UCI Paul Merage School of Business

Leadership for a Digitally Driven World™

MFIN 290: Financial Econometrics

Lecture 4-1

Final Projects – Groups and Topics

- There are two components to the final projects.
- A 20-30 min presentation to the class: evaluated by your peers (function of group size)
  - What question you are trying to answer?
  - What data do you have?
  - How are you able to identify this effect (if relevant)?
  - Key exhibits/conclusions (graphs are nice)
  - Highlight techniques from the course
  - Conclusions, options for further investigation.
- Be sure to allow time for questions!
- Include this same content in a paper. 5-10 pages is a good target, but I am not the kind of person that counts pages.

Last Time

- Time Series Introduction
- White Noise Process
- AR(1), MA(1), derivation of mean, variance, autocovariances
- Autocorrelegrams, Partial Autocorrelgrams
  - Using these to tell the difference between AR(1) and MA(1)
  - Confidence Interval construction: pointwise v. repeat inference adjusted
  - Testing for residual autocorrelation
  - Order Selection criteria

Repeat Inference

This kind of summary can be very helpful… remember the exam is open notes!

Example: Forecasting

- Let's say we have fitted an AR(2) GDP model like we had last time:

- $g_t = \hat{\alpha} + \widehat{\beta_1} g_{t-1} + \hat{\beta}_2 g_{t-2} + u_t$

- $g_t = 0.465 + 0.377 g_{t-1} + 0.2464 g_{t-2} + u_t$

- And we are interested in the forecast for the first two out of sample periods (which here are 2009Q4 and 2010Q1).

Example: Forecasting

- $g_t = 0.465 + 0.377 g_{t-1} + 0.2464 g_{t-2} + u_t$

- $g_{t+1} = 0.465 + 0.377(0.8) + 0.2464(-0.2) + u_t$
- $g_{t+1} = 0.718$

- $g_{t+2} = 0.465 + 0.377 g_{t+1} + 0.2464 g_t + u_t$

- $g_{t+2} = 0.465 + 0.377(0.718) + 0.2464(0.8) + u_t$
- $g_{t+2} = 0.933$

```
. list g obsno if obsno >=95

     +--------------+
     |    g   obsno |
     |--------------|
 95. | -1.4      95 |
 96. | -1.2      96 |
 97. |  -.2      97 |
 98. |   .8      98 |
     +--------------+
```

Prediction Intervals

- We can use the forecasted values to continue our forecasts. Obviously this has consequences for uncertainty.

- To investigate this, let's begin by reviewing prediction uncertainty and prediction intervals.

Prediction Intervals

Imagine that we have a model that predicts expected default rates, losses given default, etc. and want to know if the model's predictions are consistent with realizations out of sample.

There are two kinds of uncertainty here:

1. Uncertainty in the expected value of y given $x = (var(\hat{y}))$
2. Uncertainty in what y will be realized conditional on a correct expected value $(var(y) - var(\hat{y}) = var(\epsilon))$

The second is usually *much* larger than the first.

Best Linear Unbiased Predictor

Matrices can make life easy

$$var(\hat{y}|x) = var(xb|x) = xvar(b)x' = \sigma^2 x(x'x)^{-1}x'$$

$$var(y) = var(\hat{y}) + var(\epsilon) = \sigma^2(x(x'x)^{-1}x' + I)$$

$$cov(\hat{y}, \epsilon) = 0 \; since \; cov(x, \epsilon) = 0$$

Will go through the standard derivation from the text for reference and for intuition

