I. Stationarity and lag operator

When observing $(y_1, ..., y_T)$ we are are considering y_t as part of a stochastic process (letting t vary). An important class is *covariance stationary* (or weakly stationary or second-order stationary or just stationary) stochastic process:

Definition. $\{y_t, t = 0, \pm 1, ...\}$ is **covariance stationary** if for any t, s

$$Ey_t \equiv \mu_t = \mu < \infty, cov(y_t, y_s) \equiv \gamma(t,s) = \gamma(t-s) \text{ with } \mid \gamma(t-s) \mid < \infty.$$

This implies $var(y_t) = \gamma(0) < \infty$, constant and finite. The graph of $(u, \gamma(u))$ defines the *autocovariance function*. Often one uses the *autocorrelation function* (scale free, hereafter ACF): $\rho(u) = \frac{\gamma(u)}{\gamma(0)}$, $u = 0, \pm 1, ...$ where $-1 \leq \rho(u) \leq 1$.

Examples: Prices of stocks are not covariance stationary. However, returns are covariance stationary. Likewise, $y_t = a + bt$ is not covariance stationary.

Another important class: $\{y_t, t = 0, \pm 1, ...\}$ *is strictly stationary if for any* $t_1, ..., t_s$, *any integer* s, *the distribution of the* s*-dimensional (random) vector* $(y_{t_1}, .., y_{t_s})$ *coincide with* $(y_{t_1+h}, .., y_{t_s+h})$ *for any real* h.

Definition: White Noise. y_t is called a white noise when for any t

$$Ey_t = 0, \;\; \gamma(0) = \sigma^2 < \infty, \;\; \gamma(t-s) = 0 \;\; t \neq s.$$

Sample Counterparts. We have defined the population quantities. There are sample analogue. Given sample $(y_1, ..., y_T)$, with T sample size:

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} y_t \text{ sample mean,}$$

$$\hat{\gamma}(0) = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{\mu})^2 = \frac{1}{T}\sum_{t=1}^{T} y_t^2 - \hat{\mu}^2 \text{ sample variance,}$$

$$\hat{\gamma}(u) = \frac{1}{T}\sum_{t=1}^{T-u}(y_t - \hat{\mu})(y_{t+u} - \hat{\mu}) \text{ sample autocovariance,}$$

$$\hat{\rho}(u) = \frac{\hat{\gamma}(u)}{\hat{\gamma}(0)} \text{ sample autocorrelation,}$$

for $u = 1, .., T - 1$, setting $\hat{\gamma}(-u) = \hat{\gamma}(u)$.

Lag operator. We need a new concept. For any time series (need not be stochastic) the *lag operator* L is defined by the relation:

$$Ly_t = y_{t-1},$$

and in general for $m \geq 1$ integer, applying L m times, $L^m y_t = y_{t-m}$.

If apply L to a constant has no effect, for instance $L\,2 = 2$. The lag operator is commutative with the multiplication (by constant) operator and additive over summation operator:

$$L(2y_t) = 2Ly_t = 2y_{t-1} \quad \text{and} \quad L(y_t + x_t) = Ly_t + Lx_t = y_{t-1} + x_{t-1}.$$

So L follows the same algebraic rules of multiplicative operator. For example, for constant λ_1, λ_2

$$\begin{aligned}(1-\lambda_1 L)(1-\lambda_2 L)y_t \quad &= (1+\lambda_1\lambda_2 L^2 - \lambda_1 L - \lambda_2 L)y_t = (1-(\lambda_1+\lambda_2)L + \lambda_1\lambda_2 L^2)y_t \\ &= y_t - (\lambda_1+\lambda_2)y_{t-1} + \lambda_1\lambda_2 y_{t-2}.\end{aligned}$$

An expression like $\lambda_0 + \lambda_1 L + \lambda_2 L^2$ is referred to as a *polynomial in the lag operator* (in this case of order 2).

II. Linear processes: ARMA

An important class of stochastic processes is the *linear processes* :

$$y_t = \mu + \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + ... = \mu + \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i},$$

We say that y_t represents a *moving average* of the ϵ_t with coefficients ψ_i.

The sequence of (non-random and non-time varying) coefficients $\{\psi\}$ must satisfy $\sum_{i=1}^{\infty} \psi_i^2 < \infty$ (otherwise one has a process with infinite variance). μ is a constant coefficient and hereafter the ϵ_t is a white noise with mean zero and $E\epsilon_t^2 = \sigma^2 < \infty$.

By linearity $Ey_t = \mu + \sum_{i=0}^{\infty} \psi_i E\epsilon_{t-i} = \mu + 0 = \mu$ and by non-correlation of ε_t (variance of sum equals the sum of variances):

$$var(y_t) = var(\mu) + \sum_{i=0}^{\infty} \psi_i^2 var(\epsilon_{t-i}) = 0 + \sigma^2 \sum_{i=0}^{\infty} \psi_i^2 < \infty$$

Observe that both the expectation and variance of y_t do not depend on t.

Let us work out the autocovariance function. Take $u > 0$.

$$E(y_t - \mu)(y_{t+u} - \mu) = \sum_{i=0}^{\infty} \psi_i \sum_{j=0}^{\infty} \psi_j E\epsilon_{t-i}\epsilon_{t+u-j}$$

$$= \sigma^2 \sum_{j=0}^{\infty} \psi_{j+u}\psi_j \equiv \gamma(u),$$

because

$$E\epsilon_{t-i}\epsilon_{t+u-j} = \begin{cases} E\epsilon_{t-i}^2 = \sigma^2, & t-i = t+u-j, \\ 0 & t-i \neq t+u-j, \end{cases}$$

where $t - i = t + u - j$ equivalent to $i = j + u$. This is a *very general expression*, valid for *any* sequence of *square summable* ψ_i. We will take some special cases.

Observe that the expectation, variance, and autocovariance function of y_t do not on t. Hence, y_t is a covariance stationary process.

MA(q) processes. A Moving Average of order q (MA(q)) with integer q is defined by

$$y_t = \epsilon_t + \alpha_1 \epsilon_{t-1} + \alpha_2 \epsilon_{t-2} + ... + \alpha_q \epsilon_{t-q},$$

where the $\alpha_1, ..., \alpha_q$ are constant coefficients.

Special case of linear process: $\psi_0 = 1$, $\psi_i = \alpha_i$ for $1 \leq i \leq q$ and $\psi_i = 0$ for $i > q$.

The square summability is satisfied and mean constant (zero) so MA(q) is stationary. By using the previous formula for the variance and ACF, one has

$$\gamma(0) = \sigma^2(1+\alpha_1^2+...+\alpha_q^2), \;\; \gamma(u) = \sigma^2 \sum_{j=0}^{q-u} \alpha_j \alpha_{j+u} \text{ for } 1 \leq u \leq q \text{ and } \;\gamma(u) = 0 \text{ for } u > q.$$

Hence, the ACF is truncated, i.e., MA have little *memory*. Observe that the stationarity is achieved for *any* value of coefficients $\alpha_1, ..., \alpha_q$.

