Time Series Lecture II

EconTutor.com

Summarizing AR(1) Econometrics Tutor

AR(1) is only stationary if $|\rho| < 1$

- Things go haywire if this isn't true

AR(1) is more complicated than the MA(1):

- Autocovariances non-zero forever, although die out eventually if $|\rho| < 1$
- AR(1) thus weakly dependent if $|\rho| < 1$

Why use AR(1)?

- Most commonly modeled time series process
- Often very realistic process
  - Processes tend to persist over long stretches, unlike MA(1) where processes are not correlated for one period then suddenly uncorrelated

Estimating AR(1) Model

- The AR(1) model is estimated with OLS
- Assumptions *for consistency* (no need for unbiasedness):
  1. Contemporaneous Exogeneity: $E(u_t|y_t) = 0$
     - Same as the original OLS assumption
     - We CANNOT have strict exogeneity (why?)
  2. Stationarity and Weak Dependence
     - This is a substantially weaker assumption than iid
  3. No perfect multicollinearity
- Additional assumptions *for inference*
  1. Homoscedasticity: $Var(u|Y) = \sigma_Y^2$
  2. No serial correlation in u: $E(u_s u_t|y_t, y_s) = 0$
- With these additional assumptions, for large T, all usual hypothesis testing procedures are valid

Example AR(1): GDP Growth Rates

How good of a predictor is lagged GDP growth?

$$\Delta \log GDP_t = \beta_0 + \beta_1 \Delta \log GDP_{t-1} + U_t$$

Results:

$$\Delta \widehat{\log GDP}_t = \underset{(.001)}{.005} + \underset{(.056)}{.372} \times \Delta \log GDP_{t-1}$$

- This is *not* a causal relationship
- What "causes" GDP growth (positive or negative) to be correlated over time?
  - Roll out of new inventions
  - Upward trend in education
  - Bank failures might take time to cascade through a network
- Putting lagged GDP on RHS captures all of these relationships

Forecasting: Definitions and Notation

Information set at time t, I_t: values of all known variables *up to time t*

Example: If we observe GDP data from 1947 until 2016 Q4 then our information set is $\{GDP_{2016,Q4}, GDP_{2016,Q3} ..., GDP_{1947,Q1}\}$

Forecast of Y_{t+1} is the conditional expectation of Y given one's information set, $E(Y_{t+1}|I_t)$

When using population parameters, denote as: Y_{t+1} or $E(Y_{t+1}|I_t)$

When estimated, denote as: $\hat{Y}_{t+1|t}$ or $E(\widehat{Y_{t+1}|}I_t)$

A **k step ahead forecast** is the forecast of Y_{t+k} given I_t

Remember: Goal is not causality, but rather to capture, parsimoniously, correlations across time to make guesses about future

Example: AR(1) Forecast

What is the forecast produced by an AR(1)?

$$
\begin{aligned}
E(Y_{t+1}|I_t) =& E\left(\beta_0 + \beta_1 \underbrace{Y_t}_{\text{in } I_t} + \underbrace{U_{t+1}}_{\text{not in } I_t} |I_t\right) \\
=& \beta_0 + \beta_1 Y_t + \underbrace{E(U_{t+1}|I_t)}_{=0} \\
=& \beta_0 + \beta_1 Y_t
\end{aligned}
$$

Typically we do not know these values, so we estimate them:

$$
\hat{Y}_{t+1|t} = \hat{\beta}_0 + \hat{\beta}_1 Y_t
$$

EconTutor.com

Forecast Errors

Forecast error is the true value minus the forecast:

$$e_{t+1} = Y_{t+1} - \widehat{E(Y_{t+1}|I_t)}$$

- Typically assume that $t+1$ errors are uncorrelated with past information
- Residuals are *within-sample* error:

$$\hat{u}_t = Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t$$

- Forecast errors are about *out of sample prediction*
  - Nothing about Y_{t+1} is used in estimation
  - In fact, we do not know e_{t+1} until Y_{t+1} becomes known!
- There are **two** sources of uncertainty in forecasting: "true" change residual at $t+1$, U_{t+1} *and* the fact that forecast is estimated

Root Mean Squared Error

Root Mean Squared Forecast Error (RMSE) is given by:

$$\sqrt{E(e_{t+1}^2)} = \sqrt{E\left(\left(Y_{t+1} - E(\widehat{Y_{t+1}|I_t})\right)^2\right)}$$

The RMSFE functions like standard error for forecasts

Think about three objects:

The "true" Y_{t+1}
The population forecast $Y_{t+1|t}$
The estimated forecast $\hat{Y}_{t+1|t}$