Prediction – Single Variable

The task of predicting y_0 is related to the problem of estimating

$$E\ (y_0)\ = \beta_1\ +\ \beta_2 x_0$$

Although $E\ (y_0)$ is not random (given X), the actual outcome y_0 is random

The least squares predictor of y_0 comes from the fitted regression line

$$y_0\ = \widehat{\beta_1}\ + \widehat{\beta_2} x_0$$

It turns out that our OLS procedure generates predictions with nice properties…

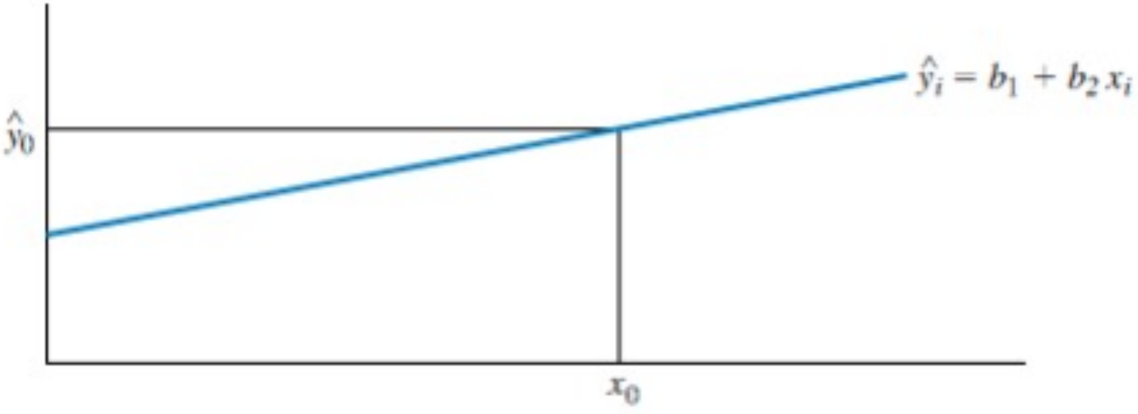
Prediction Intervals Intuition

$$var(f) = var(\widehat{y_0}) + var(\widehat{\epsilon_0}) = \sigma^2[1 + \frac{1}{N} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}]$$

This will be smaller when

- σ^2 is smaller
- N is larger
- The variance in X, $\sum(x_i - \bar{x})^2$, is larger
- The value in question is near the mean, or $(x_0 - \bar{x})^2$ is smaller

Prediction



Define the forecast error as :

$$\hat{\epsilon}_0 = y_0 - \widehat{y_0}$$
$$= (\beta_1 + \beta_2 x_0 + \epsilon_0) - (\widehat{\beta_1} + \widehat{\beta_2} x_0)$$

Prediction

The forecast error has mean zero:

$$E(\hat{\epsilon}_0) = E(y_0 - \widehat{y_0})$$
$$= E(\beta_1 + \beta_2 x_0 + \epsilon_0) - E(\widehat{\beta_1} + \widehat{\beta_2} x_0)$$
$$= \beta_1 + \beta_2 x_0 - E\left(\widehat{\beta_1} + \widehat{\beta_2} x_0\right) = 0$$

This means that $\widehat{y_0}$ is an unbiased predictor of y_0 … if we have our GM assumptions, then $\widehat{y_0}$ is also the best linear unbiased predictor (BLUP) of y_0

Best Linear Unbiased Predictor

The variance of the predicted values of $\widehat{y_0}$

$= var(\widehat{y_0})$

$$= var(\widehat{\beta_1} + \widehat{\beta_2} x_0)$$

$$= var(\widehat{\beta_1}) + x_0^2 var(\widehat{\beta_2}) + 2x_0 cov(\widehat{\beta_1}, \widehat{\beta_2})$$

$$= \sigma^2 \frac{\sum x_i^2}{N \sum (x_i - \bar{x})^2} + x_0^2 \frac{\sigma^2}{\sum (x_i - \bar{x})^2} + 2x_0 \sigma^2 \frac{-\bar{x}}{\sum (x_i - \bar{x})^2}$$

Best Linear Unbiased Predictor

The variance of the predicted values of $\widehat{y_0}$

$= var(\widehat{y_0})$

$$= \sigma^2 \frac{\sum x_i^2}{N\sum(x_i - \bar{x})^2} + x_0^2 \frac{\sigma^2}{\sum(x_i - \bar{x})^2} + 2x_0\sigma^2 \frac{-\bar{x}}{\sum(x_i - \bar{x})^2}$$