An important case is the MA(1) plus a constant:

$$y_t = \mu + \epsilon_t + \alpha_1 \epsilon_{t-1}.$$

One has $E[y_t] = \mu$, $Var[y_t] = \sigma_\varepsilon^2(1 + \alpha_1^2)$, $\rho(1) = \frac{\alpha_1}{1+\alpha_1^2}$ and $\rho(u) = 0 \;\; u = \pm 2, \pm 3, ...$

Note that if replace α_1 by $1/\alpha_1$ in $\rho(1)$ we get

$$\frac{1/\alpha_1}{1+1/\alpha_1^2} = \frac{1/\alpha_1}{\frac{\alpha_1^2+1}{\alpha_1^2}} = \frac{\alpha_1}{\alpha_1^2+1} = \rho(1),$$

so there is no difference, i.e., one can not identify α_1. We will assume that $|\alpha_1| < 1$. This assumption is linked to the *invertibility* condition, meaning the possibility to 'invert' the process and be able to write y_t as a linear function of past values y_{t-j} with $j = 1, 2, ...$ More precisely, observe that (without loss of generality, assume $\mu = 0$)

$$\begin{aligned} y_t &= \varepsilon_t + \alpha_1\varepsilon_{t-1} = \varepsilon_t + \alpha_1(y_{t-1} - \alpha_1\varepsilon_{t-2}) = \varepsilon_t + \alpha_1 y_{t-1} - \alpha_1^2\varepsilon_{t-2} \\ &= \varepsilon_t + \alpha_1 y_{t-1} - \alpha_1^2(y_{t-2} - \alpha_1\varepsilon_{t-3}) = \varepsilon_t + \alpha_1 y_{t-1} - \alpha_1^2 y_{t-2} + \alpha_1^3\varepsilon_{t-3} = ... \\ &= \varepsilon_t - \sum_{i=1}^{\infty}(-\alpha_1)^i y_{t-i} \quad \text{when} \quad |\alpha_1| < 1. \end{aligned}$$

Observe that in the previous calculation we inverted $1 + \alpha_1 L$ given that one has

$$\varepsilon_t = (1{+}\alpha_1 L)^{-1} y_t = \frac{1}{1-(-\alpha_1 L)} y_t = \left(\sum_{i=0}^{\infty}(-\alpha_1 L)^i\right) y_t = \sum_{i=0}^{\infty}(-\alpha_1)^i L^i y_t = \sum_{i=0}^{\infty}(-\alpha_1)^i y_{t-i}.$$

Extension of invertibility of MA(q) is based on looking at roots of polynomial $\alpha(L) = 1 + \alpha_1 L + ... + \alpha_q L^q = 0$. The associated MA is invertible when all the roots are greater than one in modulus. A consequence is that $Cov[\varepsilon_t, y_{t-i}] = 0$,when $i > 0$.

AR(p) processes. The previous inversion suggest to consider Autoregressive process of order p (AR(p)) with integer p:

$$y_t = \mu + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \epsilon_t,$$

where $\phi_1, ..., \phi_p$ are constant autoregressive coefficients.

The simplest case is AR(1): $\quad y_t = \mu + \phi_1 y_{t-1} + \epsilon_t.$

Observe that

$$\begin{aligned} y_t &= \mu + \phi_1(\mu + \phi y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \mu(1+\phi_1) + \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 y_{t-2} = ... \\ &= \mu(1 + \phi_1 + ... + \phi_1^{t-1}) + \varepsilon_t + \phi_1 \varepsilon_{t-1} + ... + \phi_1^{t-1}\varepsilon_1 + \phi_1^t y_0. \end{aligned}$$

Hence, the process will be explosive when $|\phi_1| > 1$. Likewise, when $\phi_1 = 1$ is non-stationary (see below). We will therefore assume that $|\phi_1| < 1$. Under this assumption, one has the MA(∞) form

$$y_t = \frac{\mu}{1-\phi_1} + \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i} = \frac{\mu}{1-\phi_1} + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \;\; \psi_i = \phi_1^i.$$

Given that $|\phi_1| < 1$, one has $\sum_{i=0}^{\infty} \psi_i^2 = 1/(1-\phi_1^2) < +\infty$, i.e., the process y_t is stationary. Likewise, one can show that an AR(p) process is stationary when all the roots of the polynomial $(1 - \phi_1 L - \phi_2 L^2 ... - \phi_p L^p)$ are greater than one in modulus.

An important property for AR(p) processes is that $Cov[\varepsilon_t, y_{t-i}] = 0$ when $i > 0$.

One can derive the moments of an AR(p)process. For instance, when one considers an AR(1) process, one has

$$E[y_t] = \frac{\mu}{1-\phi_1},\ Var[y_t] = \frac{1}{1-\phi_1^2}\sigma_\varepsilon^2,\ \rho(u) = \phi_1^u.$$

ARMA(p, q).

Autoregressive moving average process of order p, q (AR(p, q)) with integer p, q:

$$y_t = \mu + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \epsilon_t + \alpha_1 \epsilon_{t-1} + ... + \alpha_q \epsilon_{t-q}.$$

The process is *stationary* if all the roots of

$$1 - \phi_1 L - ... - \phi_p L^p = 0$$

are greater than one in modulus. It will be *invertible* if all the roots of

$$1 + \alpha_1 L + ... + \alpha_q L^q = 0$$

are greater than one in modulus. We will make these two assumptions.

With ARMA models, one must be careful to rule out the possibility of equal roots in the AR and MA component.

An important property for ARMA processes is that $Cov[\varepsilon_t, y_{t-i}] = 0$ when $i > 0$.

Moments of ARMA processes. Let us take the ARMA(1,1) example:

$$y_t = \mu + \phi_1 y_{t-1} + \epsilon_t + \alpha_1 \epsilon_{t-1},$$

We will often use the stationarity of the process. One has

$$E[y_t] = \mu + \phi_1 E[y_{t-1}] + E[\epsilon_t] + \alpha_1 E[\epsilon_{t-1}] = \mu + \phi_1 E[y_t] + 0 + 0,$$

i.e.,

$$E[y_t] = \frac{\mu}{1 - \phi_1}.$$

$$\begin{aligned} Var[y_t] &= \phi_1^2 Var[y_{t-1}] + Var[\epsilon_t] + \alpha_1^2 Var[\epsilon_{t-1}] \\ &\quad + 2\phi_1 Cov[y_{t-1}, \varepsilon_t] + 2\phi_1\alpha_1 Cov[y_{t-1}, \varepsilon_{t-1}] + 2\alpha_1 Cov[\varepsilon_{t-1}, \varepsilon_t] \\ &= \phi_1^2 Var[y_t] + (1 + \alpha_1^2)\sigma_\epsilon^2 + 0 + 2\phi_1\alpha_1 Cov[\varepsilon_{t-1} + \phi_1 y_{t-2} + \alpha_1 \varepsilon_{t-2}, \varepsilon_{t-1}] + 0 \\ &= \phi_1^2 Var[y_t] + (1 + \alpha_1^2)\sigma_\epsilon^2 + 2\phi_1\alpha_1\sigma_\epsilon^2. \end{aligned}$$

Hence,

$$Var[y_t] = \sigma_\epsilon^2 \frac{1 + \alpha_1^2 + 2\phi_1\alpha_1}{1 - \phi_1^2}.$$

Likewise,

$$
\begin{aligned}
Cov[y_t, y_{t-1}] &= Cov[\phi_1 y_{t-1} + \alpha_1 \varepsilon_{t-1} + \varepsilon_t, y_{t-1}] \\
&= \phi_1 Var[y_t] + \alpha_1 Cov[\varepsilon_{t-1}, \varepsilon_{t-1}] + Cov[\varepsilon_t, y_{t-1}] \\
&= \phi_1 Var[y_t] + \alpha_1 \sigma_\varepsilon^2, \\
&= \sigma_\varepsilon^2 \left(\phi_1 \frac{1 + \alpha_1^2 + 2\phi_1 \alpha_1}{1 - \phi_1^2} + \alpha_1 \right).
\end{aligned}
$$

Finally, for $i \geq 2$

$$
\begin{aligned}
Cov[y_t, y_{t-i}] &= Cov[\phi_1 y_{t-1} + \alpha_1 \varepsilon_{t-1} + \varepsilon_t, y_{t-i}] \\
&= \phi_1 Cov[y_{t-1}, y_{t-i}] + \alpha_1 Cov[\varepsilon_{t-1}, y_{t-i}] + Cov[\varepsilon_t, y_{t-i}] \\
&= \phi_1 Cov[y_{t-1}, y_{t-i}] + 0 + 0 \\
&= \phi_1^{i-1} Cov[y_{t-i+1}, y_{t-i}] \\
&= \phi_1^{i-1} Cov[y_t, y_{t-1}].
\end{aligned}
$$

So the ACF is a function of both AR and MA coefficients for lags $0, 1$ and depend only on the AR ones for lags greater than 1. In this case it behaves as the ACF of an AR(1). The same behaviour is observed for ARMA(p, q).