Forecast error can be decomposed:

$$e_{t+1} = \underbrace{(Y_{t+1} - Y_{t+1|t})}_{\text{Randomness in } Y} - \underbrace{(Y_{t+1|t} - \hat{Y}_{t+1|t})}_{\text{"Typical" Estimation Error}}$$

Example: AR(1) Forecast Error

What is forecast error produced by an AR(1)?

$$\begin{aligned} e_{t+1} =& Y_{t+1} - \hat{Y}_{t+1|t} \\ =& \beta_0 + \beta_1 Y_t + U_{t+1} - \left(\hat{\beta}_0 + \hat{\beta}_1 Y_t\right) \\ =& \left[\beta_0 - \hat{\beta}_0 + (\beta_1 - \hat{\beta}_1) Y_t\right] + U_{t+1} \end{aligned}$$

Notice that as $T \to \infty$, first term goes to 0

In fact, variance of $Y_{t+1|t} - \hat{Y}_{t+1|t}$ is proportional to $1/\sqrt{T}$

But U_{t+1} *always* present because there is real world randomness

EconTutor.com

Forecast Interval

The Model:

$$Y_{t+1} = E(Y_{t+1}|I_t) + u_{t+1}$$

This breaks up Y_{t+1} into two pieces: the population forecast and the true shock (sometimes called the *innovation*)

Define **forecast interval** as:

$$\hat{Y}_{t+1} \pm 1.96 \times SE(e_{t+1})$$

The range of all values of Y_{t+1} that can occur tomorrow which would not reject model forecast

Notice that we do not use $SE(\hat{Y}_{t+1})$, we use RMSFE

The dominant term in the RMSFE is almost always σ_U^2

Forecast Interval, Cont'd

What is the standard error of forecast error?

$$
\begin{aligned}
Var(e_{t+1}|I_t) &= Var(Y_{t+1} - \hat{Y}_{t+1|t}|I_t) \\
&= Var(\underbrace{E(Y_{t+1}|I_t)}_{\text{Constant at } t!} + u_{t+1} - \hat{Y}_{t+1|t}|I_t) \\
&= \sigma_U^2 + \sigma_{\hat{Y}_{t+1}}^2 + \underbrace{Cov(u_{t+1}, \hat{Y}_{t+1|t})}_{=0} \\
&= \sigma_U^2 + \sigma_{\hat{Y}_{t+1|t}}^2
\end{aligned}
$$

Hence,

$$
CI(\hat{Y}_{t+1}) = \hat{Y}_{t+1|t} \pm \left(\sigma_U^2 + \sigma_{\hat{Y}_{t+1|t}}^2\right)
$$

- We know neither σ_U^2 nor $\sigma_{\hat{e}_{t+1}}^2$
- But we can estimate them (σ_U^2 is SER, while $\sigma_{\hat{e}_{t+1}}^2$ is a bit more complicated)

Example: Forecast Error of an AR(1)

For AR(1):

$$
\begin{aligned}
Var(Y_{t+1} - \hat{Y}_{t+1|t} | I_t) &= Var(\left[\beta_0 - \hat{\beta}_0 + (\beta_1 - \hat{\beta}_1) Y_t\right] + U_{t+1} | I_t) \\
&= \underbrace{Var(\left[\beta_0 - \hat{\beta}_0 + (\beta_1 - \hat{\beta}_1) Y_t\right] | I_t) + Var(U_{t+1} | I_t)}_{\text{Split apart b/c } U_{t+1} \text{ uncorrelated with past}} \\
&= \underbrace{Var(\hat{\beta}_0 + \hat{\beta}_1 Y_t | I_t)}_{\sigma^2_{\hat{Y}_{t+1|t}}} + \sigma^2_U \\
&= \underbrace{\sigma^2_{\hat{\beta}_0} + Y_t^2 \hat{\sigma}^2_{\hat{\beta}_1} + 2 Y_t \sigma_{\hat{\beta}_0, \hat{\beta}_1}}_{\sigma^2_{\hat{Y}_{t+1|t}}} + \sigma^2_U
\end{aligned}
$$

All pieces can be estimated from OLS regression

Estimating RMSFE in Practice

In practice, especially for more complicated models, estimating forecast errors happen in 3 ways:

- Actually try to solve for the proper error (this can be complicated)
- Since $\hat{Y}_{t+1|t}$ typically has low variance, ignore it and focus on σ_U (many statistical packages do this)
- Estimate the forecast error by using data up to time t to estimate a model for each Y_{t+1} *in the data* and use these forecasts (many statistical packages also do this)

- Unlike in causal inference, standard errors here are not about statistical significance
- Accuracy of standard errors is in ensuring a good gauge of one's own ignorance.

