2019 Essex Summer School

3K: Dynamics and Heterogeneity

Robert W. Walker, Ph. D.
Associate Professor of Quantitative Methods
Atkinson Graduate School of Management
Willamette University
Salem, Oregon USA
rwalker@willamette.edu

August 7, 2019

Three Standard Time-Serial Structures [ARIMA]

AutoRegressive Integrated Moving Average (ARIMA) structures characterize most time series of interest (virtually all with the integration of their seasonal counterparts). In general, we write

- Autoregression [AR(p)]:

$$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \cdots + \rho_p e_{t-p} + \nu_t$$

- Moving Average [MA(q)]:

$$e_t = \nu_t + \theta_1 \nu_{t-1} + \theta_2 \nu_{t-2} + \cdots + \theta_p \nu_{t-q}$$

- Autoregression and Moving Average [ARMA(p, q)]:

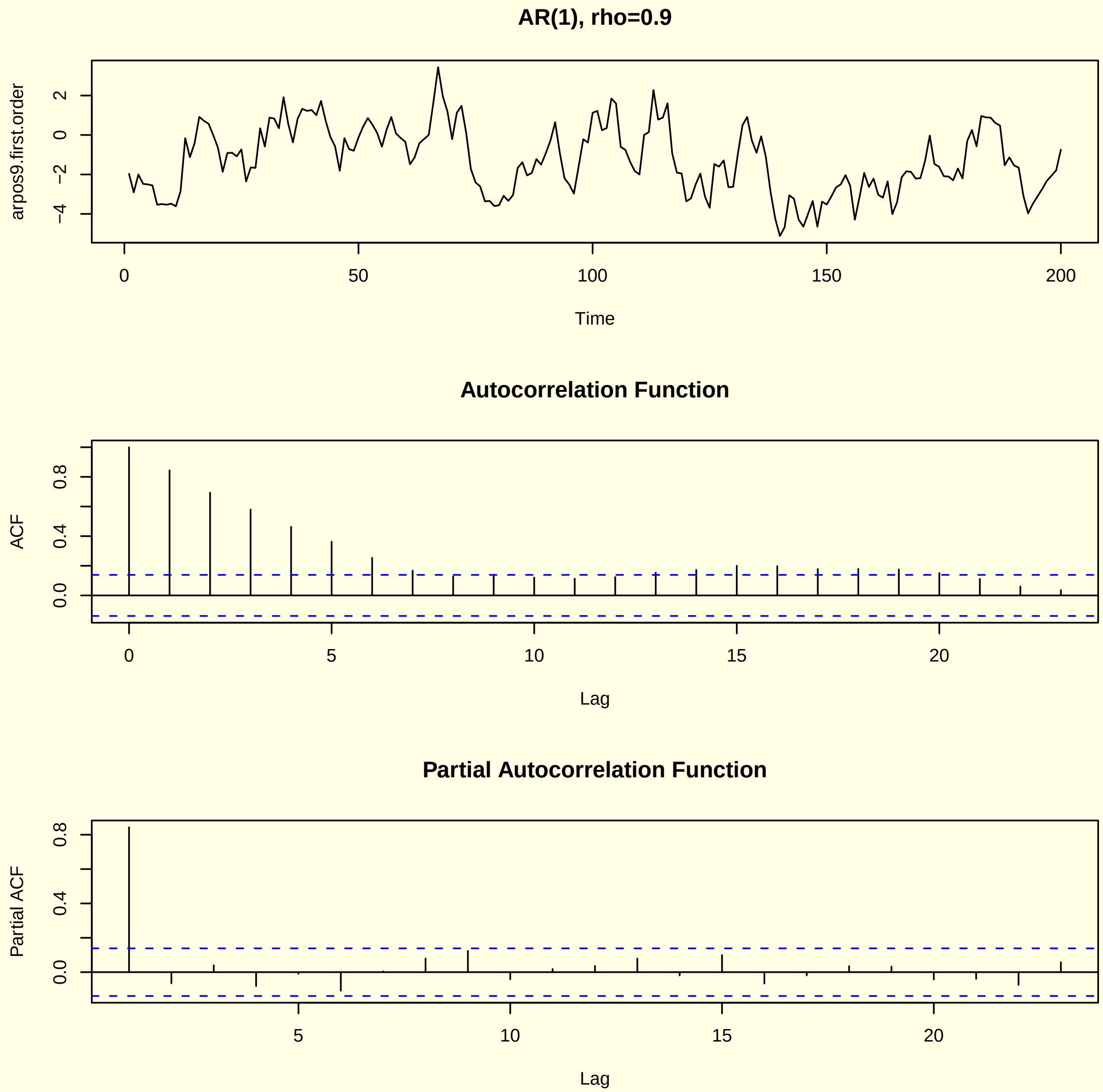
$$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \cdots + \rho_p e_{t-p} + \nu_t + \theta_1 \nu_{t-1} + \theta_2 \nu_{t-2} + \cdots + \theta_q \nu_{t-q}$$

