Time Series

Nathaniel Beck
Department of Political Science
University of California, San Diego
La Jolla, CA 92093
beck@ucsd.edu
http://weber.ucsd.edu/∼nbeck

April, 2001

TIME SERIES ISSUES

Consider a model $y_t = x_t\beta + \epsilon_t, t = 1, \ldots, T$.

Errors, ϵ_t are said to be serially correlated if ϵ_t is not independent of other $\epsilon_{t'}$. In practice we consider only certain ways in which this can occur. It is usually a problem in times series. Why? But there are situations where it occurs in cross-sections. *Spatial autocorrelation* - cf. Anselm, Spatial Econometrics. We will only consider time series issues.

• Old fashioned view: serial correlation as a nuisance that causes G-M assumptions to be violated and leads to non-optimality of OLS.

If you do OLS and errors show serial correlation than:

- OLS standard errors are wrong, perhaps very wrong
  - Query: what does it mean for a standard error to be wrong
  - Are the standard errors parameters of the model, so that there is a "correct" standard error?
- OLS estimates (of β are no longer efficient (best, optimal)
  - What does this mean?
- But OLS estimates of β are still unbiased and consistent

Note: the problem is not knife-edged. A small amount of serial correlation causes small problems, a large amount causes large problems. We would seldom, if ever, expect to see a time series with literally no serial correlation of the errors. (REMEMBER WHAT ANY STATISTICAL TEST CAN SHOW ABOUT THIS)

• Modern view - model the dynamics as part of the model, not an estimation nuisance (See Hendry and Mizon, 1978 or Beck, 1993)

Simple model

$$y_t = x_t\beta + \epsilon_t \tag{1}$$

$$\epsilon_t = \rho\epsilon_{t-1} + \nu_t. \tag{2}$$

where ν_t is iid (independent, identically distributed). These errors are said to follow a first order autoregressive (AR1) process.

We usually don't think of the second equation as part of the model.

Can transform these two into

$$y_t = x_t\beta + \rho y_{t-1} - \rho\beta x_{t-1} + \nu_t \tag{3}$$

which looks like the ADL (autoregressive distributed lag) model

$$y_t = x_t\beta + \rho y_{t-1} + \gamma x_{t-1} + \nu_t. \tag{4}$$

Autoregressive because of lagged y, distributed lag because of lagged x.

THUS we can either model the dynamics directly rather than treat them as a "nuisance."

Testing for serial correlation

Either assume specific form and test or find a more general test without assuming form - former has more POWER against that alternative

Usual assumption

AR1 $\epsilon_t = \rho\epsilon_{t-1} + \nu_t$

MA1 $\epsilon_t = \lambda\mu_{t-1} + \mu_t$

(Easy to handle and test for higher order lags)

No lagged dependent variable

Usual - Durbin- Watson (a transformation of the empirical correlation of e_t and e_{t-1}, note the move from ϵ to e, don't confuse the two! Durbin-Watson runs from 0-4, with 2 being no serial correlation (Durbin is brilliant, but why he didn't do -2 to 2 with 0 being no serial correlation is beyond me!). Need special tables, in back of any book. But no reason to bother with these days because . . .

LM tests

All LM tests look like running a regression under the null and then regressing residuals or some function of residuals on some iv's. The test statistic is TR^2 which has χ^2 distribution with appropriate df. Need to look at specific cases, though can figure out how to design in general.

Test for serially correlated errors: run OLS and then regress residual on lagged residual and other iv's, test if coefficient on lagged residual is significant

Lagged dependent variable

Usual - Durbin's h but . . .

Modern - LM - same as above, but add lagged y to the second (auxiliary) regression

Higher order lags - just include the corresponding residuals in aux. reg.

Note: LM test for MA1 and AR1 serial correlation identical!

LM useful for a wide variety of things - note the battery of tests in EVIEWS.

What to do?

If no serial correlation, OLS is fine (wonderful, in fact), even for ADL model. Issues are interpretation and deciding how to model lags.

If serial correlation and no lagged y, can use GLS (Prais-Winsten), maximum likelihood (best, will see later on) or Cochrane-Orcutt (common, but not quite as good as Prais-Winsten, but not bad)

If lagged y you have problems - why?

Because lagged y is correlated with error term, first round OLS estimates not consistent, GLS fails.

Fortunately, in practice, models with lagged dv often show no serial correlation. Why?

(But note Chris Achen argument!)

Even if they show a bit of serial correlation, cure is almost certainly worse than the disease.

Cure is IV (instrumental variables). But usually instruments are horrible.

GLS estimation

If OLS is not optimal because of error term issues (as with serially correlated errors), can often fix with Generalized Least Squares (GLS). GLS finds some transform of the original model so that the transformed model meets G-M assumptions, then do OLS on transformed model, untransform to get parameters of interest if necessary.

