable case:

$$w_n^* = \frac{\text{var}(e_{1t})^{-1}}{\text{var}(e_{1t})^{-1} + \text{var}(e_{2t})^{-1} + \ldots + \text{var}(e_{nt})^{-1}}$$

Alternative methods

Consider the regression equation

$$y_t = \alpha_0 + \alpha_1 f_{1t} + \alpha_2 f_{2t} + \quad + \alpha_n f_{nt} + v_t \qquad (2.75)$$

It is also possible to force $\alpha_0 = 0$ and $\alpha_1 + \alpha_2 + \quad + \alpha_n = 1$. Under these conditions, the α_i's would have the direct interpretation of optimal weights.

Here, an estimated weight may be negative. Some researchers would reestimate the regression without the forecast associated with the most negative coefficient.

Granger and Ramanathan recommend the inclusion of an intercept to account for any bias and to leave the α_i's unconstrained.

As surveyed in Clemen (1989), not all researchers agree with the Granger–Ramanathan recommendation and a substantial amount of work has been conducted so as to obtain optimal weights.

The SBC

Let SBC_i be the SBC from model *i* and let SBC^* be the SBC from the best fitting model.

Form $\alpha_i = \exp[(SBC^* - SBC_i)/2]$ and then construct the weights

$$w_i^* = \alpha_i / \sum_{t=1}^{n} \alpha_i$$

Since exp(0) = 1, the model with the best fit has the weight $1/\Sigma\alpha_i$. Since α_i is decreasing in the value of SBC_i, models with a poor fit with have smaller weights than models with large values of the SBC.

Example of the Spread

I estimated seven different ARMA models of the interest rate spread. The data ends in April 2012 and if I use each of the seven models to make a one-step-ahead forecast for January 2013:

	AR(7)	AR(6)	AR(2)	AR(‖1,2,7‖)	ARMA(1,1)	ARMA(2,1)	ARMA(2,‖1,7‖)
$f_{i2013:1}$	0.775	0.775	0.709	0.687	0.729	0.725	0.799

Simple averaging of the individual forecasts results in a combined forecast of 0.743.

Construct 50 1-step-ahead out-of-sample forecasts for each model so as to obtain

	AR(7)	AR(6)	AR(2)	AR(‖1,2,7‖)	ARMA(1,1)	ARMA(2,1)	ARMA(2,‖1,7‖)
$var(e_{it})$	0.635	0.618	0.583	0.587	0.582	0.600	0.606
w_i	0.135	0.139	0.147	0.146	0.148	0.143	0.141

Next, use the spread (s_t) to estimate a regression in the form of (5). If you omit the intercept and constrain the weights to unity, you should obtain:

$$s_t = 0.55f_{1t} – 0.25f_{2t} −2.37f_{3t} + 2.44f_{4t} + 0.84f_{5t} – 0.28f_{6t} + 1.17f_{7t} \quad (6)$$

Although some researchers would include the negative weights in (6), most would eliminate those that are negative. If you successively reestimate the model by eliminating the forecast with the most negative coefficient, you should obtain:

$$s_t = 0.326f_{4t} + 0.170f_{5t} + 0.504f_{7t}$$

The composite forecast using the regression method is:

$$0.326(0.687) + 0.170(0.729) + 0.504(0.799) = 0.751.$$

If you use the values of the SBC as weights, you should obtain:

	AR(7)	AR(6)	AR(2)	AR(||1,2,7||)	ARMA(1,1)	ARMA(2,1)	ARMA(2,||1,7||)
w_i	0.000	0.000	0.011	0.001	0.112	0.103	0.773

The composite forecast using SBC weights is 0.782. In actuality, the spread in 2013:1 turned out to be 0.74 (the actual data contains only two decimal places). Of the four methods, simple averaging and weighting by the forecast error variances did quite well. In this instance, the regression method and constructing the weights using the SBC provided the worst composite forecasts.

APPENDIX 2.1: ML ESTIMATION OF A REGRESSION

$$\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-\varepsilon_t^2}{2\sigma^2}\right)$$

$$\prod_{t=1}^{T}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-\varepsilon_t^2}{2\sigma^2}\right)$$

$$\ln L = \frac{-T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon_t^2$$

Let $\varepsilon_t = y_t - bx_t$

$$\ln L = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - \beta x_t)^2$$

$$\frac{\partial \ln L}{\partial\sigma^2} = -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^{T}(y_t - \beta x_t)^2 \qquad \frac{\partial \ln \mathrm{L}}{\partial\beta} = \frac{1}{\sigma^2}\sum_{t=1}^{T}(y_t x_t - \beta x_t^2)$$

ML ESTIMATION OF AN MA(1)

Now let $y_t = \beta\varepsilon_{t-1} + \varepsilon_t$. The problem is to construct the $\{\varepsilon_t\}$ sequence from the observed values of $\{y_t\}$. If we knew the true value of β and knew that $\varepsilon_0 = 0$, we could construct $\varepsilon_1, \ldots, \varepsilon_T$ recursively. Given that $\varepsilon_0 = 0$, it follows that:

$$\varepsilon_1 = y_1$$
$$\varepsilon_2 = y_2 - \beta\varepsilon_1 = y_2 - \beta y_1$$
$$\varepsilon_3 = y_3 - \beta\varepsilon_2 = y_3 - \beta(y_2 - \beta y_1)$$
$$\varepsilon_4 = y_4 - \beta\varepsilon_3 = y_4 - \beta[y_3 - \beta(y_2 - \beta y_1)]$$

In general, $\varepsilon_t = y_t - \beta\varepsilon_{t-1}$ so that if L is the lag operator

$$\varepsilon_t = y_t/(1+\beta L) = \sum_{i=0}^{t-1}(-\beta)^i y_{t-i}$$

$$\ln L = \frac{-T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}(-\beta)^i y_{t-i}\right)^2$$