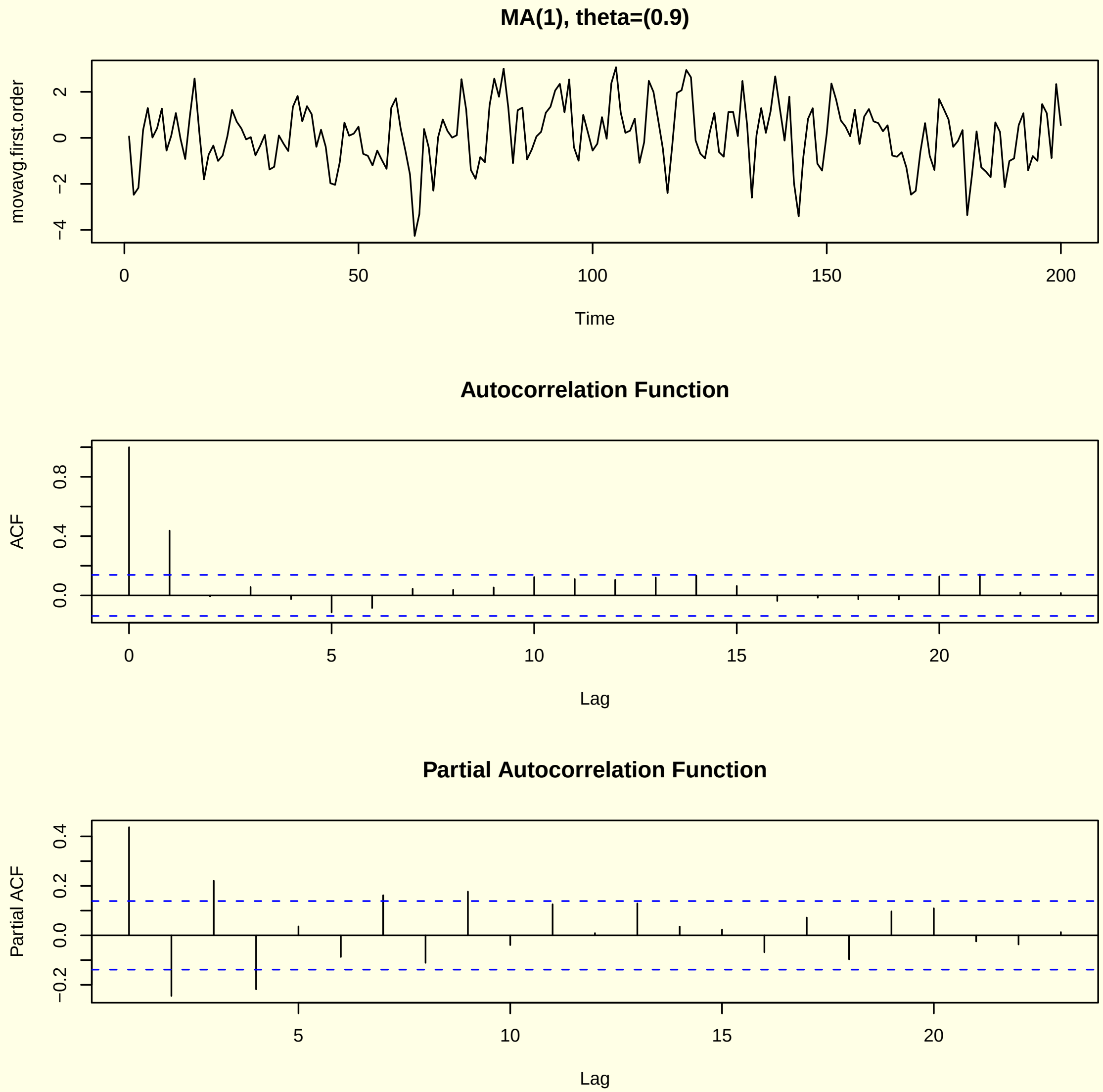
Before, I mentioned two relevant autocorrelations and discussed them briefly:

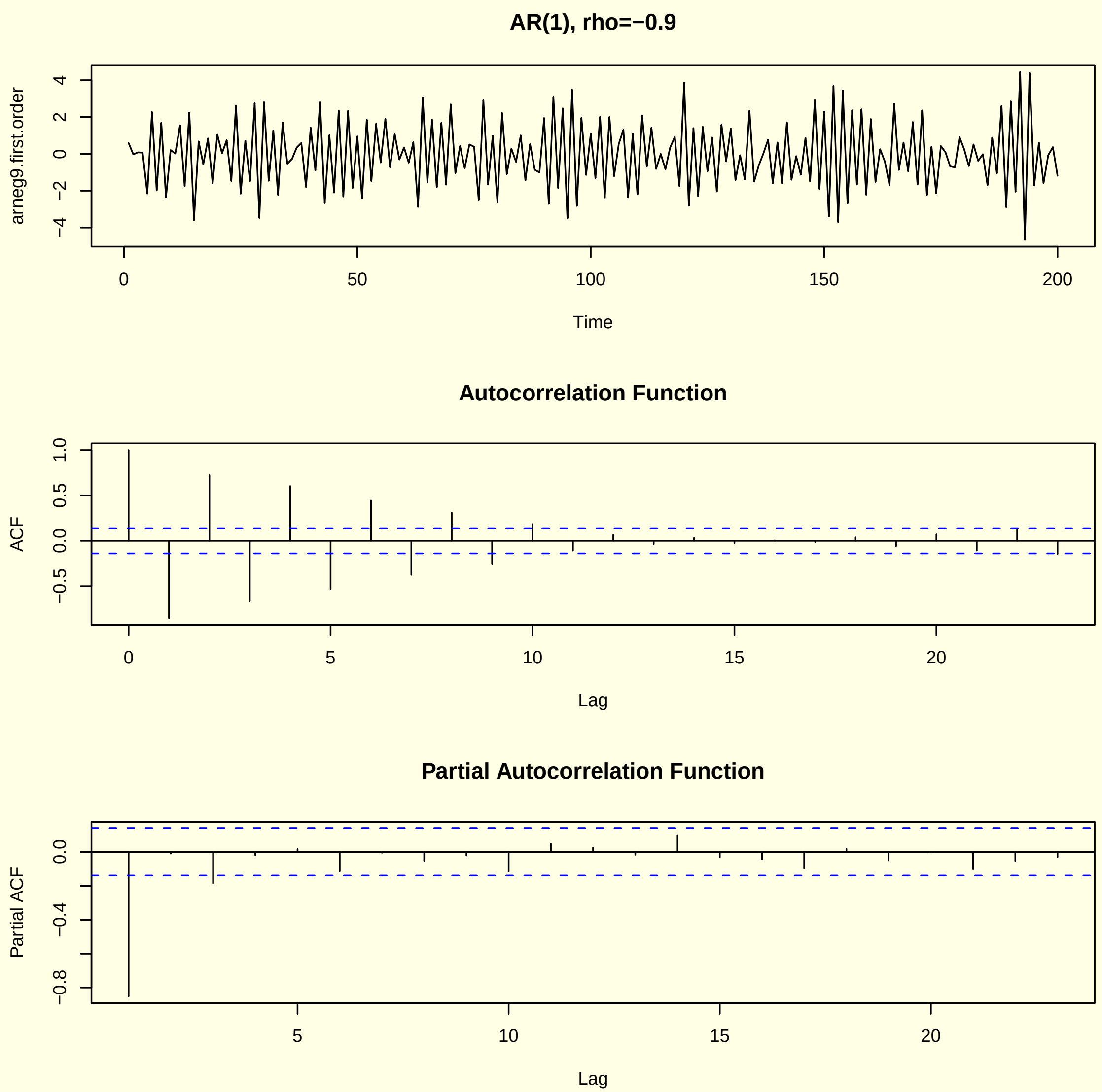
1. Autocorrelation: $\rho_s = \frac{\sum_{t=s+1}^{T}(y_t - \overline{y})(y_{t-s} - \overline{y})}{\sum_{t=1}^{T}(y_t - \overline{y})^2}$