Easiest to see in practice, let us look at for AR1 errors.

$$y_t = x_t\beta + \epsilon_t \tag{5}$$

$$y_{t-1} = x_{t-1}\beta + \epsilon_{t-1} \tag{6}$$

If we subtract off ρ times the second equation from both sides, we get

$$y_t - \rho y_{t-1} = x_t\beta - \rho\beta x_{t-1} + \{\epsilon_t - \rho\epsilon_{t-1}\} \tag{7}$$

But $\epsilon_t = \rho\epsilon_{t-1} + \nu_t$ so the term in braces is just ν_t which is iid by assumption. Hence if we transform by subtracting ρ times the prior observation from the current observation, call this $y_t^*(= y_t - \rho y_{t-1}$ and $x_t^*(= x_t - \rho x_t t - 1$, we have an nice G-M observing equation $y_t^* = \beta x_t^* + \nu_t$ and we can then get β by OLS.

UNFORTUNATELY we don't know ρ. FORTUNATELY, we can do OLS on original equation, get our consistent but non-optimal estimate of β, use that to compute residuals, and then take the correlation of e_t and e_{t-1} as our estimate of ρ and then do GLS with that estimate. Since we are using an estimate of ρ instead of a known ρ this is called Feasible GLS. (We never do other than FGLS.)

Cochrane-Orcutt drops the first observation (because can't compute y_1^* and then proceeds to iterate, estimate β, compute a ρ, use that to get a better β, use that to get a new ρ, stop when converges. Usually will converge in 1-10 iterations.

Full GLS (Prais-Winsten) keeps first observation and doesn't iterate. Is more computationally intensive (10 milliseconds instead of 1), but is now standard.

ARCH

There are other time series complications of the error, in particular AutoRegressive Conditional Heteroskedasticity invented by Rob Engle. Here it is the variance of the errors that shows time dependence, if you have a big error one month you have a big error the next. So if σ_t^2 is the variance of the t'th error (conditional on observations $1, \ldots, t-1$, then

$$\sigma_t^2 = \kappa\epsilon_{t-1}^2 + \kappa_0 \tag{8}$$

for one period, can generalize easily (Guj., p. 437).

Can test for via LM test, regress squared residuals on lagged squared residuals, if present can correct for. This has been incredibly important in financial econometrics (Citation Classic) but, I think, is less important in PS.

Issues in ADL models: no serial corr

First choice - finite vs infinite distributed lags

Contrast

$$y_t = \beta_1 x_t + \beta_2 x_{t-1} + \epsilon_t \text{(iid)} \tag{9}$$

with

$$y_t = \beta x_t + \rho y_{t-1} + \epsilon_t \text{(iid)} \tag{10}$$

We can rewrite the second as

$$y_t = \beta x_t + \rho\beta x_{t-1} + \rho^2 \beta x_{t-2} + \rho^3 \beta x_{t-3} + \ldots + \epsilon_t \tag{11}$$

IT IS THESE DOTS THAT ARE CRITICAL. THIS REWRITING TRICK WILL SERVE YOU IN GOOD STEAD

Note, if the iid assumption on errors is correct, then estimating either type of model is statistically trivial (ols).

Finite distributed lag is difficult to estimate in practice since lagged x's are likely to be highly correlated. May not be a problem for two lags, but tough for longer.

How do we choose?

Stories (err, theory)? Error terms become SHOCKS!

Put less cynically, the data often (most often) do not clearly discriminate between these various models, and there is often little theory to fall back on. Thus we often choose on matters like convenience.

There are also lots of other more complicated models, but they are not commonly used.

Partial adjustment

You have a desired y_t^* for a given x, but it is costly to move all the way to the desired level in one period (say it involves buying machines which you can't throw away the next period).

$$y_t^* = \beta x_t \quad (12)$$
$$y_t - y_{t-1} = \lambda(y_t^* - y_{t-1}) + \mu_t \quad (13)$$

yielding

$$y_t = \lambda\beta x_t + (1-\lambda)y_{t-1} + \mu_t \quad (14)$$

which is our ADL model.

Exponential distributed lag

Another story is due to Koyck. Suppose lags decline exponentially so

$$y_t = \alpha + \beta x_t + \beta\rho x_{t-1} + \beta\rho^2 x_{t-2} \cdots + \epsilon_t \quad (15)$$

and

$$y_{t-1} = \alpha + \beta x_{t-1} + \beta\rho x_{t-2} + \beta\rho^2 x_{t-3} \cdots + \epsilon_{t-1} \quad (16)$$

yielding

$$y_t = \alpha + \beta x_t + \rho y_{t-1} - \rho\alpha + \epsilon_t - \rho\epsilon_{t-1} \quad (17)$$

which looks like an ADL model with a constrained MA1 error (note the coefficient ρ).