Can simplify if we add and subtract a common factor:

$$= \sigma^2 \frac{\sum x_i^2}{N\sum(x_i - \bar{x})^2} - \sigma^2 N \frac{\bar{x}^2}{N\sum(x_i - \bar{x})^2}$$

$$+ x_0^2 \frac{\sigma^2}{\sum(x_i - \bar{x})^2} + 2x_0\sigma^2 \frac{-\bar{x}}{\sum(x_i - \bar{x})^2} + \sigma^2 N \frac{\bar{x}^2}{N\sum(x_i - \bar{x})^2}$$

Best Linear Unbiased Predictor

$$= \sigma^2 \frac{\sum x_i^2}{N\sum(x_i-\bar{x})^2} - \sigma^2 N \frac{\bar{x}^2}{N\sum(x_i-\bar{x})^2}$$

$$+x_0^2 \frac{\sigma^2}{\sum(x_i-\bar{x})^2} + 2x_0\sigma^2 \frac{-\bar{x}}{\sum(x_i-\bar{x})^2} + \sigma^2 N \frac{\bar{x}^2}{N\sum(x_i-\bar{x})^2}$$

$$=\sigma^2 \frac{\sum(x_i-\bar{x})^2}{N\sum(x_i-\bar{x})^2} + \sigma^2 \frac{(x_0-\bar{x})^2}{\sum(x_i-\bar{x})^2}$$

$$\Rightarrow \ var(\widehat{y_0}) = \sigma^2 [\frac{1}{N} + \frac{(x_0-\bar{x})^2}{\sum(x_i-\bar{x})^2}]$$

Best Linear Unbiased Predictor

Since the covariance of x and ϵ are zero under our assumptions,
The variance of the forecast is $var(f) = var(\widehat{y_0}) + var(\widehat{\epsilon_0}) =$
$= var(\widehat{y_0}) + \sigma^2$

$$= \sigma^2 \left[\frac{1}{N} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right] + \sigma^2$$

$$\sigma^2 [1 + \frac{1}{N} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}]$$

We can estimate this using $\widehat{\sigma^2}$, our usual estimate for the variance in the errors!

Prediction Intervals Intuition

$$var(f) = var(\widehat{y_0}) + var(\widehat{\epsilon_0}) = \sigma^2[1 + \frac{1}{N} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}]$$

This will be smaller when

- σ^2 is smaller
- N is larger
- The variance in X, $\sum(x_i - \bar{x})^2$, is larger
- The value in question is near the mean, or $(x_0 - \bar{x})^2$ is smaller