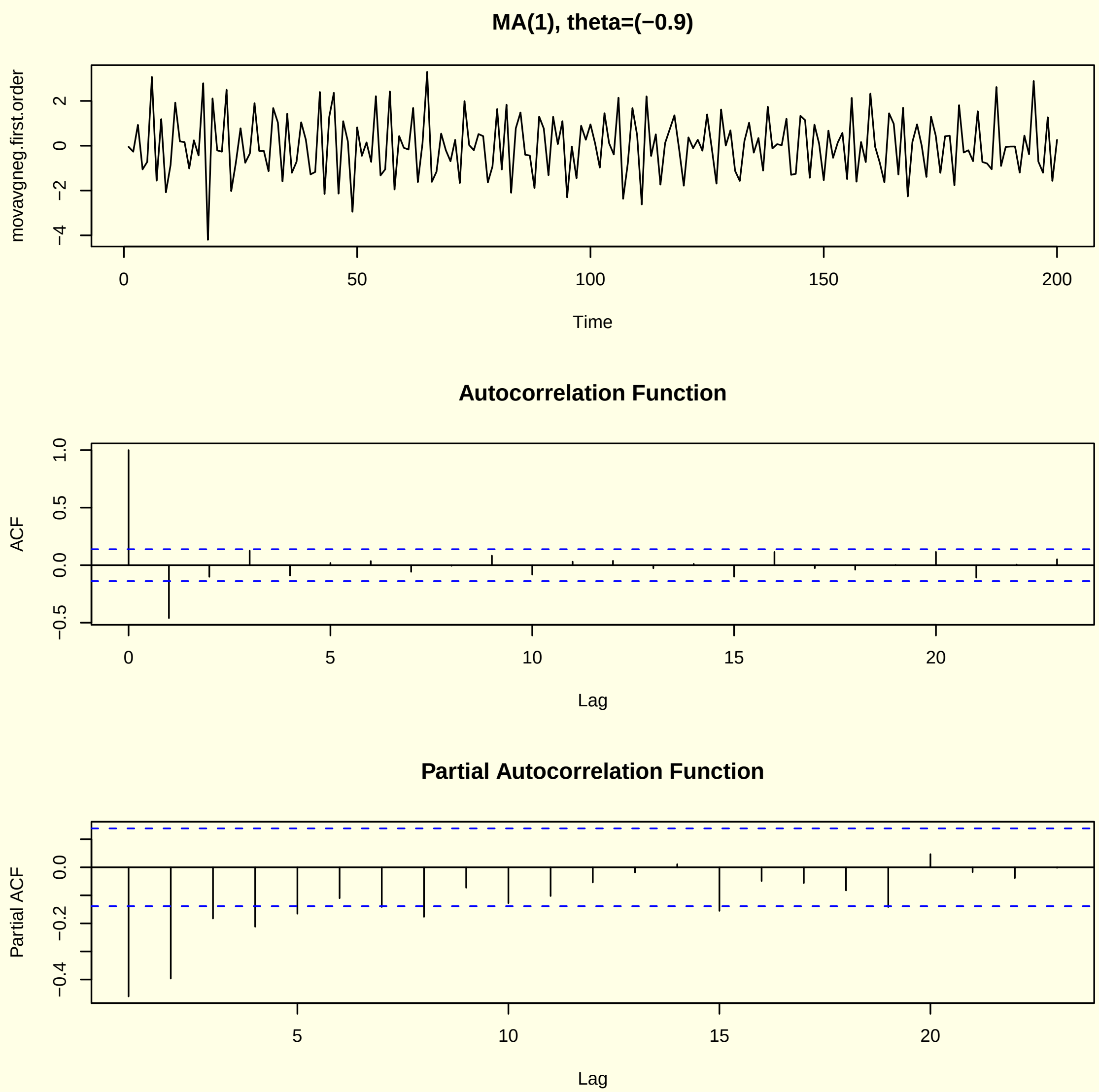
2. Partial Autocorrelation $\phi_s = \frac{\rho_s - \sum_{j=1}^{s-1}\phi_{s-1,j}\rho_{s-j}}{1 - \sum_{j=1}^{s-1}\phi_{s-1,j}\rho_j}$

