POLS 707 Paul Johnson <pauljohn@ku.edu>
Time-Series and Autocorrelation (2005)

1 Consider one unit over time.

Suppose you have the idea that X affects Y.

$$y_t = a + b \cdot x_t + u_t \quad (1)$$

By custom, we use subscript t for time points.

It is possible to have time related components in y_t, x_t, u_t, and, possibly also the coefficients would need a time subscript. So you could end up with some big horrible looking equation like

$$y_t = \rho_1 y_{t-1} + \ldots + \rho_p y_{t-p} + a + b_0 x_t + b_1 x_{t-1} + \ldots + b_m x_{t-m} + \theta_0 u_t + \theta_1 u_{t-1} \ldots + \theta_{t-q} u_{t-q} \quad (2)$$

Please remember: we are working on a study of one unit only. If you want to combine data from several units over time, you cross over into a different statistical field which is variously known as “panel data”, “cross sectional time series”, “repeated measurements”, or “longitudinal data”.

2 Overview

2.1 Focus on u_t: You get traditional Autocorrelation (same as “serially correlated errors”)

This is a simple extension of regression modeling. The error term from one time is influenced by the error term from the time before (or times before).

The model has 2 parts.

$$y_t = b_0 + b_1 x_t + u_t \quad (3)$$

and then a statement about the autoregressive error, such as:

$$u_t = \rho u_{t-1} + \varepsilon_t \quad (4)$$

or this

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t \quad (5)$$

and so forth.

There are fixes you can use if you have a good idea about the kind of autocorrelation that is present in the error term. If you understand “Weighted Least Squares,” you will understand this as a simple variant of it.

Please note. This is not the same as saying the X is correlated with its own previous values, or the Y is. Neither of these is necessarily a problem.

2.2 Focus on x_t: You Get Distributed Lag Models

Sometimes people get excited because they think that "lagged" values of X matter. So their specification might be

$$y_t = a + b_0 x_t + b_1 x_{t-1} + u_t \tag{6}$$

more generally,

$$y_t = a + \sum_{j=0}^{m} b_j x_{t-j} + u_t \tag{7}$$

These are often difficult to estimate (primarily because of multicollinearity).

To allow estimation, some clever transformations can be used. For example, impose a mathematical structure on the b's so as to reduce the number of parameters to be estimated. Supposing that all of the historical values of X play a role, but their coefficients are gradually diminishing, as in

$$y_t = a + b_0 x_t + \sum_{j=1}^{\infty} b_1^j x_{t-j} + u_t \tag{8}$$

this might make sense if $0 < b_1 < 1$ so that we think of lagged X's having less and less impact. Just so the sum doesn't conceal anything, you might write this out.

$$y_t = a + b_0 x_t + b_1^1 x_{t-1} + b_1^2 x_{t-2} + \ldots b_1^{\infty} x_{t-\infty} + u_t \tag{9}$$

In this model, the distributed lag part has only one coefficient, b_2 that needs to be estimated.

2.3 Distributed Lags with Autoregression

In Greene (Econometric Analysis, 4ed), a variant is considered in which there are lagged input variables and lagged dependent variables. He calls it "ARDL", short for Autoregressive distributed lag.

$$y_t = \sum_{k}^{p} \gamma_k y_{t-k} + \sum_{j=0}^{m} b_j x_{t-j} + u_t = \gamma_1 y_{t-1} + \ldots + \gamma_p y_{t-p} + b_0 x_t + \ldots + b_m x_{t-m} + u_t \tag{10}$$

2.4 ARIMA modeling.

AR-I-MA: "auto regressive - integrated - moving average" modeling. The idea here is to think of a time series y_t as a combination of inputs from its own past and various input variables. The original intention of ARIMA modeling was to isolate trends and predict y_t without using independent variables as input.

The AR part is the lagged y's on the right hand side. *Note that "autoregressive" in this context has a completely different meaning than in the previous section!* The MA part is the lagged unobserved error–thought of as inputs–on the right hand size. The "integrated" part is a confusing thing I don't want to distract you with it.