Model is even simpler if we think of errors as shocks which also have a lagged effect which declines exponentially, that is

$$\epsilon_t = \nu_t + \phi\nu_{t-1} + \phi^2\nu_{t-2} + \ldots \quad (18)$$
$$= \nu_t + \phi\epsilon_{t-1} \quad (19)$$

Now the most plausible assumption is that $\phi = \rho$, since why should shocks die out at a different rate than the x, since shocks are just unmeasured variables. At that point the Koyck model, transformed, is exactly the ADL model with iid errors.

Can allow first lag (or few) to be free, e.g.

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \rho y_{t-1} + \epsilon_t \tag{20}$$

What do we do with one than one iv? If all decline in effect at same rate, than no problem, Koyck transformation is fine. If decline at different rate, complicated but doable.

Finite lags

No violation of Gauss-Markov assumptions. Problem is that the lagged iv's are often highly collinear. If only using two may be no problem. If more than that, most popular technique is to parameterize the lags, using only a slight amount of structure. Simplest assumption is that the lags follow a low order (often quadratic) polynomial. This method is due to Almon.

Thus we replace

$$\beta_0 x_t + \beta_1 x_{t-1} + \cdots + \beta_k x_{t-k} \tag{21}$$

by

$$\sum_{i=0}^{k} (a + bi + ci^2) x_{t-i} \tag{22}$$

Higher polynomial degrees allow for wider range of shapes, but quadratic is far and away the most common choice.

Obviously the method is atheoretical. But it often works.

If we don't want to use some parameterization, can try alternative lag lengths and pick best based on some criteria such as Akaike Information Criteria (AIC) or Schwarz Criteria (SC). All these criteria look at sum of square errors plus a penalty for number of parameters - they vary in this penalty. The ones with most severe penalty (e.g. Schwarz) seem to work best.

The two criteria are:

$$\text{AIC} = \ln \hat{\sigma}^2 + \frac{2k}{T} \tag{23}$$

$$\text{SC} = \ln \hat{\sigma}^2 + \frac{k \ln T}{T} \tag{24}$$

where T is number of observations and k is number of parameters. You can see that SC prefers more parsimonious models.

Atheoretical approaches

Single variable - Box Jenkins (ARMA)

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_k y_{t-k} + \nu_t + \phi_1 \nu_{t-1} + \cdots + \phi_m \nu_{t-m} \tag{25}$$

This is an Autoregressive Moving Average (ARMA) process of order (k,m). Box-Jenkins have a whole methodology for picking the "best" such model and diagnostics for goodness of fit. May be of interest in policy work for prediction (works pretty well) but no theory.

Note - need stationarity (see below), if not can often get by differencing, leading to ARIMA, I stands for Integrated. Just difference to get stationarity, then do ARMA modelling. Usual standard for stationarity is whether autocorrelations decline over time.

See Gujarati, ch. 22 for details on this methodology.

Vector Autoregression (VAR)

Set of variables, no theory. Model each variable as a function of own lags and lags on other variables, using F-tests to pick appropriate lengths. Great implementation in RATS. Estimate via OLS.

$$y_t == \sum_{i=1}^{k} \beta_i y_{t-i} + \sum_{j=1}^{k} \gamma_j x_{t-j} + \kappa_1 \tag{26}$$

$$x_t == \sum_{i=1}^{k} \delta_i x_{t-i} + \sum_{j=1}^{k} \phi_j y_{t-j} + \kappa_2 \tag{27}$$

with obvious generalization for more than two variables. No theory, complete asymmetry.

(Granger causality: x causes y if x belongs in first equation. More generally, x Granger causes y if x helps you predict y over and above what you could predict from knowing the history of y alone.)

Unit Roots and Non-Stationarity

A series is stationary if its properties (mean, variance, covariances) don't change with time. We need stationarity because when we average to compute means and variances, if not stationary then are adding up different things to compute mean.

In particular, y is covariance stationary (we seldom worry about higher moments, possibly because of "normal" thinking) if (note the lack of subscripts on right hand side)

1. $E(y_t) = \mu$
2. $\text{Var}(y_t) = \sigma^2$
3. $\text{Cov}(y_t, y_{t-s}) = \phi_s$

A series is I(1) (integrated of order 1) if it is not stationary but its first difference is, with I(2) and so forth defined in the obvious way.

Consider the series y_t with

$$y_t = y_{t-1} + \epsilon_t \tag{28}$$

$$= \epsilon_t + \epsilon_{t-1} + y_{t-2} \tag{29}$$

$$= \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} + \dots \tag{30}$$

Note that

$$\text{Var}(y_t) = \sum_{i=0}^{t-1} \text{Var}(\epsilon_{t-i}) \tag{31}$$

$$= t\sigma^2 \tag{32}$$

so series is not stationary (variance is increasing) and variance of $y_t \xrightarrow[t\to\infty]{} \infty$ which clearly causes problems for standard analysis.

