Data Analysis:
Time Series Models

Mark Hendricks

August Review

UChicago Financial Mathematics

Outline

Autoregression

Stationarity

Volatility Models

AR(1)

An AR(1) refers to a model where the dependent variable depends on a single lag:

$$
\begin{aligned}
y_t &= \alpha + \rho y_{t-1} + \epsilon_t \\
\text{corr}\left(y_{t-1}, \epsilon_t\right) &= 0
\end{aligned}
$$

This follows the form of our usual linear model,

$$y = \beta_0 + \beta_1 x + \epsilon$$

The AR(1) restricts the regressor to be a time-lag of the dependent variable.

Stationarity

A time-series is said to be stationary if...

- $\mathbb{E}[y_t]$ is independent of t.
- $\text{var}(y_t)$ is a finite, positive constant, independent of t.
- $\text{cov}(y_t, y_{t-k})$ is a finite function of k—not a function of t.

An AR(1) process is stationary if $|\rho| < 1$.

Properties of AR(1)

The AR(1) model implies...

- Correlations follow

$$\operatorname{corr}(y_t, y_{t-k}) = \rho^k$$

- Conditional mean is

$$\mathbb{E}_t[y_{t+1}] = \alpha + \rho y_t$$

  where $\mathbb{E}_t$ denotes an expectation conditional on time-t data. (In this case, the series $\{y_t\}$.)

- Unconditional mean and variance are

$$\mathbb{E}[y] = \frac{\alpha}{1-\rho}, \qquad \operatorname{var}(y) = \frac{\sigma_\epsilon^2}{1-\rho}$$

Estimating an AR model

While OLS is commonly used, it can be problematic.

- The exogeneity assumption does not hold. The regressor, y_{t-1}, is correlated with past errors, ϵ_{t-1}.
- OLS will be biased.
- OLS standard errors will be wrong.

Still, OLS will often be consistent and we can use robust standard errors to get correct inference.

AR(1) with AR(1) residuals

If the AR(1) has autocorrelated errors, the problems with OLS become more serious.

$$
\begin{aligned}
y_t =& \alpha + \beta y_{t-1} + \epsilon_t \\
\epsilon_t =& \gamma \epsilon_{t-1} + u_t
\end{aligned}
$$

Now, OLS estimation of β is inconsistent.

- The problem stems from correlation between y_{t-1} and ϵ_{t-1} causing correlation between y_{t-1} and ϵ_t.
- Even with lots of data, OLS can't distinguish the impact on y_t coming from ϵ_t versus that coming from y_{t-1}.

Monte Carlo Simulation

Generate data with

$$
\begin{aligned}
y_t &= 0 + .9 y_{t-1} + \epsilon_t \\
\epsilon_t &= \gamma \epsilon_{t-1} + u_t
\end{aligned}
$$

- Try to estimate the parameters using OLS and the observed series $\{y\}$.
- Run the simulation first with white noise, $\gamma = 0$.
- Then run the simulation with autocorrelated errors.

Simulated OLS performance - white noise

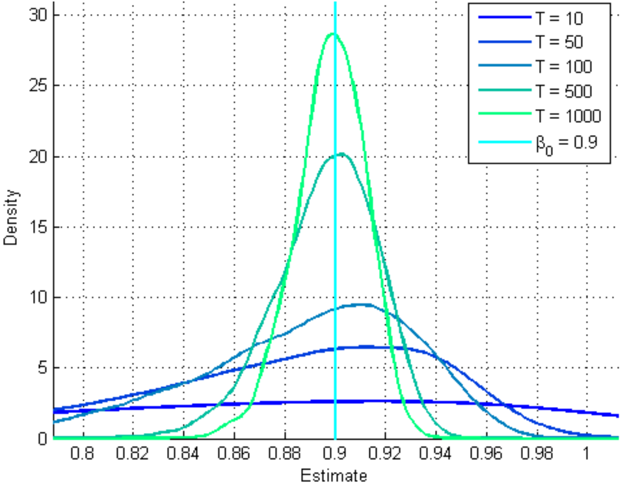


Figure: Simulated distribution of β estimates.
Source: Matlab.

Simulated OLS performance - AR errors

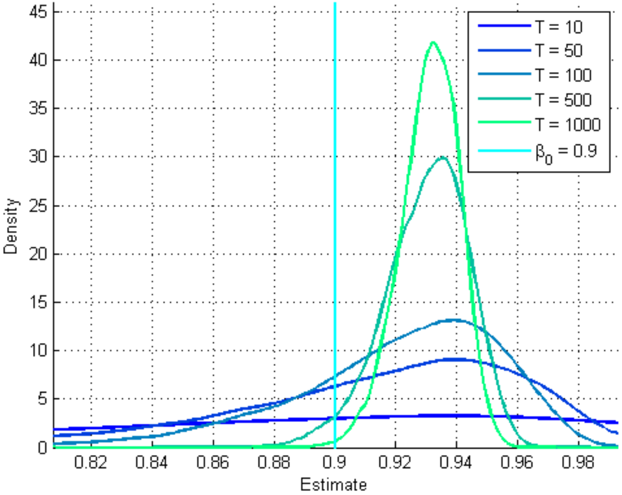


Figure: Simulated distribution of β estimates.
Source: Matlab.