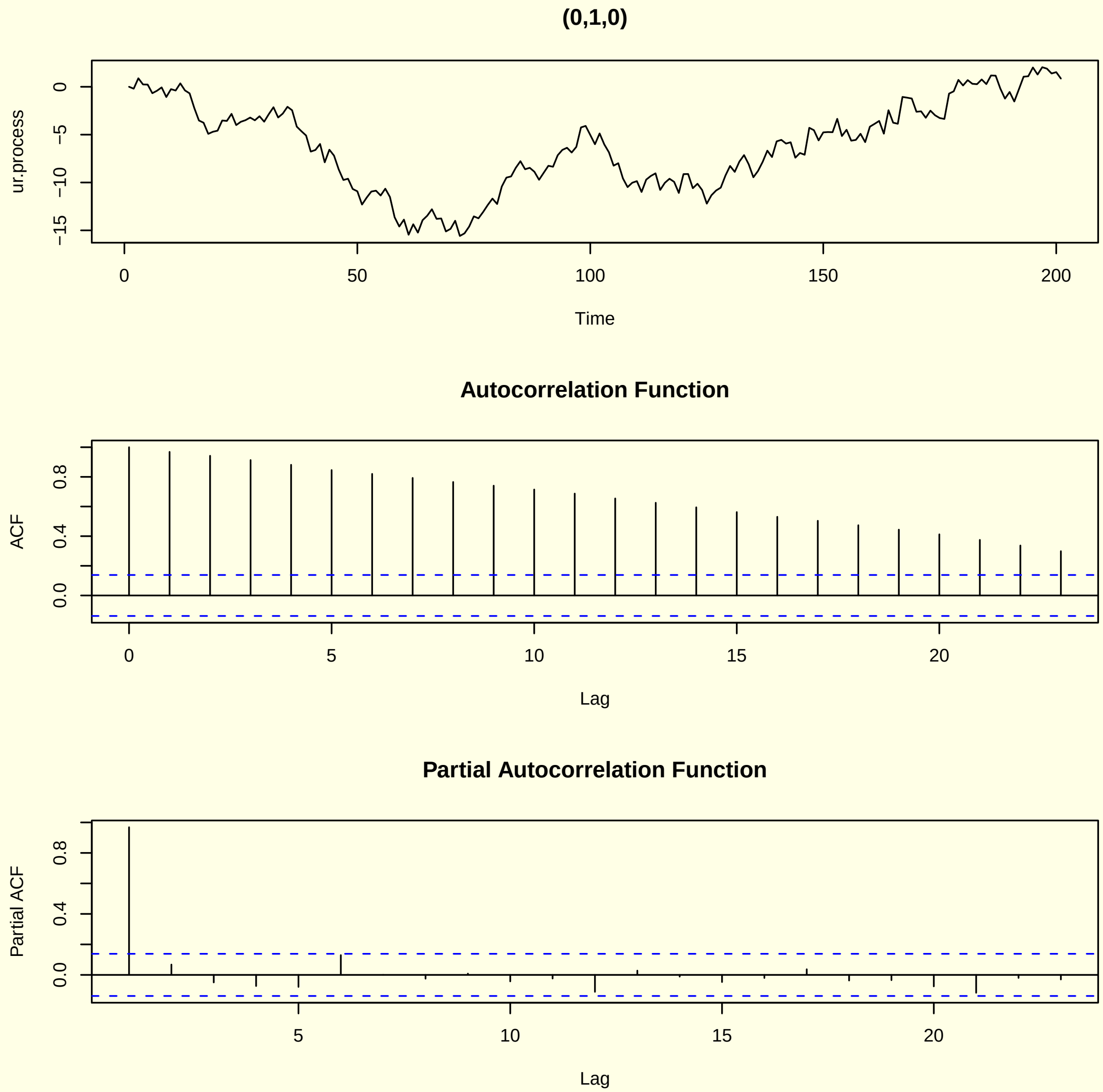
In ARIMA modeling, these are two critical components as each process has a characteristic signature. An autoregressive process typically exhibits geometric decay in the autocorrelation function and spikes in the partial; moving average processes exhibit the reverse. Nonstationary series decay very slowly (the I in ARIMA).


