Auto Correlation in Regression

Paul E. Johnson[1] [2]

[1]Department of Political Science

[2]Center for Research Methods and Data Analysis, University of Kansas

2012

Overview

Outline

1. Introduction
2. Time Series Analysis
3. Autocorrelated Error: The Love Story Called AR(1)
   - Define AR(1)
   - Testing for Autocorrelation
4. If Autocorrelation, Then What? GLS!
   - GLS (In General)
   - Specialized Versions of the GLS Algorithm
5. Topics for Further Study

Time Series Is Its Own Field of Statistics

- IF you want to study time series data, there is a separate series of courses you should take
  - Econometrics, dynamics, panel-data analysis, longitudinal regression analysis
  - Purpose of this lecture is to motivate main questions for those additional courses
- Notational changes
- Refer to "rows of data" by subscript t instead of subscript i
  - some lingering effect of errors at $t-1$, $t-2$, and so forth

Autocorrelation in a Nutshell

- Suppose rows and columns are time points. OLS regression assumes

$$Var(e) = E(e \cdot e'|X) = \begin{bmatrix} \sigma_e^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_e^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_e^2 & 0 & 0 \\ \dots & & \dots & \dots & 0 \\ 0 & 0 & 0 & 0 & \sigma_e^2 \end{bmatrix} \quad (1)$$

- Note 2 critical simplifications are used
  1. All non-diagonal elements are 0
  2. All diagonal elements are equal to each other

Autocorrelation in a Nutshell (2)

- Heteroskedasticity throws away simplification #2: allows differing σ_i^2 values,

$$Var(e) = E[e \cdot e'|X] = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 & 0 \\ 0 & 0 & \ddots & \cdots & 0 \\ 0 & 0 & 0 & \sigma_{N-1}^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_N^2 \end{bmatrix}$$

but heteroskedasticity still leaves 0's on all off-diagonal elements

Autocorrelation in a Nutshell (3)

- Autocorrelated errors: there are correlations "across time"

$$Var(e) = E[e \cdot e'|X] = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \cdots & \sigma_{2N} \\ \sigma_{31} & \sigma_{32} & \ddots & \cdots & \sigma_{3N} \\ \vdots & \ddots & \ddots & \sigma_{N-1}^2 & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & & \sigma_N^2 \end{bmatrix}$$

- $\sigma_{21} = Cov(e_1, e_2)$, the covariance of the random error across observations on a unit
- $Var(e)$ is symmetric, because $\sigma_{21} = \sigma_{12}$
- But otherwise, it can be arbitrarily complicated

Consequences of Ignoring Autocorrelated Errors

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 falsely 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.

Spatial Autocorrelation

- Autocorrelation is not just for “time series” anymore.
- Work on spatial autocorrelation has intensified.
  - Refer to data points in a grid or a map
  - Hypothesize that disturbances at one cell may “disperse” themselves across other cells.

Outline

Consider one unit over time.

- One time series, x_t affects y_t, for $t = 1, 2, 3, \ldots T$
- The simplest model, the one with no complications, is fit with OLS

$$y_t = a + b \cdot x_t + e_t \tag{2}$$

- That's just one of the many possibilities, of course. Consider big horrible looking equation like

Time Series as a Field of Study

$$y_t = \rho_1 y_{t-1} + \ldots + \rho_p y_{t-p} + a + b_0 x_t + \ldots + b_m x_{t-m} + \theta_0 e_t + \theta_1 e_{t-1} \ldots +$$

- Lagged dependent variables, y_{t-1}, y_{t-2}, ...: "autoregression models"
- Lagged exogenous variables, x_{t-1}, x_{t-2}, ...: "distributed lag models"
- Lagged error terms, e_{t-1}, e_{t-2}, ...: "autocorrelated error models" and "Autoregressive Conditional Heteroskedasticity"

Cross Sectional Time Series

- Original Time Series studies focused on 1 data series, in isolation (psychologists would say “idiographic”)
- Collect several time series. What do you have? Names for same kind of problem:
  - Pooled Time Series
  - Cross Sectional Time Series
  - Panel Data Analysis
  - Repeated Measures
  - Longitudinal Analysis

Outline

Focus on e_t: Autocorrelation = "serially correlated errors"

- Suppose we don't have lagged y's or x's on the right hand side.
- Do allow lagged errors.

$$y_t = b_0 + b_1 x_t + e_t \tag{3}$$

- We consider only the distortion caused by lagged error terms, which are often called "disturbances".
- This is not the same as saying the X is correlated with its own previous values, or that Y is. This only concerns the lingering effects of past "shocks".