Note that if ϵ is stationary, this series is I(1) since $\Delta(y_t) = \epsilon_t$.

Properties

Integrated series have nasty properties in addition to this "infinite" variance. In particular they wander very far from the mean of the series and have no tendency to return to the mean (as opposed to stationary series, which wander around their mean).

Thus, for example, it will be hard to estimate a model like Equation 1 if x or y is integrated. If y is integrated but x is not then ϵ must be integrated (the sum of two stationary series is integrated). If x is integrated than y cannot be stationary since the sum of an integrated and stationary series is integrated. (Integration dominates: think about adding together two series, one of which nicely wanders about mean, other wanders far off for huge periods of time.)

It is a little hard to test for unit roots, since if a series has a unit root than the standard distribution theory is wrong (not being able to deal with infinite variances and such). But the logic is easy. Regress y_t on y_{t-1} and look at coefficient. If near one, suspect unit root. But instead of $\frac{\rho-1}{\text{se}}$ having a t distribution, it has a "Dickey-Fuller" distribution. EVIEWS takes care of this, if not any new text has D-F tables.

Note: One can either regress y_t on y_{t-1} and ask if the coefficient is near one or regress Δy_t on y_{t-1} and ask if the resulting coefficient is near zero, make sure it is obvious that two forms are identical.

One must be careful. Note that one gets "integration" if one fails to reject $H_0 : \rho = 1$ which is backwards from the way we usually proceed. So with small samples, may get integration result because of large standard error.

Can series like presidential approval, which run from 0-100, be integrated. It can't have infinite variance!

What if only looking at one upturn of a long cycle - would think it integrated, but a longer time slice would show stationary?

Near integration, fractional integration and stationary series that change VERY slowly (long memoried). These are new areas of research.

Cointegration

The only hope is if x and y are *cointegrated*. Loosely, two integrated series are cointegrated if each is integrated but their difference is stationary (so that each wanders pretty wildly, but they wander close to each other).

A set of variables is said to be cointegrated if each one alone is integrated (say I(1)) but a linear combination of them is stationary.

Consider two variables, x and y. First test each to see if integrated. If so, then regress y on x. If cointegrated residuals will stationary (and the estimated $\hat{\beta}$ will be the "co-integrating value." Note that since y and $x\beta$ wander together, the two variables are in a long run equilibrium, in that if x changes y will also change. This can be very useful.

Estimation

To estimate, we can use Engle-Granger two step.

1. Estimate cointegrating relationship by regressing y on x using OLS. (Why is this okay?)
2. Take $\hat{\beta}$ from this, and compute $e_t = y_t - x_t\beta$. This is the error, the amount the system is out of equilibrium (it is also the residual from the cointegrating regression).
3. Now estimate the short-run relationship $\Delta y_t = \Delta x_t \gamma + \rho e_{t-1}$.

Note: Similar to estimating a simple short-run model Δy on Δx except for lagged e term. Since the lagged e term is the amount the system was out of equilibrium ("error"), this is called an "error correction" model.

We can also do this more easily using the DHSY form of the error correction model

$$\Delta y_t = \Delta x_t \beta + \rho\{y_{t-1} - x_{t-1}\gamma\} + \epsilon_t \tag{33}$$

Note the term in braces is how much the system was out of equilibrium last period and the first term measures the short run relationship. This DHSY error correction model is a very sensible model whenever y and x are in an equilibrium relationship, even if they are stationary.

Note that ρ gives you the speed of adjustment, what proportion of the "error" is corrected each period. When interpreting, remember it is per period, so will be bigger with yearly data, smaller with monthly data, etc.

Why is simple short run model inferior? Throws away information. Only tells you how short term changes in x affect y, no information about the two being in long run equilibrium. (If they are not, then $\rho = 0$).

Can estimate the DHSY form by OLS.

Structural Stability

There are other forms of non-stationarity. One of the more interesting is a structural break, where the parameters change at $1 < t_0 < T$.

If have an idea about t_0, can test for via Chow test (just estimate two models, one for $1 - t_0$ and one for $t_0 + 1 - T$ and compare SSE's of these regressions with SSE for overall regression via usual F-test.

If do not know break point, can try this proceedure over all possible t_0 and see if one stands out.

Alternatively and better, can do *CUSUM SQUARE* test. Consider an estimate $\hat{\beta}_t$ which is based on observations $1, \ldots, t$. Then the forecast error for the next period, which is called the *recursive residual* is $y_{t+1} - x_{t+1}\hat{\beta}_t$.

The cusum squared test takes the sum of the squares of the recursive residuals (divided by t) and plots it against t. If no breaks should look like 45 degree line. Durbin has computed bands to see if this plot indicates a structural break - this plot and test is implemented in EVIEWS.