III. Model Selection, Estimation and Diagnostics of ARMA(p, q)

In practice, there are three spteps:

- Model selection: how to choose p and q.
- Estimation: in the most efficient way as possible.
- Diagnostic testing: to check that the result is satisfactory using estimated residuals.

If diagnostic checking not satisfactory, one should start again with new choice of p, q. One has to increase p or q or both.

Model selection. Choose p, q according to *principle of parsimony*: good fit with minimum number of parameters.

The Strategy coined by Box-Jenkins (1976) is to choose p, q based on the ACF of the residuals. Of course, one has to take into account sampling variability. When $\rho(u) = 0$ for $u > q$ like MA(q),

$$T^{\frac{1}{2}}(\hat{\rho}(u) - \rho(u)) \to_d N(0, 1 + 2\sum_{j=1}^{q} \rho^2(j)), \;\; T \to \infty.$$

The method works well for MA but not for AR, where there is no cut-off point. For AR, one has to use the *partial ACF* (PACF): for AR(p)

$$y_t = \phi_1^{(n)} y_{t-1} + ... + \phi_n^{(n)} y_{t-n} + \epsilon_t,$$

then the last coefficient $\phi_n^{(n)}$ defines the *partial ACF* at lag n. This means that if we consider say AR($p+1$), when true is AR(p), then necessarily $\phi_n^{(n)} = 0$ for $n \geq p+1$. We have a cut-off. When using estimated quantities, for estimated partial ACF $\hat{\phi}_n^{(n)}$, for $n > p$,

$$T^{\frac{1}{2}} \hat{\phi}_n^{(n)} \to_d N(0, 1), \;\; .$$

Information Criteria. In practice, data are neither AR nor MA models. One needs a criteria that allows for ARMA models. The main criteria are those based on the likelihood function (or the criterion used to estimate the parameters). They are called information criteria. There are several criteria. The most important one is the **Akaike Information Criteria** defined as

$$\text{AIC} = -\frac{2}{T}\log(\textit{Likelihood}) + \frac{2}{T} \times (\text{number of parameters}).$$

One has to minimize the AIC.

Estimation. There are several estimation methods. One could use an OLS method for AR(p) models. This is the main method to estimate AR models (use Eicker and White estimators to estimate the variance of the OLS estimator).

However, the method does not work for MA or ARMA models because $\varepsilon_{t-1},\ \varepsilon_{t-2}, ...,$ are not observable. Observe that when one ignores the MA part and want to estimate the AR coefficient by the OLS, one gets inconsistent estimators.

A method that works well is the maximum likelihood. The theory developed in the previous lecture works for time series. This is the main estimation method for ARMA types models.

One could also use the method of moments by using the covariance function. The method is called the Yule-Walker. But it is not used in practice.

Diagnostic checking. Having chosen the order of the model and estimated it, we use estimated residuals $\hat{\epsilon}_t$ to assess deviation from postulated hypothesis of white-noise and/or *i.i.d.*. One has to use the Portmanteau test: Consider the first R sample ACF of residuals. Then,

$$Q = T \sum_{u=1}^{R} \hat{\rho}_{\hat{\epsilon}}(u)^2 \rightarrow_d \chi^2(R - p - q)$$

under the assumption that the model is an ARMA(p,q). This test is known as the Box-Pierce test. R should be larger than $p + q$ and often has some interpretation. For instance, when one estimates daily models and estimates an ARMA(1,1) model, practitioners take $R = 3, 4, 5, 10, 20$, etc...

In finite sample, it is better to use the modified version called the Ljung-Box test

$$Q^* = T(T+2) \sum_{u=1}^{R} \frac{1}{T-u} \hat{\rho}^2_{\hat{\epsilon}}(u) \rightarrow_d \chi^2(R - p - q).$$

When one rejects the null, it means that there is still correlation in the residuals, which means that one should increase p, q or both.

I. Non Stationary Time Series

Macroeconomic and financial data are often non stationary. We have three interesting forms of non stationarity:

1) Trend. An example is given by

$$y_t = \mu + \delta t + \varepsilon_t.$$

2) Unit roots (also called stochastic trend). An example is given by

$$y_t = y_{t-1} + \varepsilon_t.$$

Often, macroeconomic and financial data are described by the second example. So, we will focus on this example. However, one could face another problem, which is empirically relevant:

3) Breaks. An example is given by

$$y_t = \mu_1 + \alpha_1 y_{t-1} + \varepsilon_t, \quad t = 1, 2, ..., T_1$$

and

$$y_t = \mu_2 + \alpha_2 y_{t-1} + \varepsilon_t, \quad t = T_1 + 1, T_1 + 2, ..., T.$$

We will ignore this problem in this lecture.

The simplest unit root model is given by

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

and is called a random walk. Observe that $Var[y_t] = t\sigma_\varepsilon^2 + Var[y_0]$. Likewise,

$$E[y_{t+1} \mid y_t, y_{t-1}...] = y_t$$

which is also called a martingale.

One could have a constant and then the process is called a random walk with drift

$$y_t = \mu + y_{t-1} + \varepsilon_t.$$

It is well accepted that exchange rates are random walks, without drift, but the conditional variance of ε_t is time-varying (see SAFE II). Likewise, log-prices of stocks have unit roots, but with drift because you expect that price of the stock will increase:

$$\log(p_t) = \mu + \log(p_{t-1}) + \varepsilon_t$$

which implies that the log-return is given by

$$r_t = \log(p_t) - \log(p_{t-1}) = \mu + \varepsilon_t.$$

We could have however more general examples like ARIMA models: y_t has a unit root and $\Delta y_t \equiv y_t - y_{t-1}$ is a stationary ARMA(p,q) processes.

An important problem is that we do not know whether a time series has a unit root or not. For instance, is it the case for inflation? We need a test.

The first test is known as the Dickey and Fuller test. Consider the AR(1) model

$$y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \text{ i.i.d. } \mathcal{N}(0, \sigma^2).$$

The unit root model corresponds to the case $\phi_1 = 1$. The Dickey and Fuller (DF) test deals with testing $H_0 : \phi_1 = 1$ against $H_a : \phi_1 < 1$. In other words, DF tests a unit root model against a stationary process. This alternative is the interesting one for economic and financial data (the case $\phi_1 > 1$ corresponds to explosive time series which is not interesting for our purpose).

DF changed the model in the following way: they considered the model

$$\Delta y_t = y_t - y_{t-1} = \mu + \varepsilon_t, \quad \varepsilon_t \text{ i.i.d. } \mathcal{N}(0, \sigma^2), \tag{1}$$

that they tested against the model

$$\Delta y_t = y_t - y_{t-1} = \mu + \gamma_1 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \text{ i.i.d. } \mathcal{N}(0, \sigma^2). \tag{2}$$

Hence, the null is $\gamma_1 = 0$ while the alternative is $\gamma_1 < 0$.