Prediction Intervals

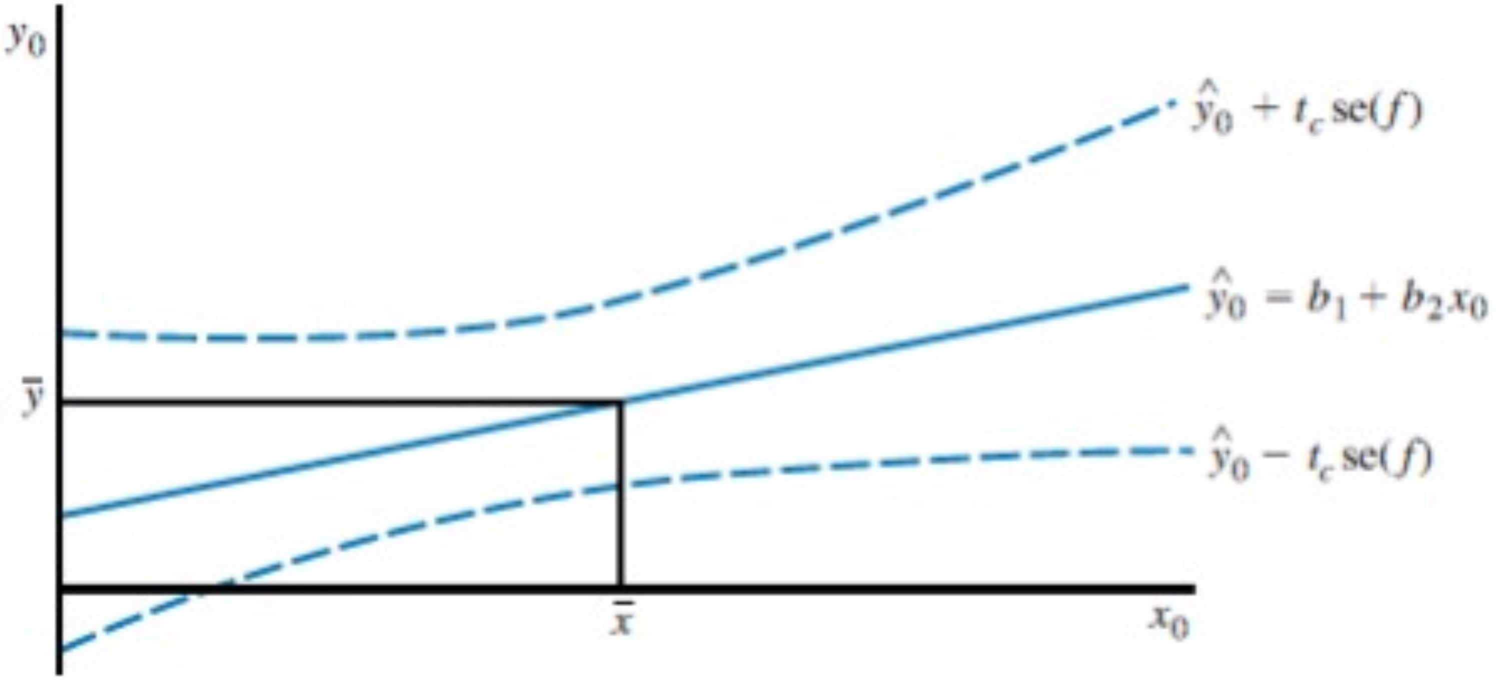
Just like anything else, with the mean and the estimated variance of the predictions, we can construct a t statistic!

$$t = \frac{y - y_0}{\sqrt{var(\hat{f})}} = \frac{(y - y_0)}{SE(\hat{f})} \sim t(N-2)$$

And get critical values and construct prediction intervals

$\Pr(-t_c \leq \frac{(y - y_0)}{SE(f)} \leq \mathrm{t_c}) = 1 - \alpha$

Prediction Intervals



Example: Forecasting

- $g_t = \hat{\alpha} + \widehat{\beta_1} g_{t-1} + \hat{\beta}_2 g_{t-2} + u_t$
- $g_t = 0.465 + 0.377 g_{t-1} + 0.2464 g_{t-2} + u_t$
- $g_{t+1} = 0.465 + 0.377(0.8) + 0.2464(-0.2) + u_t$
- $g_{t+1} = 0.718$

- $g_{t+2} = 0.465 + 0.377 g_{t+1} + 0.2464 g_t + u_t$

- $g_{t+2} = 0.465 + 0.377(0.718) + 0.2464(0.8) + u_t$
- $g_{t+2} = 0.933$

```
. list g obsno if obsno >=95

     +-------------+
     |    g  obsno |
     |-------------|
 95. | -1.4     95 |
 96. | -1.2     96 |
 97. |  -.2     97 |
 98. |   .8     98 |
     +-------------+
```

Example: Forecasting

- Forecasting error in the first period given by:
- $v_{t+1} = g_{t+1} - \hat{g}_{t+1} = (\alpha - \hat{\alpha}) + (\beta_1 - \widehat{\beta_1})g_t + (\beta_2 - \widehat{\beta_2})g_{t-1} + u_{t+1}$
- If the random error (u_{t+1}) is large relative to the estimation error, we can essentially ignore it (I will do that here for exposition).
- $v_{t+1} \cong u_{t+1}$

Example: Forecasting

- Forecasting error in the second period includes the random shock in the second period, and the error that occurs by using the forecasted values of $\hat{g}_{t+1}$ instead of the true g_{t+1}.