Outline

Autoregression

Stationarity

Volatility Models

Random Walk

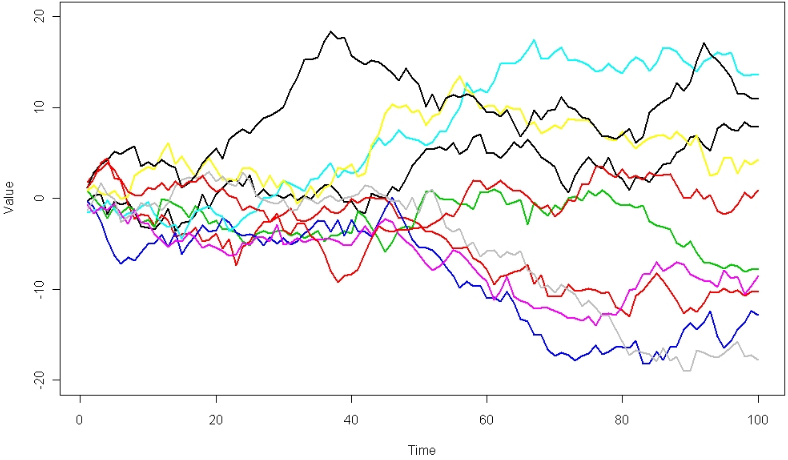
Consider the following random walk process:

$$
\begin{aligned}
y_t =& y_{t-1} + \epsilon_t \\
0 =& \text{corr}(\epsilon_t, y_{t-1}), \quad \epsilon_t \sim iid., \quad \mathbb{E}[\epsilon_t] = 0
\end{aligned}
$$

Note that the random walk is non-stationary.

- Conditional forecast is $\mathbb{E}_t[y_{t+1}] = y_t$.
- Unconditional forecast is $\mathbb{E}[y_t] = 0$.
- Variance is undefined!

Simulated random walks



Trends in data

Often data has a trend. Consider two models of this:

$$
\begin{aligned}
y_t =& \mu + y_{t-1} + \epsilon_t \\
y_t =& \mu + \beta t + \epsilon_t
\end{aligned}
$$

- The first is a random walk with drift.
- The second is a trend-stationary process.
- In the form above, neither is stationary.

Spurious correlation

Regressions using data with trends can give to spurious results.

- Two series with trends may show strongly significant regression results just due to the underlying trend in both.
- For instance, two completely independent random walks with drift may seem highly correlated.
- Even worse, random walks without trends will lead to spurious regression.

De-trending or differencing?

Unfortunately, the remedy depends on whether the data is trend-stationary or a random walk.

- If we take the difference of a random walk, we have

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

  which suits our OLS assumptions.

- But if we take the difference of a trend-stationary series, we get

$$y_t - y_{t-1} = \beta + \epsilon_t - \epsilon_{t-1}$$

  which involves past error terms and will violate OLS assumptions.

Unit-root tests

To distinguish whether differencing or de-trending is needed, consider a nested model:

$$y_t = \mu + \beta t + y_{t-1} + \epsilon_t$$

Differencing (and introducing the artificial parameter γ,) gives

$$y_t - y_{t-1} = \alpha_0 + \alpha_1 t + (\gamma - 1) y_{t-1} + \epsilon_t$$

- If estimation reveals $\gamma = 1$, we have a random walk with drift.
- If estimation reveals γ significantly less than 1, we may prefer to detrend.
- We cannot use OLS inference to understand γ. Use tests such as Dickey-Fuller.

Seasonality

Many time-series have seasonal variation.

- To adjust, one might include seasonal dummy variables.
- Or start with seasonally adjusted data, and interpret the regression results accordingly.

Outline

Autoregression

Stationarity

Volatility Models

ARCH

An **ARCH** model is a parametric way to deal with heteroscedasticity.

- An ARCH(1) model is

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + \epsilon_t$$
$$\epsilon_t = u_t\sqrt{\alpha_0 + \alpha_1\epsilon_{t-1}^2}$$

where $u_t \sim \mathcal{N}(0, 1)$.

- This is often written as

$$\sigma_t^2 = \alpha_0 + \alpha_1\epsilon_{t-1}^2$$

where σ_t^2 now refers to variance conditional on lagged ϵ.

Using ARCH

ARCH gives a flexible, parametric description of heteroscedasticity.

- ARCH(q) generalizes by making the conditional variance a function of an MA(q) instead of MA(1).

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \ldots + \alpha_q \epsilon_{t-q}^2$$

- The ARCH model can be estimated with OLS, but it is not efficient relative to *nonlinear* estimators such as maximum likelihood.

GARCH

The **GARCH** model generalizes the ARCH model by including lagged values of variance in the equation describing how variance evolves:

$$\sigma_t^2 = \alpha_0 + \gamma_1 \sigma_{t-1}^2 + \alpha_1 \epsilon_{t-1}^2$$

- This is a GARCH(1,1) model as it includes 1 lag of σ^2 and 1 lag of ϵ.
- GARCH(p, q) include p lags of σ^2 and q lags of ϵ.
- Working with GARCH(p, q) becomes extremely complicated as p and q get large, so GARCH(1,1) is the most common.
- Even GARCH(1,1) models are typically estimated with nonlinear methods.

Testing GARCH

Testing whether data follows a GARCH process is done with the LM test described above.

- Regress sum of square errors on lagged values of the sample errors.
- Take the R-squared from this regression of squared errors.
- This statistic has a χ^2 distribution.

References

- Cochrane, John. *Asset Pricing.* 2001.
- Greene, William. *Econometric Analysis.* 2011.
- Hamilton, James. *Time Series Analysis.* 1994.
- Tsay, Ruey. *Analysis of Financial Time Series.* 2010.