Given that the null corresponds to a non-stationary process while the alternative is stationary, the distribution of the test is non-standard, i.e., it is not normal or chi-square, etc... One has to simulate it.

In addition, one could consider different null hypotheses

$$\Delta y_t = \gamma_1 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \text{ i.i.d. } \mathcal{N}(0, \sigma^2) \tag{3}$$

or

$$\Delta y_t = \mu + \gamma_1 y_{t-1} + \delta t + \varepsilon_t, \quad \varepsilon_t \text{ i.i.d. } \mathcal{N}(0, \sigma^2). \tag{4}$$

When $\gamma_1 = 0$, Eq. (2) corresponds to the random walk and drift case, Eq. (3) corresponds to the random walk and no drift case, while Eq. (4) corresponds to the trend with constant case.

The DF test is simple. Do the OLS with their usual variance estimator. Consider the T test

$$T = \frac{\hat{\gamma}_1}{\sqrt{V\hat{a}r(\hat{\gamma}_1)}}.$$

When the process is stationary, T follows asymptotically a $\mathcal{N}(0, 1)$. However, this is not the case here. When has to simulate the distribution to obtain the critical values. We denote by τ_{nc}, τ_c, and τ_{ct} the critical values of the test when one consider as the null Eq. (3), Eq. (2), or Eq. (4) respectively. The critical values are reported in table 7.9 of Johnston and DiNardo.

Unfortunately, in practice, the model is often more general than Eq. (2) because there is serial correlation in ε_t. One possibility is to use a Newey-West type method (the test is known as the Phillips and Perron test). A simple approach is done by taking the model

$$\Delta y_t = \mu + \gamma_1 y_{t-1} + \sum_{i=1}^{p} \alpha_i \Delta y_{t-i} + \varepsilon_t.$$

In other words, one augments the model by adding lags of δy_t. The test is known as the augmented DF test (ADF). The asymptotic distribution of the ADF test is the same one as the one of DF. The main limitation of AD is that one assumes that p is known. This is not a practical problem.

Conclusion: For modeling, one has to start to test for non-stationarity. If there is a unit-root, one has to differentiate the data, i.e. consider Δy_t, and then find the best ARMA model for Δy_t. If there is not unit-root, one should find the best ARMA model for y_t.

II. Forecasting ARIMA Processes

Forecasting a time series y_{t+n} means that we are at time t and we want to find a variable f_t available at time t that minimizes the mean-square error

$$MSE = E[(y_{t+n} - f_t)^2].$$

One can show that the solution is

$$E[y_{t+n} \mid \text{Information at time t}].$$

Therefore, our goal is to compute $E[y_{t+n} \mid \text{Information at time t}]$. We will alse be interested in computing the forecast error given by

$$e_{t+n,t} = y_{t+n} - E[y_{t+n} \mid I_t],$$

where $I_t =$ Information at time t.

— Diebold —

true root is nevertheless very close to 1, because the Dickey-Fuller bias plagues estimation in levels. We need to use introspection and theory, in addition to formal tests, to guide the difficult decision of whether to impose unit roots, and we need to compare the forecasting performance of different models with and without unit roots imposed.

In certain respects, the most important part of unit root theory for forecasting concerns estimation, not testing. It is important for forecasters to understand the effects of unit roots on consistency and small-sample bias. Such understanding, for example, leads to the insight that at least *asymptotically* we are probably better off estimating forecasting models in levels with trends included, because then we will get an accurate approximation to the dynamics in the data regardless of the true state of the world, unit root or no unit root. If there is no unit root, then, of course, it is desirable to work in levels, and if there is a unit root, the estimated largest root will converge appropriately to unity, and at a fast rate. In contrast, differencing is appropriate only in the unit root case, and inappropriate differencing can be harmful, even asymptotically.

3. APPLICATION: MODELING AND FORECASTING THE YEN/DOLLAR EXCHANGE RATE

Here we apply and illustrate what we have learned by modeling and forecasting the yen/dollar exchange rate. For convenience, we call the yen/dollar series y, the log level $\ln y$, and the change in the log level $\Delta \ln y$. We have end-of-month data from 1973.01 through 1996.07; we plot $\ln y$ in the top panel of Figure 12.7, and $\Delta \ln y$ in the bottom panel.[10] The plot of $\ln y$ looks very highly persistent; perhaps it has a unit root. Conversely, $\Delta \ln y$ looks thoroughly stationary, and in fact rather close to white noise. Figure 12.8, which shows the correlogram for $\ln y$, and Figure 12.9, which shows the correlogram for $\Delta \ln y$, confirm the impression we gleaned from the plots. The sample autocorrelations of $\ln y$ are all very large and fail to damp, and the first sample partial autocorrelation is huge, whereas all the others are insignificantly different from 0. The correlogram of $\Delta \ln y$, however, looks very different. Both the sample autocorrelation and partial autocorrelation functions damp quickly; in fact, beyond displacement 1 they are all insignificantly different from 0. All of this suggests that $\ln y$ is $I(1)$.

10. Throughout, we work with the log of the exchange rate because the change in the log has the convenient interpretation of approximate percentage change. Thus, when we refer to the level of the exchange rate, we mean the log of the level ($\ln y$), and when we refer to the change, we mean the change of the log exchange rate ($\Delta \ln y$).

FIGURE 12.7 Log Yen/Dollar Exchange Rate (Top Panel); Change in Log Yen/Dollar Exchange Rate (Bottom Panel)