AR(1), rho=0.9
arpos9.first.order
−4
−2
0
2
0
50
100
150
200
Time
Autocorrelation Function
ACF
0.0
0.4
0.8
0
5
10
15
20
Lag
Partial Autocorrelation Function
Partial ACF
0.0
0.4
0.8
5
10
15
20
Lag


MA(1), theta=(0.9)
movavg.first.order
Time
Autocorrelation Function
ACF
Lag
Partial Autocorrelation Function
Partial ACF
Lag


AR(1), rho=−0.9
arneg9.first.order
Time
Autocorrelation Function
ACF
Lag
Partial Autocorrelation Function
Partial ACF
Lag


MA(1), theta=(−0.9)
movavgneg.first.order
−4
−2
0
2
0
50
100
150
200
Time
Autocorrelation Function
ACF
−0.5
0.0
0.5
1.0
0
5
10
15
20
Lag
Partial Autocorrelation Function
Partial ACF
−0.4
−0.2
0.0
5
10
15
20
Lag


(0,1,0)
ur.process
0
−5
−10
−15
0
50
100
150
200
Time
Autocorrelation Function
ACF
0.0
0.4
0.8
0
5
10
15
20
Lag
Partial Autocorrelation Function
Partial ACF
0.0
0.4
0.8
5
10
15
20
Lag

Stata

Though these plots were generated in R, we could do the same thing in Stata. For a quick summary with a little graphic, have a look at the `corrgram`. For (pretty) plots, Stata has two commands to recreate this, `ac` and `pac`. The former generates the autocorrelations while the latter creates the partial autocorrelations. We will have a go at this in the lab.

TSCS and Time Series

- Common structure restrictions may be difficult to deal with and limit our ability to gain much from combining individual time series.
- Most will be pretty simple structures.
- Mixed orders of integration present special problems.

Diagnosing Serial Correlation Individually

If we can reject a range of pathologies, we can justify inference rationally?

- First question is the integrity of the estimand; does the conditional mean make sense?

- Unit root tests come in a host of forms with nulls of a unit root and nulls of stationarity. The processes have different implications. Unfortunately, in TSCS/CSTS settings, tests are pretty unreliable. That said,

  ⋆ Levin and Lin: `levinlin` with $H_0 : I(1)$.
  ⋆ Im, Pesaran, and Shin: `ipshin` with $H_0 : I(1)$.
  ⋆ KPSS: `kpss` with $H_0 : I(0)$.
  ⋆ Fisher: `xtfisher` works with unbalanced panels
  ⋆ Simple `xtreg` with lagged y, if $\beta_{y_{t-1}} \approx 1$ then there is a worry.