- Ignoring the estimation error, this simplifies to:
- $v_{t+2} = (g_{t+1} - \hat{g}_{t+1})\widehat{\beta_1} + u_{t+2} = v_{t+1}\widehat{\beta_1} + u_{t+2}$
- $v_{t+2} = u_{t+1}\widehat{\beta_1} + u_{t+2}$

Example: Forecasting

- For three periods we have two sets of forecasting errors we care about since the model is 2nd order:
- $v_{t+3} = (g_{t+2} - \hat{g}_{t+2})\widehat{\beta_1} + \ (g_{t+1} - \hat{g}_{t+1})\widehat{\beta_2} + u_{t+3}$
- $v_{t+3} \ = v_{t+2}\widehat{\beta_1} + v_{t+1}\ \widehat{\beta_2} + u_{t+3}$
- $v_{t+3} \ = (u_{t+1}\widehat{\beta_1} + u_{t+2})\ \widehat{\beta_1} + u_{t+1}\ \widehat{\beta_2} + u_{t+3}$
- $v_{t+3} \ = u_{t+1}(\widehat{\beta_1^2} + \widehat{\beta_2}) + u_{t+2}\ \widehat{\beta_1} + u_{t+3}$

Example: Forecasting

- Each of these expressions can give us variances at different points in time.

- Defining $var(u_t) = \sigma_u^2$ for all t, we have:

$$v_{t+1} = u_{t+1} \qquad \Rightarrow var(v_{t+1}) = \sigma_u^2$$

$$v_{t+2} = u_{t+1}\widehat{\beta_1} + u_{t+2} \qquad \Rightarrow var(v_{t+2}) = \sigma_u^2(1 + \widehat{\beta_1}^2)$$

$$v_{t+3} = u_{t+1}(\widehat{\beta_1^2} + \widehat{\beta_2}) + u_{t+2}\,\widehat{\beta_1} + u_{t+3} \Rightarrow var(v_{t+3}) = \sigma_u^2(\left(\widehat{\beta_1}^2 + \widehat{\beta_2}\right)^2 + \widehat{\beta_1}^2 + 1)$$

Example: Forecasting

These will increase as we move into the future as the errors compound. They can get very wide, very quickly.

Should always review how these do and will evolve to differentiate between candidate time series models.

Exercise:

Should compare the forecasts and forecast intervals (assuming zero estimation error) for the first three periods using sample data for an AR(1) and AR(2) GDP model…will do one of them now…

Example

```
. reg g l.ehat l2.ehat

      Source |       SS           df       MS      Number of obs   =        96
-------------+----------------------------------   F(2, 93)        =     19.06
       Model |  11.6417916         2  5.82089582   Prob > F        =    0.0000
    Residual |  28.4081042        93  .305463486   R-squared       =    0.2907
-------------+----------------------------------   Adj R-squared   =    0.2754
       Total |  40.0498958        95  .421577851   Root MSE        =    .55269

------------------------------------------------------------------------------
           g |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ehat |
         L1. |   .3770015     .100021     3.77   0.000     .1783797    .5756233
         L2. |   .2462394    .1028688     2.39   0.019     .0419623    .4505165
             |
       _cons |   1.261312    .0564417    22.35   0.000      1.14923    1.373394
------------------------------------------------------------------------------

. predict ehat_l2, resid
(2 missing values generated)
```