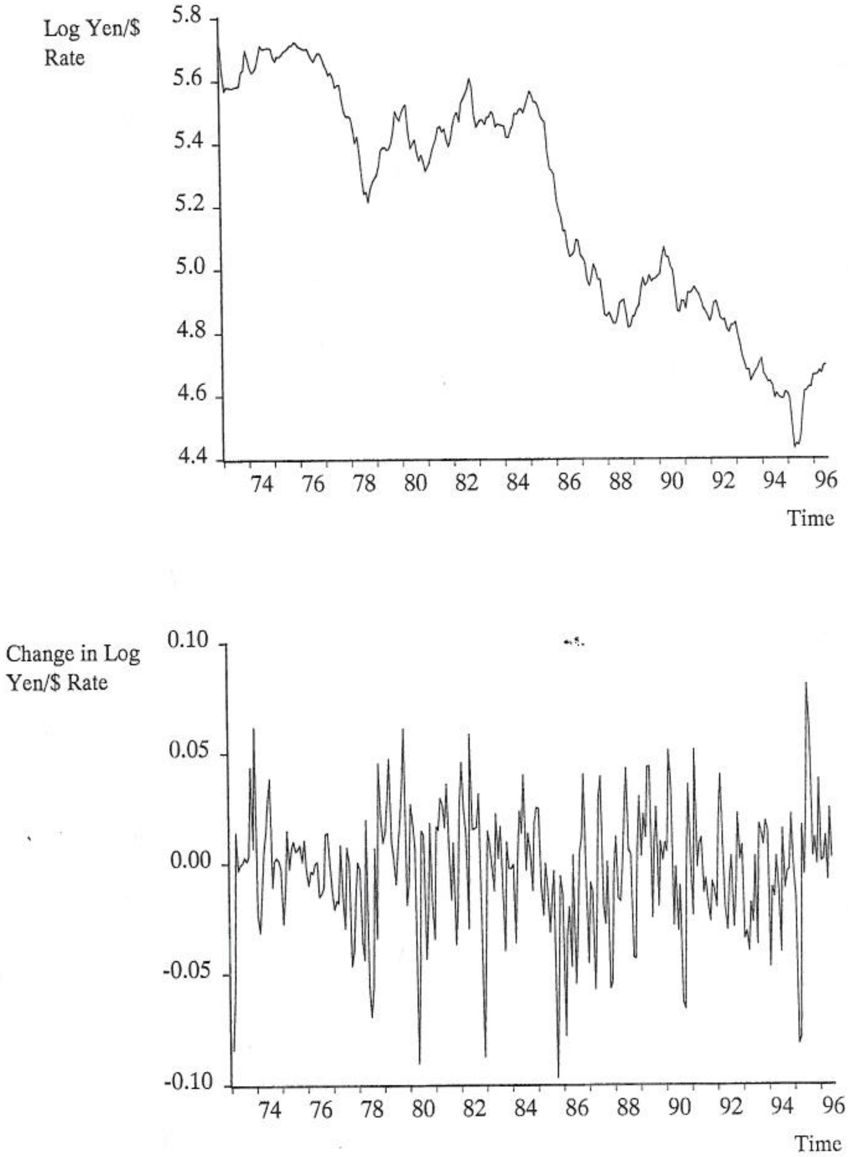


FIGURE 12.8 Log Yen/Dollar Exchange Rate: Sample Autocorrelations (Top Panel); Sample Partial Autocorrelations (Bottom Panel)

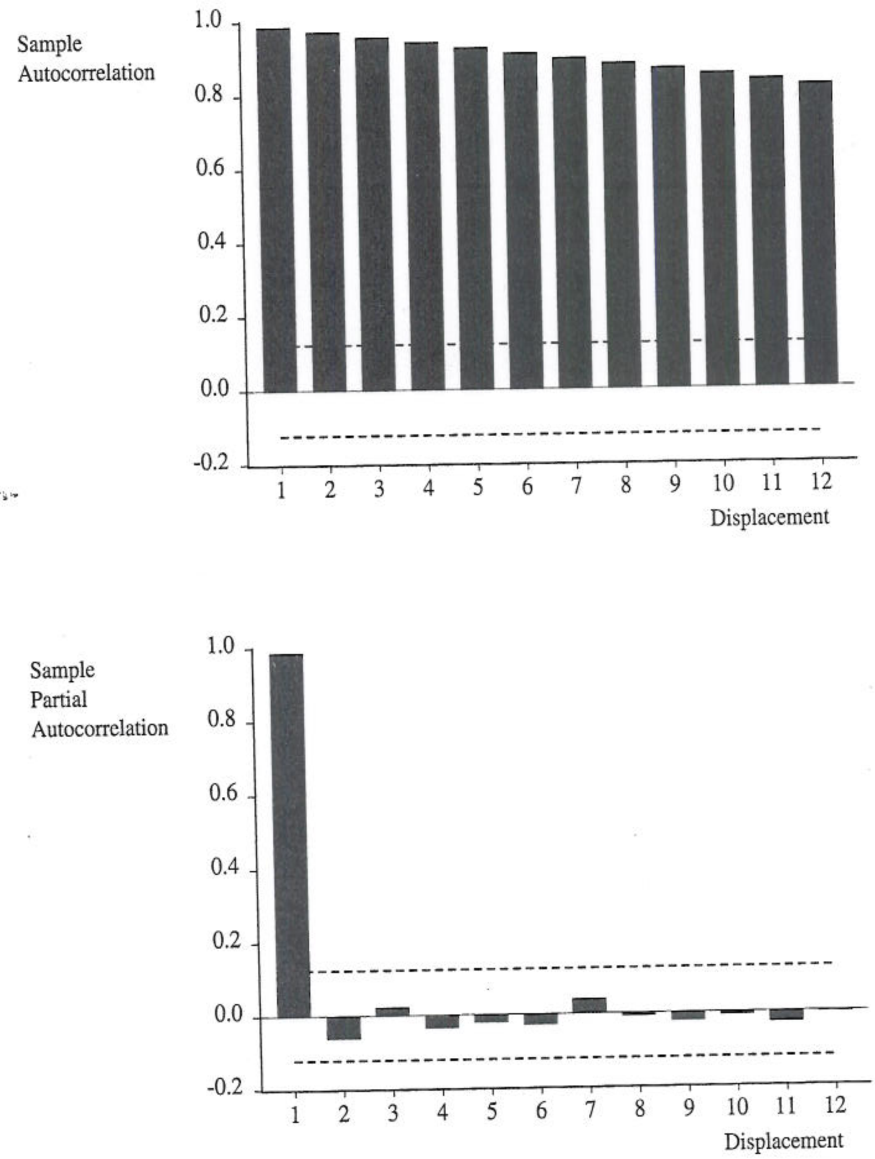


FIGURE 12.9 Log Yen/Dollar Exchange Rate, First Differences: Sample Autocorrelations (Top Panel); Sample Partial Autocorrelations (Bottom Panel)

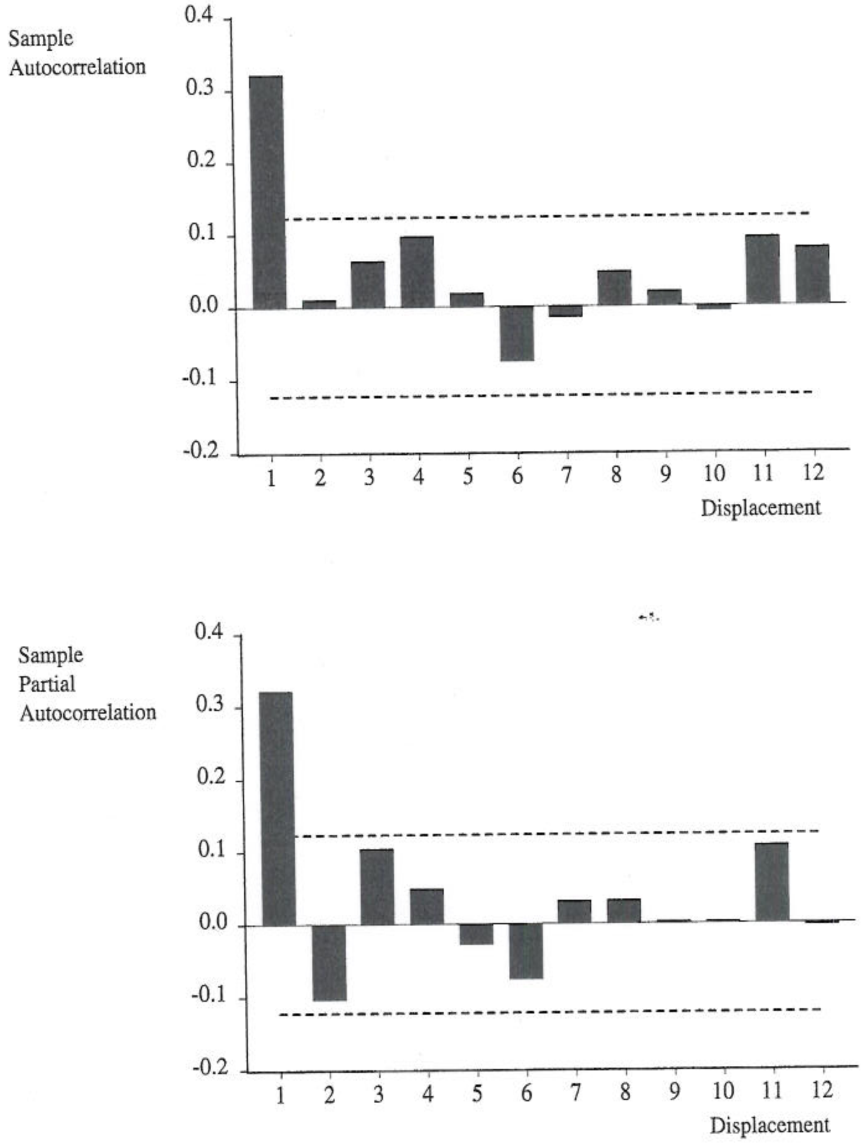


TABLE 12.1 Log Yen/Dollar Rate, Levels: AIC Values for Various ARMA Models

		MA Order			
		0	1	2	3
AR Order	0		–5.171	–5.953	–6.428
	1	–7.171	–7.300	–7.293	–7.287
	2	–7.319	–7.314	–7.320	–7.317
	3	–7.322	–7.323	–7.316	–7.308