- Given this:
  - ⋆ Plots (Every structure has different theoretical ACF/PACF)
  - ⋆ Durbin-Watson d and Durbin's h with endogenous variables
  - ⋆ Dickey-Fuller tests and many others. $\Delta y_t = \rho y_{t-1} + \theta_L \Delta y_{t-L} + \lambda_t + u_t$
  - ⋆ Breusch-Godfrey test and the like (Fit regression, isolate residuals, regress residual on X and lags of residual, $nR^2 \sim \chi^2_p$).

- The above alongside:
  (1) is the temporal process common or distinct? and
  (2) if distinct, how and why?

Panel Unit Root Testing in Stata

As of Stata 11, a battery of panel unit-root tests have emerged. There are many and they operate under differing sets of assumptions.

- Levin-Lin-Chu (`xtunitroot llc`): trend nocons (unit specific) demean (within transform) lags. Under (crucial) cross-sectional independence, the test is an advancement on the generic Dickey-Fuller theory that allows the lag lengths to vary by cross-sections. The test relies on specifying a kernel (beyond our purposes) and a lag length (upper bound). The test statistic has a standard normal basis with asymptotics in $\frac{\sqrt{N_T}}{T}$ (T grows faster than N). The test is of either all series containing unit roots (H_0) or all stationary; this is a limitation. It is recommended for moderate to large T and N.

1. Perform separate ADF regressions:

$$\Delta y_{it} = \rho_i \Delta y_{i,t-1} + \sum_{L=1}^{p_i} \theta_{iL} \Delta y_{i,t=L} + \alpha_{mi} d_{mt} + \epsilon_{it}$$

with d_{mt} as the vector of deterministic variables (none, drift, drift and trend). Select a max L and use t on $\hat{\theta}_{iL}$ to attempt to simplify. Then use $\Delta y_{it} = \Delta y_{i,t-L}$ and d_{mt} for residuals

- Harris-Tzavalis (`xtunitroot ht`): trend nocons (unit specific) demean (within transform) altt (small sample adjust) Similar to the previous, they show that $T \to \infty$ faster than N (rather than T fixed) leads to size distortions.

- Breitung (`xtunitroot breitung`): trend nocons (unit specific) demean (within transform) robust (CSD) lags.
Similar to LLC with a common statistic across all i.

- Im, Pesaran, Shin (`xtunitroot ips`): trend demean (within transform) lags. They free ρ to be ρ_i and average individual unit root statistics. The null is that all contain unit roots while the alternative specifies at least some to be stationary. The test relies on sequential asymptotics (first T, then N). Better in small samples than LLC, but note the differences in the alternatives.

- Fisher type tests (`xtunitroot fisher`): dfuller pperron demean lags.

- Hadri (LM) (`xtunitroot hadri`): trend demean robust

All but the last are null hypothesis unit-root tests. Most assume balance but the fisher and IPS versions can work for unbalanced panels.

Day VI: Review, Summary, and to Missing Data

Stationarity Issues

- Essence of stationarity is threefold: means, variances, and crosses are not time-dependent.
- There is a quite famous spurious regressions result in econometrics that owes to the statistician Yule in 1926.
- Basically, the regression of $I(1)$ series on one another has non-α rejection rates.
- Applied to panels, a mix of orders of integration will give t statistics non-t properties.
- In the end, I suspect the best advice is to partition data on the basis of likely orders of integration and proceed from there.

ADL/Canonical models

We can consider some very basic time series models.

- Koyck/Geometric decay:
  short run and long-run effects are parametrically identified (given $\mathcal{M}$).

- Almon (more arbitrary decay):

$$y_{it} = \sum_{t_A=0}^{T_F} \rho_{t_A} x_{t-t_A} + \epsilon_t$$

  with coefficients that are ordinates of some general polynomial of degree $T_F >> q$. The $\rho_{t_A} = \sum_{k=0}^{T_F} \gamma_k t^k$.

- Prais-Winston, etc. are basically FGLS implementations of AR(1).