Example: Forecasting

$$\sqrt{var(v_{t+1})} = \sqrt{\sigma_u^2} = \sigma_u = 0.552$$

$$\sqrt{var(v_{t+2})} = \sqrt{(\sigma_u^2 \left(1 + \widehat{\beta_1}^2\right))} = 0.590 \ (\text{cov}\ (\sigma_u^2, \widehat{\beta_1}) = 0)$$

$$\sqrt{var(v_{t+3})} = \sqrt{(\sigma_u^2((\widehat{\beta_1}^2 + \widehat{\beta_2})^2 + \widehat{\beta_1}^2 + 1))} = 0.628$$

Quarter	Forecast G	SE (Forecast)	95% Forecast Interval ($\pm$1.985 SEs)
2009Q4	0.718	0.552	(-0.379,1.816)
2010Q1	0.933	0.590	(-0.239,2.106)
2010Q2	0.994	0.628	(-0.254,2.242)

Back to Time Series

- In the previous lecture, we derived expressions for the autocovariances of AR(1) and MA(1) relationships and described how these could be used to differentiate between the two.
- The AC and PAC functions for AR(1) and MA(1) were particularly well-behaved, though there are obviously series out there without this property.
- In order for us to conduct meaningful time series analysis, we need to place some restrictions on the processes in question that we have up to now ignored…

Stationary Series

- We call a time series y_t”weakly stationary” if:

1. $E(y_t)$ is independent of t
2. $Var(y_t)$ is a finite, positive constant independent of t
3. $Cov(y_t, y_s)$ is a finite function of $|t - s|$, but not t or s

- This is also called “covariance stationary”; should give some clue as to the relative importance of the components …

Stationary Series

- There also exists "strict stationarity", which requires only one very strong thing: the full joint distribution of the time series is independent of t.

- This distinction almost never comes up… it's possible to come up with examples of series that are covariance stationary but not strictly stationary (i.e., higher moments may depend on t or not exist), but they are rare.

- Why do we have two definitions? Any non-linear function of a strictly stationary series is strictly stationary (This should make sense with some thought on independence…).

Non-Stationary Series

- Examples of series that are not covariance stationary:
- EX 1:
- $y_t = \mu + \varepsilon_t \qquad for\ t \leq k$
- $y_t = \mu + \lambda + \varepsilon_t \quad for\ t > k$

- Which of the properties does this NOT have?

Non-Stationary Series

- Examples of series that are not covariance stationary:
- EX 1:
- $y_t = \mu + \varepsilon_t \qquad for\ t \leq k$
- $y_t = \mu + \lambda + \varepsilon_t \quad for\ t > k$

- $E(y_t)$ depends on $t \Rightarrow$ mean depends on $t \Rightarrow$ not covariance stationarity

- *This happens in our data all the time!*

Non-Stationary Series

- EX2: “Random Walk”
- $y_t = y_{t-1} + \varepsilon_t$
-
- Why?
- $Var(y_t) = Var(y_{t-1} + \varepsilon_t)$
- $Var(y_t) = Var(y_{t-1}) + Var(\varepsilon_t) + 2Cov(y_t, \varepsilon_t)$
- $Var(y_t) = Var(y_{t-1}) + Var(\varepsilon_t) > Var(y_{t-1})$
-
- Variance grows with t $\Rightarrow$ not independent of t

Non-Stationary Series

- EX3:
- $y_t = t + \beta y_{t-1} + \varepsilon_t$

- Trended series are not stationary. This is a common (the most common?) example.

Identification

- While non-stationarity is usually thought of as a problem, it can also be helpful!
- Sometimes non-stationarity can be exploited: look for the effects of policy, announcements, etc. “Impact Analysis/Regression Discontinuity Design”.
- Identify periods before and after policy changes, look at the jump in the time series.
- This is literally what is done with a difference in difference estimator.

Ergodicity

- In addition to Stationarity, we need one more property… Ergodicity
- In English, Ergodicity basically requires that if events are separated by enough time, they become asymptotically independent.
- In a time series, this means that every observation contains at least *some* new information.