Now we fit forecasting models. We base all analysis and modeling on ln y, 1973.01–1994.12, and we reserve 1995.01–1996.07 for out-of-sample forecasting. We begin by fitting deterministic-trend models to ln y; we regress ln y on an intercept and a time trend, allowing for up to ARMA(3,3) dynamics in the disturbances. Tables 12.1 and 12.2 show the *AIC* and *SIC* values for all the ARMA(p,q) combinations. *AIC* selects an ARMA(3,1) model, whereas *SIC* selects an AR(2). We proceed with the more parsimonious model selected by *SIC*. The estimation results appear in Table 12.3 and the residual plot in Figure 12.10; note, in particular, that the dominant inverse root is very close to 1 (0.96), whereas the second inverse root is positive but much smaller (0.35).

Out-of-sample forecasts appear in Figures 12.11 through 12.13. Figure 12.11 shows the history, 1990.01–1994.12, and point and interval forecasts, 1995.01–1996.07. Although the estimated highly persistent dynamics imply very slow reversion to trend, it happens that the end-of-sample values of ln y in 1994 are very close to the estimated trend. Thus, to a good approximation, the forecast simply extrapolates the fitted trend. Figure 12.12 shows the history together with a very long-horizon forecast (through 2020.12), to illustrate that the confidence intervals eventually flatten at ±2 standard errors. Finally, Figure 12.13 displays the history and forecast together with the realization. Most of the realization is inside the 95% confidence intervals.

TABLE 12.2 Log Yen/Dollar Rate, Levels: SIC Values for Various ARMA Models

		MA Order			
		0	1	2	3
AR Order	0		–5.130	–5.899	–6.360
	1	–7.131	–7.211	–7.225	–7.205
	2	–7.265	–7.246	–7.238	–7.221
	3	–7.253	–7.241	–7.220	–7.199

TABLE 12.3 Log Yen/Dollar Exchange Rate: Best-Fitting Deterministic-Trend Model

LS // Dependent Variable is LYEN
Sample(adjusted): 1973:03 1994:12
Included observations: 262 after adjusting endpoints
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.904705	0.136665	43.20570	0.0000
TIME	-0.004732	0.000781	-6.057722	0.0000
AR(1)	1.305829	0.057587	22.67561	0.0000
AR(2)	-0.334210	0.057656	-5.796676	0.0000

R-squared	0.994468	Mean dependent var	5.253984
Adjusted R-squared	0.994404	S.D. dependent var	0.341563
S.E. of regression	0.025551	Akaike info criterion	-7.319015
Sum squared resid	0.168435	Schwarz criterion	-7.264536
Log likelihood	591.0291	F-statistic	15461.07
Durbin-Watson stat	1.964687	Prob(F-statistic)	0.000000

Inverted AR Roots	.96	.35

FIGURE 12.10 Log Yen/Dollar Exchange Rate: Best-Fitting Deterministic-Trend Model, Residual Plot

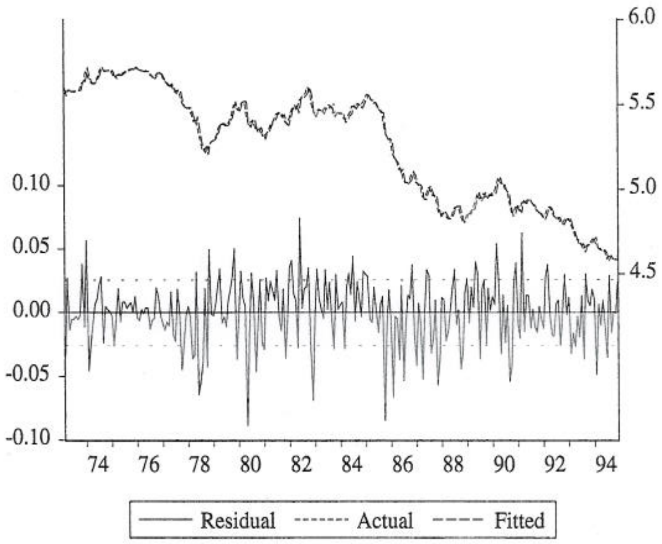


FIGURE 12.11 Log Yen/Dollar Rate: History and Forecast, AR(2) in Levels with Linear Trend

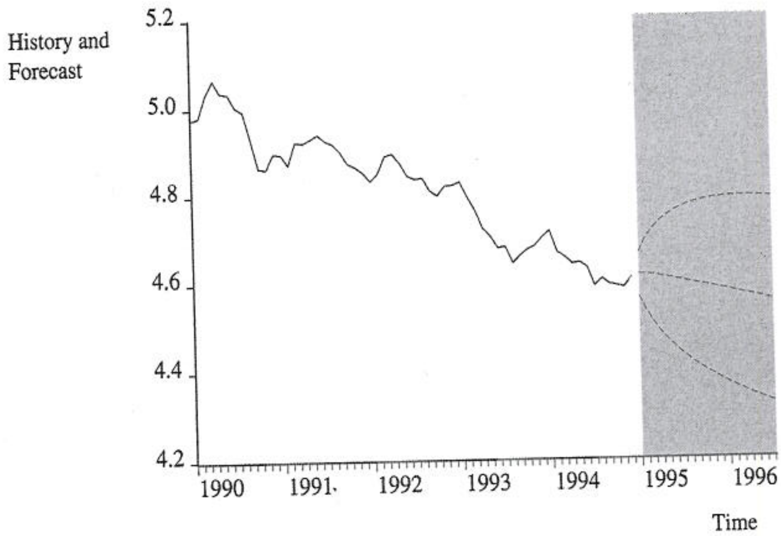


FIGURE 12.12 Log Yen/Dollar Rate: History and Long-Horizon Forecast, AR(2) in Levels with Linear Trend