Introducing the AR(p) Model

A p^{th} order autoregressive model ($AR(p)$) is given by:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + U_t$$

Clearly AR(1) is special case

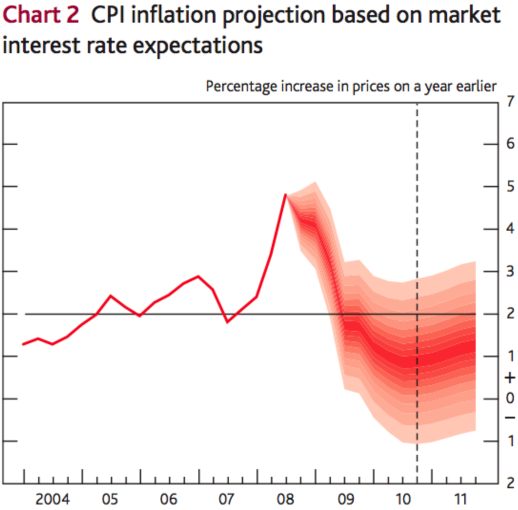
The coefficients still lack causal interpretation

Why might more lags be good?

- Seasonality with
- Business cycles
- Different processes correlated with y can diffuse at different rates

We can estimate $AR(p)$ with multiple regression

A Real Life Example: The Bank of England's Fan Chart



Forecast Error for k-Step Ahead Forecasts

In the graphs above, plot is way further out in the future than 1 period

Uncertainty seems to fan out... but how/why?

We can use the time dependent structure to determine how uncertainty grows over time.

Example with AR(1):

$$\begin{aligned} Var(\hat{Y}_{t+2}|I_t) &= Var(\rho\hat{Y}_{t+1} + U_{t+2}|I_t) \\ &= \rho Var(\hat{Y}_{t+1}|I_t) + \sigma_U^2 \end{aligned}$$

Often times, ignore estimation uncertainty. Then for $AR(1)$:

$$SE(e_{t+k}) = \sigma_U\sqrt{1 + \rho^2 + ... + \rho^{2(k-1)}}$$

Picking Lag Length

The Model:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + U_t$$

How do we pick p?

The dilemma: **bias-variance tradeoff**

Including more lags can only *help* in prediction (lowers bias)

But more lags $\Rightarrow$ less data $\Rightarrow$ noisier forecasts

One way to choose: t and F tests

Do F test for each additional p until final p is statistically insignificant

Tends to lead to overly large models (bias-variance!)

In causality, bias is really bad—but what about in forecasting?

Forecasters often rely on **information criteria** that allow for smoother tradeoff between bias and variance

The Bayes Information Criterion

Define the criteria:

$$BIC(p) = \log\left(\frac{SSR(p)}{T}\right) + (p+1)\frac{\log T}{T}$$

- Bayes information criterion (BIC) chooses p that **minimizes** above function
  - This trades off lowering bias against raising variance
  - Why this function? Because as $T \to \infty$ then $\hat{p}_{BIC} \to p$
- First term measures fit (related to bias)
  - Increasing p will always lower SSR
  - Think of this term as related to the R^2
- Second term is a penalty term (related to variance)
  - Increasing p will always increase this term
  - Think of this as related to the adjusted R^2

The Akaike Information Criterion

An alternative criterion:

$$AIC(p) = \log\left(\frac{SSR(p)}{T}\right) + (p+1)\frac{2}{T}$$

The AIC also has two terms:

First term is the same

Second term is *smaller* (as long as $T > 4$)

- This punishes variance less severely
- Useful if you have reason to believe that long lags are very important or care more about bias than variance

Still want to minimize the AIC

This will *not* converge to p

- But it will always be more conservative

Example Using Unemployment Data

Model:

$$\Delta U_{t+1} = \beta_0 + \beta_1 \Delta U_t + \ldots + \beta_p \Delta U_{t-p}$$

p	BIC	AIC	R^2
0	-2.24	-2.26	0
1	**-2.80**	-2.84	0.450
2	-2.78	-2.84	0.456
3	-2.75	-2.83	0.457
4	-2.76	**-2.86**	0.481
5	-2.73	-2.85	0.482
6	-2.71	-2.85	0.490

The BIC prefers a simple AR(1) model!