Ergodicity

- A time series process $\{z_t\}_{-\infty}^{\infty}$ is ergodic if for any two bounded functions that map vectors in the n and m dimensional real vector spaces to real scalars, $f: R^n \to R^1$ and $g: R^m \to R^1$, we have

$$\lim_{k\to\infty} |E(f(z_t, z_{t+1}, \dots, z_{t+n})g(z_{t+k}, z_{t+k+1}, \dots, z_{t+k+m}))| =$$
$$|E(f(z_t, z_{t+1}, \dots, z_{t+n}))||E(g(z_{t+k}, z_{t+k+1}, \dots, z_{t+k+m}))|$$

- The (absolutely value of the) expectation of the product (of any two bounded functions of any two points in the space) is the product of the expectations.

The Ergodic Theorem

- Why do we care?
- The Ergodic Theorem. This is essentially a law of large numbers for time series.

- If a series $\{z_t\}_{-\infty}^{\infty}$ is stationary and ergodic with $E[|z_t|]$ a finite constant and $E[z_t] = \mu$, and we define $\bar{z}_T = \frac{1}{T}\sum_{t=1}^{T} z_t$, then $\bar{z}_T$ converges almost surely to μ
- Why is this important?

The Ergodic Theorem

- If a series $\{z_t\}_{-\infty}^{\infty}$ is stationary and ergodic with $E[|z_t|]$ a finite constant and $E[z_t] = \mu$, and we define $\bar{z}_T = \frac{1}{T}\sum_{t=1}^{T} z_t$, then $\bar{z}_T$ converges almost surely to μ

- Allows us to use sample moments, ensures our estimators will converge, etc.

Spurious Regression: Intuition

- Take the model
- $Y = X\beta + \varepsilon$

Where y and x are both time series, usual assumptions on the error term.

Intuitively, unless the Y's are a white noise process, the β cannot be zero.

- Imagine Y has a trend through time.

Any non-zero drift in the X's will produce large significant coefficient estimates….

In that case, it doesn't really make sense to run a hypothesis test on the coefficients in this model…

Spurious Regressions in Economics

- Example:

Egyptian infant mortality rate (Y), 1971-1990, annual data, on Gross aggregate income of American farmers (I) and Total Honduran money supply (M), where the values of the key statistics are: R2 = .918, F = 95.17.

Spurious Regressions in Economics

- Granger, Newbold (1973/1974)

- Before this paper, often used to see regressions of different economic statistics on one another with very high t-stats and very low p values. R^2 statistics over 95-99% were also common.

- The problem is that most economic series are *trending.*

- The fact that you found a relationship between the S&P 500 and butter production in SE Asia should not make you want to open a hedge fund…

Spurious Regressions in Economics

- Granger, Newbold (1973/1974)

- They simulate *independent* series:

$$y_t = \phi y_{t-1} + u_t$$
$$x_t = \phi' x_{t-1} + v_t$$

- They run a simulation and calculate the fraction of the time the simple statistics would (incorrectly) reject a null hypothesis of $\beta = 0$ in the regression of y on x
-
- Simple to follow and see that there is an issue here.

Exploiting Non-Stationarity

- Often, non-stationarity is thought of in a negative context.

- *This need not be the case*. If we understand the cause of the non-stationarity, we can sometimes use it to identify effects.

- Regression Discontinuity Design can inform the effects of policy changes, of certain actions, etc. Explicitly relies on changes in the sample moments over time to identify the relationships.

First Differencing

- First differencing makes random walks stationary, but not all non-stationary series can be made stationary by differencing!
- Often the first thing people try, but only helpful only in a subset of circumstances. Recommend writing down the consequences of a differenced model on the coefficients that you should recover under various assumptions.
- We will do just that now …

First Differencing

- Imagine everything is stationary and ergodic, and we can run the following regression without upsetting Granger and Newbold…

- $Y_t = X_t\beta + \varepsilon_t$

- If we difference,
- $Y_t - Y_{t-1} = X_t\beta + \varepsilon_t - Y_{t-1}$
- $Y_t - Y_{t-1} = (X_t - X_{t-1})\beta_{FD} + \varepsilon_t - \varepsilon_{t-1}$
- Should get the “same” value for β and β_{FD} … but FD will be less efficient.