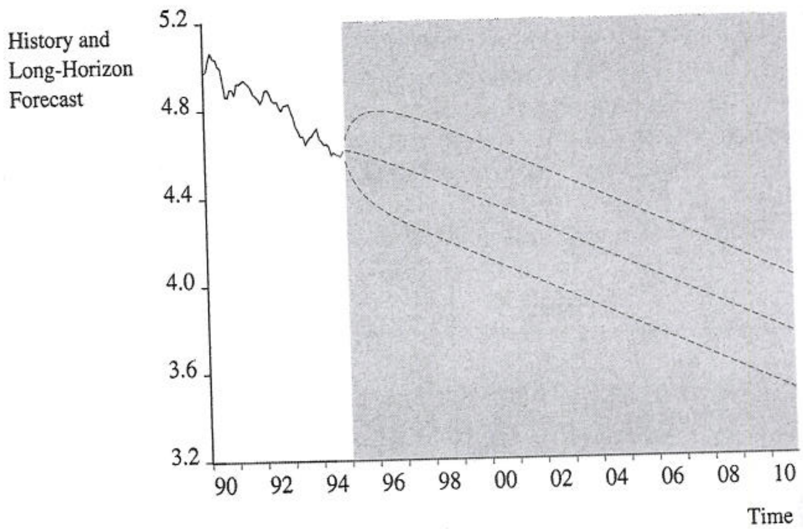
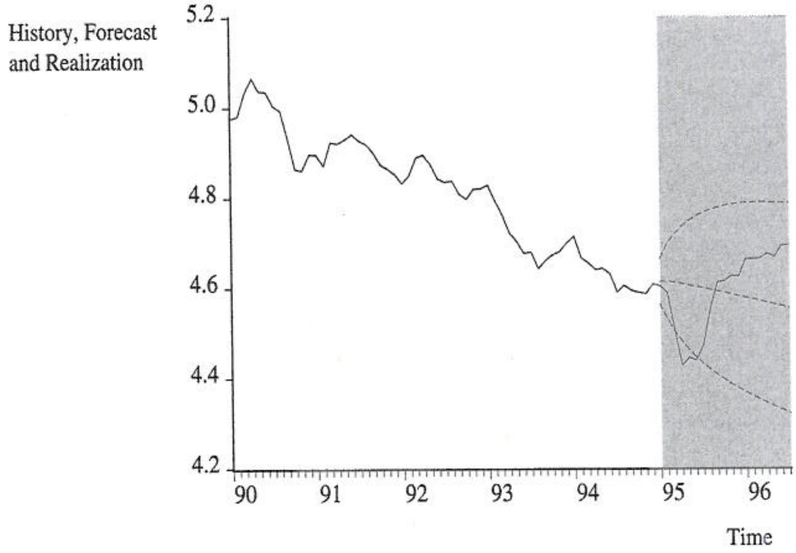


FIGURE 12.13 Log Yen/Dollar Rate: History, Forecast, and Realization, AR(2) in Levels with Linear Trend



In light of the suggestive nature of the correlograms, we now perform a formal unit root test, with trend allowed under the alternative hypothesis. Table 12.4 shows the results with three augmentation lags.[11] There is no evidence whatsoever against the unit root; thus, we consider modeling $\Delta \ln y$. We regress $\Delta \ln y$ on an intercept and allow for up to ARMA(3,3) dynamics in the disturbance. The *AIC* values appear in Table 12.5, and the *SIC* values in Table 12.6. *AIC* selects an ARMA(3,2), and *SIC* selects an AR(1). Note that the models for $\ln y$ and $\Delta \ln y$ selected by *SIC* are consistent with each other under the unit root hypothesis—an AR(2) with a unit root in levels is equivalent to an AR(1) in differences—in contrast to the models selected by *AIC*. For this reason and, of course, for the usual parsimony considerations, we proceed with the AR(1) selected by *SIC*. We show the regression results in Table 12.7 and Figure 12.14; note the small but nevertheless significant coefficient of 0.32.[12]

11. We considered a variety of augmentation lag orders, and the results were robust—the unit root hypothesis cannot be rejected. For the record, *SIC* selected one augmentation lag, whereas *AIC* and *t*-testing selected three augmentation lags.

12. The ARMA(3,2) selected by *AIC* is, in fact, very close to an AR(1) because the two estimated MA roots nearly cancel with two of the estimated AR roots, which would leave an AR(1).

TABLE 12.4 Log Yen/Dollar Exchange Rate: Augmented Dickey-Fuller Unit Root Test

Augmented Dickey-Fuller Test Statistic	-2.498863	1% Critical Value	-3.9966
		5% Critical Value	-3.4284
		10% Critical Value	-3.1373

Augmented Dickey-Fuller Test Equation
LS // Dependent Variable is D(LYEN)
Sample(adjusted): 1973:05 1994:12
Included observations: 260 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LYEN(-1)	-0.029423	0.011775	-2.498863	0.0131
D(LYEN(-1))	0.362319	0.061785	5.864226	0.0000
D(LYEN(-2))	-0.114269	0.064897	-1.760781	0.0795
D(LYEN(-3))	0.118386	0.061020	1.940116	0.0535
C	0.170875	0.068474	2.495486	0.0132
@TREND(1973:01)	-0.000139	5.27E-05	-2.639758	0.0088

R-squared	0.142362	Mean dependent var	-0.003749
Adjusted R-squared	0.125479	S.D. dependent var	0.027103
S.E. of regression	0.025345	Akaike info criterion	-7.327517
Sum squared resid	0.163166	Schwarz criterion	-7.245348
Log likelihood	589.6532	F-statistic	8.432417
Durbin-Watson stat	2.010829	Prob(F-statistic)	0.000000

TABLE 12.5 Log Yen/Dollar Rate, Changes: AIC Values for Various ARMA Models

		MA Order			
		0	1	2	3
AR Order	0		−7.298	−7.290	−7.283
	1	−7.308	−7.307	−7.307	−7.302
	2	−7.312	−7.314	−7.307	−7.299
	3	−7.316	−7.309	−7.340	−7.336

TABLE 12.6 Log Yen/Dollar Rate, Changes: SIC Values for Various ARMA Models

		MA Order			
		0	1	2	3
AR Order	0		−7.270	−7.249	−7.228
	1	−7.281	−7.266	−7.252	−7.234
	2	−7.271	−7.259	−7.238	−7.217
	3	−7.261	−7.241	−7.258	−7.240

Out-of-sample forecasting results appear in Figures 12.15 through 12.17. Figure 12.15 shows the history and forecast. The forecast looks very similar—in fact, almost identical—to the forecast from the deterministic-trend model examined earlier. That is because the stochastic-trend and deterministic-trend models are, in fact, extremely close to one another in this case; even when we don't impose a unit root, we get an estimated dominant root that is very close to unity. Figure 12.16 shows the history and a very long horizon forecast. The long horizon forecast reveals one minor and one major difference between the forecasts from the deterministic-trend and stochastic-trend models. The minor difference is that by the time we are out to 2010, the point forecast from the deterministic-trend model is a little lower, reflecting the fact that the estimated trend slope is a bit more negative for the deterministic-trend model

TABLE 12.7 Log Yen/Dollar Exchange Rate: Best-Fitting Stochastic-Trend Model

LS // Dependent Variable is DLYEN
Sample(adjusted): 1973:03 1994:12
Included observations: 262 after adjusting endpoints
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.003697	0.002350	-1.573440	0.1168
AR(1)	0.321870	0.057767	5.571863	0.0000

R-squared	0.106669	Mean dependent var	-0.003888
Adjusted R-squared	0.103233	S.D. dependent var	0.027227
S.E. of regression	0.025784	Akaike info criterion	-7.308418
Sum squared resid	0.172848	Schwarz criterion	-7.281179
Log likelihood	587.6409	F-statistic	31.04566
Durbin-Watson stat	1.948933	Prob(F-statistic)	0.000000

Inverted AR Roots .32

FIGURE 12.14 Log Yen/Dollar Exchange Rate: Best-Fitting Stochastic-Trend Model, Residual Plot