Add an Equation for the Autoregressive Error

- AR(1), auto regression of order 1:

$$e_t = \rho e_{t-1} + u_t \tag{4}$$

- Error at time t includes
  - a portion ρ of the previous (unmeasured) error, and
  - a new random error, which "nice"
    1. $E[u_t] = 0$ (unbiased)
    2. $E[u_t^2] = \sigma_u^2$ (homoskedastic) and not autocorrelated $E[u_t, u_s] = 0$
- AR(2) would introduce 2 lagged errors, as in

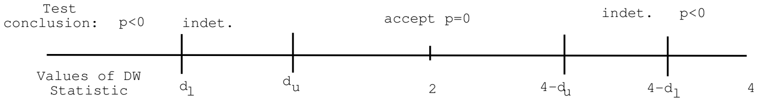
$$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + u_t \tag{5}$$

The Durbin Watson Test for AR(1)

- This is only for AR(1).
- AND it is not correct when there are "lagged y" values on the right hand side.
- General rule of thumb: DW should be "near 2" in order to reject the possibility that serial correlation exists, which means you affirm the claim ρ=0.

DW Interpretation

- The theory is $y_t = b_0 + b_1 x_t + e_t$ and $e_t = \rho e_{t-1} + u_t$
- If $\rho = 0$, then this is just the "regular old OLS model". So the null hypothesis is $\rho = 0$.
- But the DW statistic equals 2 if the null is true.
- How close to 2 does it have to be? DW comes with 2 diagnostic limits, d_l and d_u.



DW far from 2 ($< d_l$ or $> 4d_o$) means that the null can be rejected.
DW very close to 2 ($d_U < DW < 4 - d_U$), null accepted
If $d_l <$ DW $< d_u$ (or $4 - d_u < DW < 4 - d_l$), then the test is inconclusive.

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

Outline

FGLS: Feasible GLS

A Two Step Estimation Procedure

1. Theorize a correlation structure and "work out" an estimate of the error term's variance/covariance matrix.
2. Use Generalized Least Squares to calculate estimates that best fit.
3. Repeat the procedure. The GLS fit -> estimates of the error covariances -> new GLS estimates.

Generalized Least Squares

- Weighted Least Squares weights each error $(y_t - \hat{y}_t)$ by w_t

$$S(\hat{b}) = \sum_{t=1}^{T} W_t (y_t - \hat{y}_t)^2 = \sum_{t=1}^{T} (y_t - \hat{y}_t) w_t^2 (y_t - \hat{y}_t) \qquad (6)$$

- Think of that like this: Multiply all "mix and match combinations":

$$\begin{array}{cc} & \begin{array}{ccccc} (y_1 - \hat{y}_1)(& y_2 - \hat{y}_2)(& y_3 - \hat{y}_3) & \ldots & (y_T - \hat{y}_T) \end{array} \\ \begin{array}{c} (y_1 - \hat{y}_1) \\ (y_2 - \hat{y}_2) \\ (y_3 - \hat{y}_3) \\ \vdots \\ (y_T - \hat{y}_T) \end{array} & \begin{bmatrix} w_1^2 & 0 & 0 & 0 & 0 \\ 0 & w_2^2 & 0 & 0 & 0 \\ 0 & 0 & \ddots & \cdots & 0 \\ \vdots & 0 & 0 & w_{T-1}^2 & 0 \\ 0 & 0 & 0 & 0 & w_T^2 \end{bmatrix} \end{array} \qquad (7)$$

- A sum with only T terms–the 0's in the weight matrix

Generalized Least Squares (cont.)

- The Weight matrix for autocorrelated errors does not have all of those 0's
- So the carry out the "multiply and add" exercise:

$$\begin{matrix} & (y_1-\hat{y}_1)(& y_2-\hat{y}_2)(& y_3-\hat{y}_3) & \dots & (y_T-\hat{y}_T) \\ \begin{matrix}(y_1-\hat{y}_1)\\(y_2-\hat{y}_2)\\(y_3-\hat{y}_3)\\ \vdots \\(y_T-\hat{y}_T)\end{matrix} & \multicolumn{5}{c}{\begin{bmatrix} w_1^2 & w_{12} & w_{13} & \dots & w_{1T} \\ w_{21} & w_2^2 & w_{23} & \dots & w_{2T} \\ w_{31} & w_{32} & \ddots & \dots & w_{3T} \\ \vdots & \vdots & \ddots & w_{T-1}^2 & \vdots \\ w_{T1} & w_{T2} & w_{T3} & \dots & w_T^2 \end{bmatrix}} \end{matrix} \quad (8)$$

Leads to a sum with $T \times T$ terms.