If I ignore the "integrated" part, then I just have an ARMA model, with p lagged y's and q lagged inputs (ϵ):

$$y_t = \rho_1 y_{t-1} + ... + \rho_p y_{t-p} + \epsilon_t + \tau_1 \epsilon_t + ... + \tau_q \epsilon_{t-q} \tag{11}$$

Sometimes people make the notation more fancy by using the lag operator notation,

$$y_{t-1} = L(y_t) \tag{12}$$

If you use that notation, then the big model above can be written:

$$y_t - \rho_1 L(y_t) + ... + \rho_p L^{\rho}(y_t) = \epsilon_t + \tau_1 L(\epsilon_t) + ... + \tau_q L^q(\epsilon_t) \tag{13}$$

which is the same as:

$$y_t(1 - \rho_1 L_t + \rho_p L^{\rho}) \;\; = \;\; \epsilon_t(1 + \tau_1 L + ... + \tau_q L^q) \tag{14}$$

and you should be able to see some additional representations.

This kind of thing was pioneered by Box and Jenkins and there are many adherents of it. In 2000, I made a pretty serious attempt in POLS909 to master this kind of model, and concluded it is very risky and unstable.

The error term here is just thought of as unmeasured inputs. If you have measures of variables you want to study, you have what is sometimes called an ARIMAX model:

$$y_t = \rho_1 y_{t-1} + ... + \rho_p y_{t-p} + \beta_0 x_t + \tau_0 \epsilon_t + \tau_1 \epsilon_t + ... + \tau_q \epsilon_{t-q} \tag{15}$$

If the x variable (or variables) are dummy variables representing policy effects, this is sometimes called “interrupted time series analysis” or “intervention analysis” or “state-space modeling”.

Many smart people I know, including Prof. Schrodt and Prof. Herron, seem to agree (for different reasons) that, although ARIMA modeling is widely practiced, you should be cautious about it. Prof. Schrodt once showed me some models indicating that OLS parameter estimates have better properties than ARIMA estimates.

The main problem with ARIMA is that the specification–the choice of p and q– is very subjective. Two educated people can follow the same principles and conclude that quite different models are called for.

3 Autocorrelation in the error term (more details)

This is the part of time series analysis that fits together well with an intermediate regression class.

3.1 Recall the OLS assumption that $E(e_t, e_{t-j}) = 0$, for any j. Error term for one observation is not dependent on the error term for other observations.

3.2 The Problem: e is influenced by its past values.

Suppose it is still true that $E(e_t) = 0$, but there is an “autocorrelation” problem in e_t.

In the stylish notation, the first-order autocorrelation model (often referred to as AR(1)):

$$AR(1): \;\; e_t = \rho e_{t-1} + v_t \tag{16}$$

Here ρ (latin "rho") is a coefficient, and v_t is a "new, nice, pleasant and ordinary error term," by which I mean it has a constant variance and it has no autocorrelation, $E(v_t, v_{t-j}) = 0$.

You should understand that the error term at a given time reflects a "weighted average" of past values of v_t.

$$e_t = \rho(\rho e_{t-2} + v_{t-1}) + v_t = \rho^2 e_{t-2} + \rho v_{t-1} + v_t \tag{17}$$

$$e_t = \rho^2(\rho e_{t-3} + v_2) + \rho v_{t-1} + v_t = \rho^3 e_{t-3} + \rho^2 v_{t-2} + \rho v_{t-1} + v_t \tag{18}$$

Repeat that a few times and you see that, as long as $-1 < \rho < 1$, the past values are "discounted".

To better understand the implications of this structure, calculate the variance of the error term.

$$Var(e_t) = E(e_t \cdot e_t) = E[(v_t + \rho v_{t-1} + \rho^2 v_{t-2} + \ldots)(v_t + \rho v_{t-1} + \rho^2 v_{t-2} \ldots)] \tag{19}$$

$$= E[v_t^2 + \rho^2 v_{t-1}^2 + \rho^4 v_{t-2}^2 + \cdots + \rho v_t v_{t-1} + \rho^3 v_t v_{t-2} \cdots] \tag{20}$$

$$= E(v_t^2) + \rho^2 E(v_{t-1}^2) + \rho^4 E(v_{t-2}^2) + \cdots + \rho E(v_t v_{t-1}) + \rho^3 E(v_t v_{t-2}) \cdots] \tag{21}$$

Recall, we assumed that v_t is a well behaved error term, so

$$E(v_t^2) = E(v_{t-j}^2) = \sigma_v^2 \tag{22}$$

and

$$E(v_t \cdot v_{t-s}) = 0 \, (s \neq 0) \tag{23}$$

so variance reduces to

$$Var(e_t) = E(e_t e_t) = \sigma_v^2 + \rho^2 \sigma_v^2 + \rho^4 \sigma_v^2 + \ldots = \frac{\sigma_v^2}{1-\rho^2} \tag{24}$$