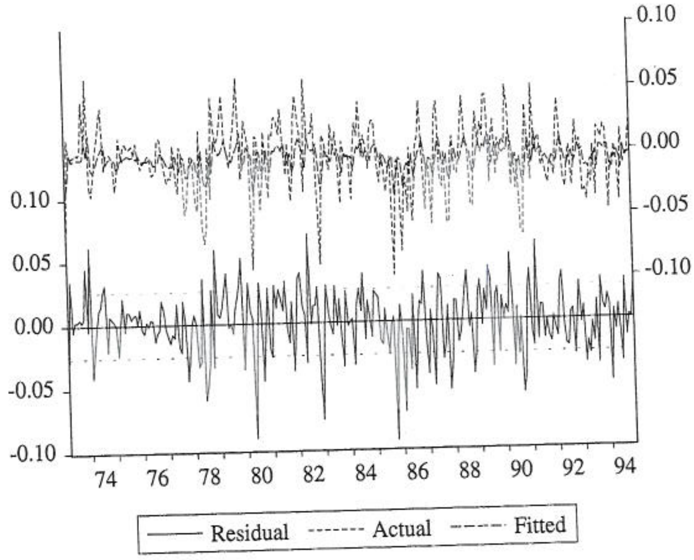


FIGURE 12.15 Log Yen/Dollar Rate: History and Forecast, AR(1) in Differences with Intercept

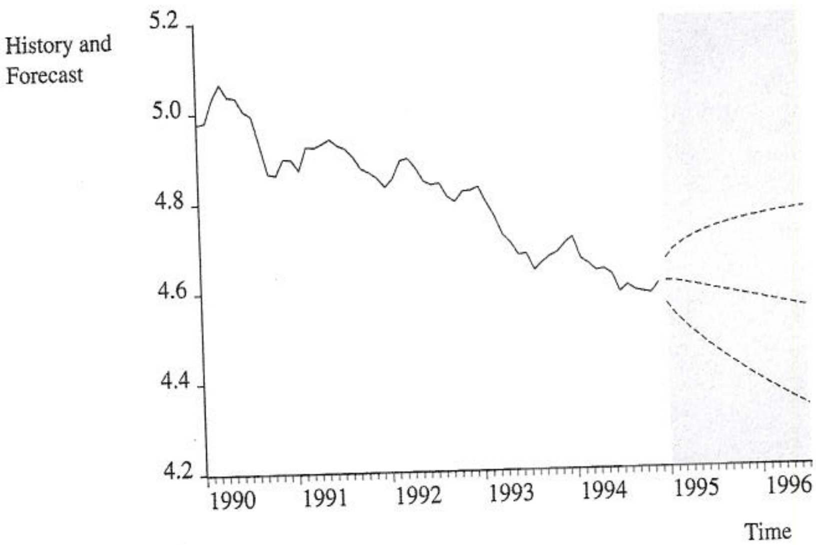
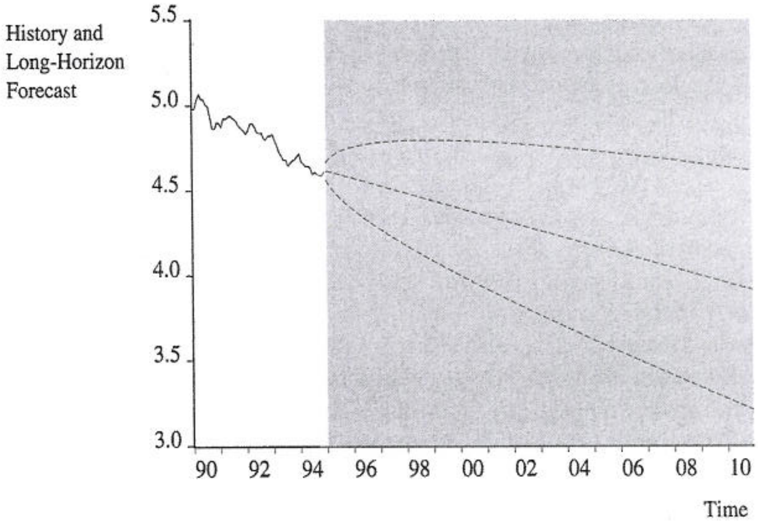


FIGURE 12.16 Log Yen/Dollar Rate: History and Long Horizon Forecast, AR(1) in Differences with Intercept



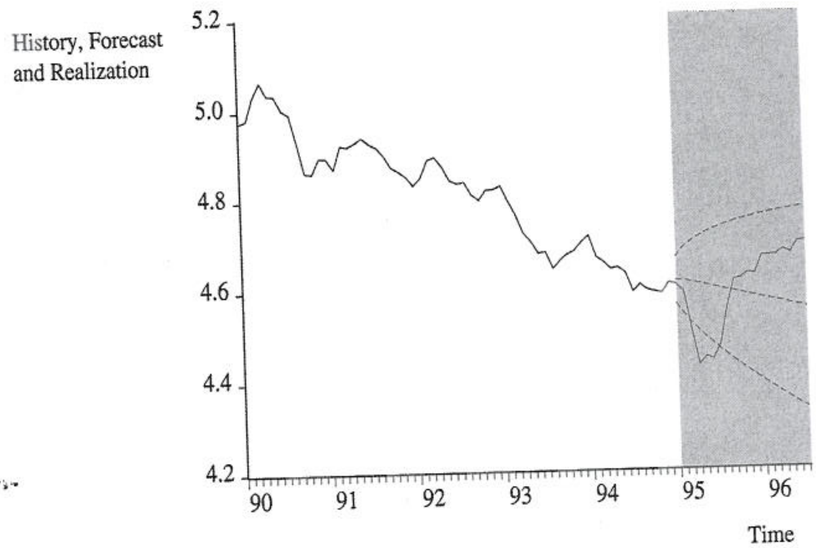
than for the stochastic-trend model. Statistically speaking, however, the point forecasts are indistinguishable. The major difference concerns the interval forecasts: The interval forecasts from the stochastic-trend model widen continuously as the horizon lengthens, whereas the interval forecasts from the deterministic-trend model do not. Finally, Figure 12.17 shows the history and forecast together with the realization 1995.01–1996.07.

Comparing the AR(2) with trend in levels (the levels model selected by *SIC*) and the AR(1) in differences (the differences model selected by *SIC*), it appears that the differences model is favored in that it has a lower *SIC* value. The AR(1) in differences fits only slightly worse than the AR(2) in levels—recall that the AR(2) in levels had a near unit root—and saves one degree of freedom.[13] Moreover, economic and financial considerations suggest that exchange rates should be close to random walks because if the change were predictable, one could make a lot of money with very little effort, and the very act of doing so would eliminate the opportunity.[14]

13. *Caution:* In a sense, the AR(1) model in differences may not save the degree of freedom, insofar as the decision to impose a unit root was itself based on an earlier estimation (the augmented Dickey-Fuller test), which is not acknowledged when computing *SIC* for the AR(1) in differences.

14. As for the trend (drift), it may help as a local approximation, but be wary of too long an extrapolation. See the Problems and Complements at the end of this chapter.

FIGURE 12.17 Log Yen/Dollar Rate: History, Forecast, and Realization, AR(1) in Differences with Intercept



Ironically enough, in spite of the arguments in favor of the stochastic-trend model for ln y, the deterministic-trend model does slightly better in out-of-sample forecasting on this particular dataset. The mean-squared forecast error from the deterministic-trend model is 0.0107, whereas that from the stochastic-trend model is 0.0109. The difference, however, is likely statistically insignificant.

4. SMOOTHING

We encountered the idea of time series smoothing early on, when we introduced simple moving-average smoothers as ways of estimating trend.[15] Now we introduce additional smoothing techniques and show how they can be used to produce forecasts.

Smoothing techniques, as traditionally implemented, have a different flavor than the modern model-based methods that we have used in this book. Smoothing techniques, for example, don't require "best-fitting models," and they don't generally produce "optimal forecasts." Rather, they are simply a way to tell a computer to draw a smooth line through data, just as we would do

15. See the Problems and Complements of chapter 4.