Generalized Least Squares (cont.)

- Horrible, yes? The GLS Sum of Squares is a gigantic tangle if you write it all out

$$\text{row 1: } (y_1 - \hat{y}_1)w_1^2(y_1 - \hat{y}_1) + (y_1 - \hat{y}_1)w_{12}(y_2 - \hat{y}_2) + (y_1 - \hat{y}_1)w_{13}(y_3$$

$$\text{row 2: } (y_2 - \hat{y}_2)w_{21}(y_1 - \hat{y}_1) + (y_2 - \hat{y}_2)w_2^2(y_2 - \hat{y}_2) + (y_2 - \hat{y}_2)w_{23}(y$$

$$\text{row T: } (y_T - \hat{y}_T)w_{T1}(y_1 - \hat{y}_1) + (y_T - \hat{y}_T)w_{T2}(y_2 - \hat{y}_2) + \ldots + (y_T - \hat{y}_T)w_T^2($$

$$= \sum_{t=1}^{T}\sum_{s=1}^{T}(y_t - \hat{y}_t)w_{ts}(y_s - \hat{y}_s) \qquad (10)$$

- Use matrix algebra, this is a lot easier to write down.

$$(y - \hat{y})' \, W \, (y - \hat{y}) \qquad (11)$$

Generalized Least Squares (cont.)

- A big swirl of computations is involved in deriving the best fitting values of the coefficients.
- The solution looks like

$$\hat{b} = (X'WX)^{-1}X'Wy \tag{12}$$

And Then Iterate (Maybe)

- An estimate of $\hat{b}$ begets a new estimate of W
- Then a new estimate of W begets a new estimate of $\hat{b}$.
- Repeat until estimates of $\hat{b}$ converge.

Cochrane-Orcutt for AR(1)

The most famous special purpose GLS estimator is the Cochrane-Orcutt procedure for AR(1).

Step 1: Get an estimate of ρ. To do so,

a) estimate an OLS regression of y_i on x_i

b) call the residuals $\widehat{e}_t$

c) estimate ρ in this model:

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

u_t is a "nice" error term.

Cochrane-Orcutt (Step 2)

Step 2: Use $\hat{\rho}$ to create new weighted observed variables

$$y_t^* = y_t - \hat{\rho} y_{t-1} \text{ and } x_t^* = x_t - \hat{\rho} x_{t-1} \tag{13}$$

Regress: y_t^* on x_t^*. That provides an estimate of $\hat{b}_1$.
And it provides a new set of residuals, so we can repeat step 1, then step 2, and so forth.

Why Calculate $y_t^* = y_t - \rho y_{t-1}$and $x_t^* = x_t - \hat{\rho} x_{t-1}$?

That's why Cochrane and Orcutt became famous.

- Restate the assumptions we have already made:

$$y_t = b_0 + b_1 x_t + e_t$$

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

- Multiply through by the unknown constant ρ:

$$\rho y_{t-1} = \rho b_0 + \rho b_1 x_{t-1} + \rho e_{t-1} \tag{14}$$

- Subtract the third from the first:

$$y_t - \rho y_{t-1} = b_0 - \rho b_0 + b_1 (x_t - \rho x_{t-1}) + e_{t-1} - \rho e_{t-1} \tag{15}$$

Why y_t^* and x_t^*?

Restate:

$$y_t - \rho y_{t-1} = b_0 - \rho b_0 + b_1(x_t - \rho x_{t-1}) + e_{t-1} - \rho e_{t-1}$$

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

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

Step 2 is implemented, then, by calculating new variables y* and x* from the obvious equivalents in (15):

$$y_t^* = b_0(1-\rho) + b_1 x_t^* + ... + u_t$$

Outline

Focus on x_t: Distributed Lag Models

- Sometimes people get excited because they think that "lagged" values of x_t matter.

$$y_t = a + b_0 x_t + b_1 x_{t-1} + u_t \tag{16}$$

- Its possible somebody wants to add a whole slew of lagged x's.

$$y_t = a + \sum_{j=0}^{m} b_j x_{t-j} + u_t \tag{17}$$

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

- