Time Series Analysis

Enders, Chapter 2: Stationary Time Series Models

Jude C. Hays

February 2, 2021

Outline

Lag Operators

- Oftentimes it is much more convenient to write difference equations using the lag operator L, which is defined as a linear operator such that

$$L^i y_t \equiv y_{t-i}$$

- Lag operators allow us to write compactly the difference equation $y_t = a_0 + a_1 y_{t-1} + ... + a_p y_{t-p} + \varepsilon_t$ as

$$(1 - a_1 L - a_2 L^2 - ... - a_p L^p) y_t = a_0 + \varepsilon_t$$

or simply

$$A(L) y_t = a_0 + \varepsilon_t$$

Lag Operators

- And, importantly for our purposes today, we can express the equation
  $y_t = a_0 + a_1 y_{t-1} + \ldots + a_p y_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \ldots + \beta_q \varepsilon_{t-q}$ as

$$A(L)y_t = a_0 + B(L)\varepsilon_t$$

- This representation has the compact particular solution

$$y_t = a_0/A(L) + B(L)\varepsilon_t/A(L)$$

- If we do not need to know the values of the coefficients in the particular solution, you will likely see the lag operator notation used to write out time series models.

White Noise Process

The autoregressive moving-average (ARMA) model underlies much of time-series analysis. The methods for estimating these models were developed in Box and Jenkins (1976).

- We begin with white noise processes, which are a critical component of ARMA models.
- We use the notation $\{\varepsilon_t\}$ to represent the entire sequence $\{\varepsilon_0, \varepsilon_1, \varepsilon_2, ..., \varepsilon_t\}$
- The sequence $\{\varepsilon_t\}$ is a white noise process if

$$E(\varepsilon_t) = E(\varepsilon_{t-1}) = ... = 0$$

$$E(\varepsilon_t^2) = E(\varepsilon_{t-1}^2) = ... = \sigma^2$$

$$E(\varepsilon_t, \varepsilon_{t-s}) = E(\varepsilon_{t-j}, \varepsilon_{t-j-s}) = ...0 \text{ for all } j \text{ and } s$$

ARMA (p,q) Models

- Next, consider a p^{th} order difference equation

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

- Let x_t take the following form

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

- We call this a q^{th} order moving average process.
- Note that while $\{\varepsilon_t\}$ is a white noise process, $\{x_t\}$ is not.

Covariance-stationary Processes

If y_t is a linear stochastic difference equation, the stability condition is a necessary condition for the time-series $\{y_t\}$ to be stationary.

- A stochastic process with finite mean and variance is **covariance-stationary** if for all t, s, and j

$$E(y_t) = E(\varepsilon_{t-s}) = \mu$$

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

$$E[(y_t - \mu), (y_{t-s} - \mu)] = E[(y_{t-j} - \mu), (y_{t-j-s} - \mu)] = \gamma_s$$

- In words, this implies that the mean and autocovariances of the time series do not depend on time.

Covariance-stationary Processes

In order for a time series to be stationary...

- The homogeneous solution must be zero.
- The characteristic roots must lie within the unit circle.

As for the stochastic part of the particular solution, which will take the form $x_t = \sum_{i=0}^{\infty} \beta_i \varepsilon_{t-i}$

- The mean of $\{x_t\}$ must be finite and time-independent, which it is given that $\{\varepsilon_t\}$ is a white noise process.
- The variance of $\{x_t\}$ must be finite and time-independent, which it as long as $\sum (\beta_i)^2$ is finite.
- The covariances of $\{x_t\}$ must be finite and time-independent, which they are as long as $\sigma^2(\beta_s + \beta_1\beta_{s+1} + \beta_2\beta_{s+2} + ...)$ is finite.

Identification, Estimation, Diagnostics

1. Compare sample autocorrelation and partial autocorrelation functions with theoretical ARMA processes.
2. Choose a parsimonious specification with coefficient estimates that imply a covariance-stationary process.
3. Check model fit using AIC/SBC.
4. Check the residual to make sure they are "white noise."
5. Check out of sample forecasts and coefficient stability.

The ACF and PACF

- The *autocorrelation* between y_t and $y_t - s$ is defined as $\rho_s \equiv \frac{\gamma_s}{\gamma_0}$.
- The *partial autocorrelation* between y_t and $y_t - s$ (ϕ_{ss}) eliminates the effects of the intervening values of $y_t - 1$ and $y_t - s + 1$.

The ACF and PACF

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

$$\begin{aligned}
&\gamma_s = E[(y_t - \mu)(y_{t-s} - \mu)] \\
&= E[(\varepsilon_t + a_1\varepsilon_{t-1} + (a_1)^2\varepsilon_{t-2} + \ldots)(\varepsilon_{t-s} + a_1\varepsilon_{t-s-1} + \ldots)] \\
&= \sigma^2 (a_1)^s [1 + (a_1)^2 + (a_1)^4 + \ldots] \\
&= \sigma^2 (a_1)^s / [1 - (a_1)^2]
\end{aligned}$$

$$y_t^* = \sum_{j=1}^{s-1} \phi_{sj} y_{t-j}^* + \phi_{ss} y_{t-s}^* + e_t$$

Theoretical ACF and PACF Patterns

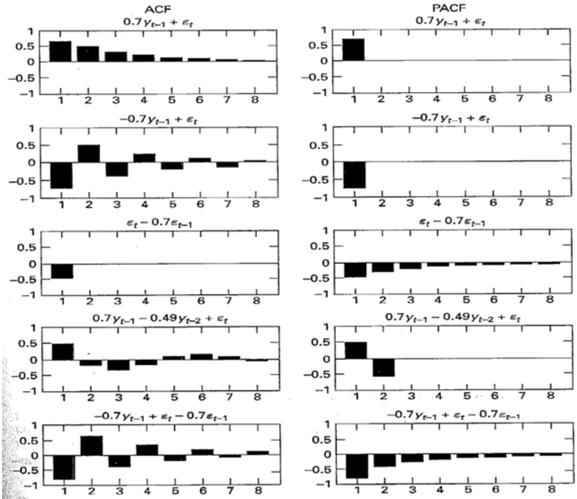


FIGURE 2.2 Theoretical ACF and PACF Patterns

Theoretical ACF and PACF Properties

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 geometric 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 lag 1. Sign $\rho_1 = \text{sign}(a_1 + \beta)$	Oscillating decay after lag 1. $\phi_{11} = \rho_1$
ARMA(1, 1) $a_1 < 0$	Oscillating decay beginning after lag 1. Sign $\rho_1 = \text{sign}(a_1 + \beta)$	Geometric decay beginning after lag 1. $\phi_{11} = \rho_1$ and $\text{sign}(\phi_{ss}) = \text{sign}(\phi_{11})$.
ARMA(p, q)	Decay (either direct or oscillatory) beginning after lag q.	Decay (either direct or oscillatory) beginning after lag p.

The AIC and SBC

- The Akaike Information Criterion (AIC) and Schwartz Bayesian Criterion (SBC) are measures of model fit that can be used to compare non-nested models (e.g., AR(1) and MA(3)).

$$AIC = T\ln(\text{sum of squared residuals}) + 2n$$

$$SBC = T\ln(\text{sum of squared residuals}) + n\ln(T)$$

- Smaller numbers are better! As the model fit improves the AIC and SBC $\rightarrow -\infty$.

The AIC and SBC

- The SBC has superior large-sample properties. Both the AIC and SBC will select higher order models than the true data-generating process (DGP), but the SBC is asymptotically consistent, while the AIC is biased in favor of over-parameterized models.
- However, the AIC can perform better than SBC in small samples.
- Hopefully, both measures select the same model specification.
- If not, check the residuals, out-of-sample forecasting performance, and parameter stability. (You should do this regardless.)

Checking Residuals

- If the residuals are normal, independent and identically distributed. Only 5% of the standardized residuals (ε_t/σ) should lie outside of the -2 to +2 band.
- Check the ACF and PACF for the residuals.
- Calculate the Ljung-Box statistic.

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

- This statistic tests the null hypothesis that the residuals were generated by a white-noise process. Under the null hypothesis, it is distributed χ^2 with $s-p-q-1$ degrees of freedom.

Out-of-Sample Forecasting Performance

- To assess the forecasting performance of a model, we evaluate its out-of-sample forecast errors.

$$e_T(1) = y_{T+1} - E_T(y_{T+1})$$

- If our model is an ARMA(2,1), the forecast error is

$$e_T(1) = y_{T+1} - (\hat{a}_0 + \hat{a}_1 y_T + \hat{a}_2 y_{T-1} + \hat{\beta}_1 \hat{\varepsilon}_T)$$

- We can compare the forecasting performance in terms of both bias and efficiency.

Out-of-Sample Forecasting Performance

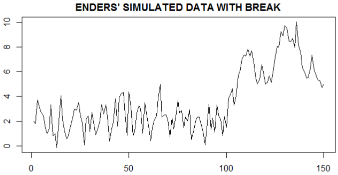
- A useful comparison is the mean squared prediction error, which combines forecasting bias and efficiency performance.

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

  where H is the number of one-step-ahead forecasts generated with model i.

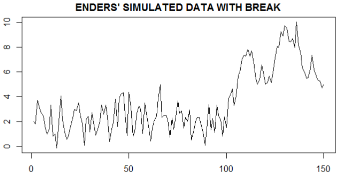
- The Granger-Newbold and Diebold-Mariano tests statistically evaluate the null hypothesis of equal forecast accuracy. The former assumes a quadratic loss function (MSPE), while the latter allows the loss function to be general.

Parameter Stability



- Structural breaks in the DGP can lead to wildly inaccurate parameter estimates.
- The true DGP for the figure above is $y_t = 1 + 0.5y_{t-1} + \varepsilon_t,$ for $t < 100$ and $y_t = 2.5 + 0.65y_{t-1} + \varepsilon_t, \text{for} t \geq 100.$

Parameter Stability



- Estimating an AR(1) model using the entire sample gives

$$y_t = 0.44 + 0.88 y_{t-1}$$

- Structural break in constant is mistaken for persistence.
- We typically look for structural breaks by estimating our models recursively and evaluating the evolution of the parameter estimates and forecasting accuracy of the model over time.

The Autoregressive Distributed Lag Model

We can think of time series regression as an alternative to the ARIMA approach.

- There are a number of models that we lump together under the category of time series regressions.
- The most general of these models is the autoregressive distributed lag (ADL) model

$$y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{j=1}^{n} \sum_{i=0}^{q} \beta_{ji} x_{jt-i} + \varepsilon_t$$

- Consider the simple case where $p = q = n = 1$

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$$

Restricted Versions of the ADL Model

- The other common time series regression models are restricted version of the ADL.
- Among these models are

Partial Adjustment Model: $y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \varepsilon_t$

Finite Distributed Lag Model: $y_t = \alpha_0 + \beta_0 x_t + \beta_1 y_{t-1} + \varepsilon_t$

Static Model: $y_t = \alpha_0 + \beta_0 x_t + \varepsilon_t$

Common Factor Model:

$$y_t = \beta_0 x_t + u_t, u_t = \alpha_0 + \alpha_1 u_{t-1} + \varepsilon_t, \alpha_1 = -\beta_1/\beta_0$$

Doing Time Series Regression the Right Way

- Estimate the general model (i.e., the ADL or ECM) and test restrictions.
- Calculate all the dynamic quantities of interest (short-run effects, long-run effects, mean-lag length, median-lag length).
- Note that all of these quantities are contained in the *impulse response function*.