First Differencing

- With a single variable
- $var(\hat{\beta}) = \frac{\sigma^2}{var(x)}$

- $var\ (\hat{\beta}_{FD}) = \frac{var(\varepsilon_t - \varepsilon_{t-1})}{var(X_t - X_{t-1})}$

- Usually numerator is higher (2x if independent), denominator is lower.
- Should get the same betas, but one estimator is efficient? Sounds like a Hausman Specification test!

Correcting Non-Stationarity

- 1) RW with drift:
- $y_t = \mu + y_{t-1} + e_t$

- 2) Trend Stationary:
- $y_t = \mu + \beta \times t + e_t$

- 3) RW:
- $y_t = y_{t-1} + e_t$

Differencing (1) or (3) gives a white noise series. Differencing (2) generates an MA process in the residuals…

Detrending (regressing on time, extracting residuals) will obviously work on (2), but will obviously not help in (1) or (3), though may be (erroneously) statistically significant given the spurious inference mentioned earlier… so what to do?

Testing for Stationarity

- So stationarity is important, and knowing which kind of non-stationarity we have is necessary to apply the right fix as we just saw.

- How do we identify when we have nonstationary series? Let’s combine the earlier examples (and add in some flexibility with an additional parameter, γ):

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

Testing for Stationarity

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

- Subtract y_{t-1} from each side,
- $y_t - y_{t-1} = \mu\gamma + \beta\gamma t + (\gamma - 1)y_{t-1} + e_t$

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

Testing for Stationarity

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

- If $\gamma = 1$ & $\alpha_1 = 0 \Rightarrow$ RW (and we difference to resolve)
- If $\gamma < 1$ & $\alpha_1 \neq 0 \Rightarrow$ trending (and we have to detrend)

- This structure lies behind common stationarity tests.

Dickey Fuller Test 1

- No Constant and No Trend
- $y_t - y_{t-1} = (\gamma - 1)y_{t-1} + e_t$
- If $\gamma = 1 \ \Rightarrow$ RW (and we difference to resolve)
- Null Hypothesis here is non-stationarity (a RW in particular). If we reject => conclude is not a RW at desired level of confidence

Inference in DF Tests

- Problem is that under the null, the estimates from the DF and ADF regressions no longer come from a t-distribution… if the series is non-stationary, the variance of y increases with the sample size!

- This test has specific critical values, and the statistics here are usually referred to as “tau” statistics instead of t for that reason.

Dickey Fuller Test 2

- With Constant and No Trend
- $y_t - y_{t-1} = \alpha_0 + (\gamma - 1)y_{t-1} + e_t$
- If $\gamma = 1 \Rightarrow$ RW (and we difference to resolve)

Dickey Fuller Test 3

- With Constant and With Trend

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

- If $\gamma = 1$ & $\alpha_1 = 0 \Rightarrow$ RW (and we difference to resolve around the trend)
- If $\gamma < 1$ & $\alpha_1 \neq 0 \Rightarrow$ trending (and we have to detrend)

Augmented Dickey Fuller Test

- With Constant and With Trend

- $y_t - y_{t-1} = \alpha_0 + \alpha_1 t + (\gamma - 1)y_{t-1} + \sum \beta_s (y_{t-s} - y_{t-s-1}) + e_t$

- If $\gamma = 1$ & $\alpha_1 = 0 \Rightarrow$ RW (and we difference to resolve around the trend)
- If $\gamma < 1$ & $\alpha_1 \neq 0 \Rightarrow$ trending (and we have to detrend)
- Allows for autocorrelation in the dependent variable ⇔ autocorrelation in the error terms.

Choosing DF Tests

- If the series is wandering around with mean zero => form 1
- If the series is wandering around with a non-zero mean => form 2
- If the series is wandering around a linear trend => form 3