The AIC adds more lags, as expected

Focusing on the R^2 alone (and doing t-tests) would pick 4 lags

In the real world, bias-variance tradeoff depends on problem at hand

Adding Covariates: Autoregressive Distributed Lag Models

A (p, r) order autoregressive distributed lag model ($ADL(p, r)$) is given by:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + \gamma_1 X_{t-1} + \ldots + \gamma_r X_{t-r} + U_t$$

- Brings in covariate X
  - We do not have a causal interpretation for the coefficients on X either
- When will an ADL model be useful?
  - Many factors predict unemployment—inflation, GDP, etc.
- Estimate with OLS, and RMSFE extends naturally
  - Extending BIC/AIC is more advanced

Example: Inflation and Unemployment

Estimated Model:

$$\widehat{Inf}_t = .005 + .420 Inf_{t-1} + .002 U_{t-1} - .0004 U_{t-2}$$

Unemployment only weakly correlated with inflation (at the quarterly level)

Forecasting proceeds exactly as before—plugging in lagged values

How to calculate forecast intervals now?

Exactly as before–more complicated, so often times we will just use σ_U^2 and ignore estimation error

Granger Causality

Model:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + \gamma_1 X_{t-1} + \ldots + \gamma_r X_{t-r} + U_t$$

- "Granger Causality" (**really** bad term) tests for whether X has predictive power for Y
- Question: does X have predictive content *beyond* the predictive power of lagged values of Y
  - This is **not** causality
  - Classic example: the price of frozen orange juice futures predicts the weather in Florida better than lagged NOAA measures.
  - If X determines Y *exactly* then X will not predict Y more than lagged Y!
- How to test? Run a joint F test on all coefficients on lagged X's but not on lagged Ys.
  - Include at least many lags of Y as X
  - Additional controls are allowed

Seasonality

Substantial data exhibits **seasonality**: periodicity in relationships that is not captured by simple autoregressive processes

Examples:

- Construction employment is different in summer and winter months
- There is a spending boom every December
- There is a presidential election every 4 years

Not including known periodic relationships can hamper forecasting ability

Solutions:

- **Bad:** Include many, many lags (BIC will not like this)
- **Better:** Include dummies that capture the frequency of seasonality (quarter dummies, month dummies, presidential cycle dummies, etc.)

Simple Non-Stationarity: Deterministic Trends

Trend Model:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \gamma t + U_t$$

When Y_t is a function of t (could also be non-linear) we say that Y_t has a **deterministic trend**

- Data with a trend will *not* be stationary
- Ignoring a trend will tend to give y_{t-1} too little predictive content

Two ways to remove trends:

- Include t as a regressor
- Detrend all data by regressing all variables on t and keeping residuals

What happens to a trend in differences regression?

- Plays the role of the constant!

Later: Breaks in trends and stochastic trends

Spurious Relationships

Model:

$$
\begin{aligned}
Y_t &= \beta_0 + \beta_1 X_{t-1} + \beta_2 t + U_t \\
X_t &= \delta_0 + \delta_1 t + V_t
\end{aligned}
$$

- A **spurious relationship:** Ignoring trends will give *too much* predictive power to x
- Why? Consider regressing Y on X w/o t:

Spurious Relationships

Model:

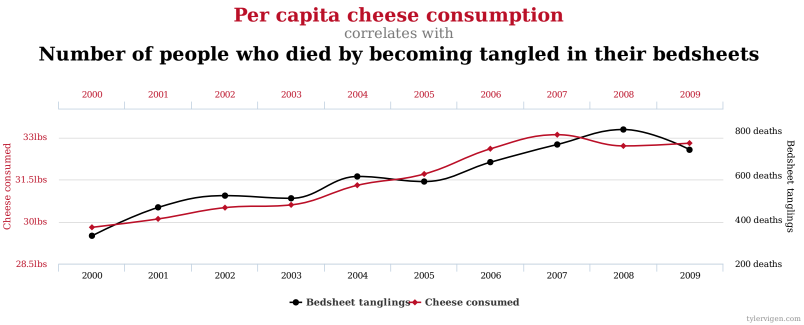
$$Y_t = \beta_0 + \beta_1 X_{t-1} + \beta_2 t + U_t$$
$$X_t = \delta_0 + \delta_1 t + V_t$$

Question: Assume that U and V are independent and are uncorrelated to X_t for all t. Find the bias of the OLS estimator when you regress Y_t on X_{t-1}.

Spurious Relationship: Example

Spurious Correlation Bias (in simple model):

$$\hat{\beta}_1 \to \beta_1 + \beta_2/\delta_1$$



- Bias depends on trends
- Related issue: trends violate finite 4th moments assumption