Similarly, the covariance across times can be calculated

$$E(e_t e_{t-1}) = \rho \frac{\sigma_v^2}{1-\rho^2} = \rho Var(e_t e_t) = \rho \sigma_e^2 \tag{25}$$

$$E(e_t e_{t-s}) = \rho^s \frac{\sigma_v^2}{1-\rho^2} \tag{26}$$

Using that information, it is quite feasible to write out a "variance-covariance" matrix for the error terms, e_t

$$\frac{\sigma_v^2}{1-\rho^2} \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & & \\ \rho & 1 & \rho & \rho^2 & \rho^3 & \\ \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^3 & & & \ddots & & \rho^2 \\ & \rho^3 & \rho^2 & \rho & 1 & \rho \\ & & \rho^3 & \rho^2 & \rho & 1 \end{bmatrix} \tag{27}$$

You think your error term has more detailed time-dependence? No problem. Maybe you want:

AR(2):

$$AR(2): \; e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + v_t \tag{28}$$

We can have a higher order, if you want.

3.3 Regression consequences

If you estimate a regression in which the error term is AR(1) with OLS, you should know this:

1. OLS estimates of the b's are unbiased and consistent
2. OLS gives the wrong (biased) estimates of the standard errors of the b's. Thus the t-tests are bogus. The t-values are bigger than they should be, and you are likely to falesly reject the null hypothesis.
3. OLS is inefficient. There is an alternative estimation procedure (GLS) that gives estimates that are also unbiased and consistent, but also have lower variance.

3.4 GLS estimates

GLS is rather like WLS. You do some inspecting to guess what sort of AR you have, then you apply a correction. It is much easier to describe the GLS process using matrix algebra. There is a separate handout on the principles of GLS which demonstrates the full argument.

There are several procedures specifically aimed at estimating AR(1) models. The most famous is the Cochrane-Orcutt procedure, which first estimaes ρ from the data, then it plugs that estimated value of ρ to calculate estimates of the b's.

Step 1: Get an estimate of ρ. In Cochrane-Orcutt, the first step is to estimate an OLS model, take the residuals, calling them $\widehat{e}_t$, and then estimate ρ in this model:

$$\widehat{e}_t = \rho * \widehat{e}_{t-1} + u_t$$

where u_t is some pleasant error term we assume.

Step 2: Take the estimate of ρ from that regression, and then reweight the observations so they have constant variance and uncorrelated observations. You do this by a sneaky trick:

$$y_{t-1} = b_0 + b_1 X1_{t-1} + e_{t-1}$$

$$\rho y_{t-1} = \rho b_0 + \rho b_1 X1_{t-1} + \rho e_{t-1} \tag{29}$$

If you subtract equation 29 from 3, look what you get:

$$y_t - \rho y_{t-1} = b_0 - \rho b_0 + b_1(X1_t - \rho X1_{t-1}) + ... + e_t - \rho e_{t-1} \tag{30}$$

Holy cow! Look at the error term. It is equal to our nice friend v_t.

$$v_t = e_t - \rho e_{t-1}$$

Step 2 is implemented, then, by just calculating new variables y* and x* from the obvious equivalents in 30:

$$y_t^* = b_0(1-\rho) + b_1 X1_t^* + ... + v_t$$

where

$$y_t^* = y_t - \rho y_{t-1}$$

and

$$X1_t^* = X1_t - \rho X1_{t-1}$$

4 Testing for Autocorrelation (serial correlation)

Regression estimates usually include an estimate of the Durbin Watson statistic, which is a test for AR(1). Only for AR(1). AND it is not correct when there are "lagged y" values on the right hand side.

4.1 Interpretation

Interpretation of the DW is somewhat tricky.

General rule of thumb: DW should be "near 2" if you want to reject the possibility that serial correlation exists, which means you affirm the claim ρ=0.

How close to 2 does it have to be? That's the hard part. DW comes with 2 diagnostic limits, d_l and d_u.

If the null hypo is that $\rho = 0$, then a DW<$d_l$means that the null can be rejected. If the DW > d_u, then the null can be rejected as well. However, if d_l < DW < d_u then the test is inconclusive.

The indeterminacy is due to the possibility that autocorrelation in $X1_t$ may be causing the apparent autocorrelation in e_t.