Prais-Winsten/Cochrane-Orcutt

$$y_{it} = X_{it}\beta + \epsilon_{it}$$

where

$$\epsilon_{it} = \rho\epsilon_{i,t-1} + \nu_{it}$$

and $\nu_{it} \sim N(0, \sigma_\nu^2)$ with stationarity forcing $|\rho| < 1$. We will use iterated FGLS. First, estimate the regression recalling our unbiasedness condition. Then regress $\hat{\epsilon}_{it}$ on $\hat{\epsilon}_{i,t-1}$. Rinse and repeat until ρ doesn't change. The transformation applied to the first observation is distinct, you can look this up.... In general, the transformed regression is:

$$y_{it} - \rho y_{i,t-1} = \alpha(1-\rho) + \beta(X_{it} - \rho X_{i,t-1}) + \nu_{it}$$

with ν white noise.

Beck

- Static model: Instantaneous impact.

$$y_{i,t} = X_{i,t}\beta + \nu_{i,t}$$

- Finite distributed lag: lags of x finite horizon impact (defined by lags).

$$y_{i,t} = X_{i,t}\beta + \sum_{k=1}^{K} X_{i,t-k}\beta_k + \nu_{i,t}$$

- AR(1): Errors decay geometrically, X instantaneous. (Suppose unmeasured x and think this through).

$$y_{i,t} = X_{i,t}\beta + \nu_{i,t} + \theta\epsilon_{i,t-1}$$

- Lagged dependent variable: lags of y [common geometric decay]

$$y_{i,t} = X_{i,t}\beta + \phi y_{i,t-1} + \nu_{i,t}$$

- ADL: current and lagged x and lagged y.

$$y_{i,t} = X_{i,t}\beta + X_{i,t-1}\gamma + \phi y_{i,t-1} + \epsilon_{i,t}$$

- Panel versions of transfer function models from Box and Jenkins time series. (each x has an impact and decay function)

Interpretation of dynamic models

- Do it.
- Whitten and Williams `dynsim` uses `Clarify`[1] to do this.
- Their paper is "But Wait, There's More! Maximizing Substantive Inferences from TSCS Models". Easy to find on the web and on the website.

[1] If you do not know what Clarify is, please ask: estimate, set, simulate.

On Granger Causation

Granger causation[2] in a panel data setting. This is a rather complex topic because of heterogeneity.

- An identical causal relationship could exist for each $i \in N$
- No causal relationship could exist for any $i \in N$.
- Something in between the above two extremes.

Matching the theoretical claim and the empirical analysis here, like everywhere else, is absolutely crucial. Also, in some ways, this is just our earliest ANOVA example but now we will use the lags of the dependent variable to establish the alternative hypotheses.

[2] In the remainder of this discussion, when I use the word cause, I mean it in a Granger sense which may differ dramatically from more common understandings of causal.

Implementation

- Look at stationarity. We want to make certain that this holds.
- Test a hypothesis, many forms can be implied.

Conditional on stationarity, the procedure goes like this. For example, to test the hypothesis that, for all i, x does not Granger cause y, we choose a lag length, call it k. We then regress lags of y and k lags of x on y in a model with unit specific slopes and unit specific intercepts. We do not include contemporaneous x because Granger causality is all about temporal priority. This is the unrestricted model. The hypothesis implies the restriction that all the coefficients on all of the lagged x are zero for each cross-section and for all k lags of x.

In other words, the unrestricted model is:

$$y_{it} = \rho_{i,1} y_{i,t-1} + \rho_{i,2} y_{i,t-2} + \ldots + \rho_{i,m} y_{i,t-m} + \beta_{i,1} x_{i,t-1} + \ldots + \beta_{i,k} x_{i,t-k} + \epsilon_{it}$$

while the restricted model is

$$y_{it} = \rho_{i,1} y_{i,t-1} + \rho_{i,2} y_{i,t-2} + \ldots + \rho_{i,m} y_{i,t-m} + \epsilon_{it}$$

because the null hypothesis implies that all the β are zero for all lags $t-1$ to $t-k$ lags of x.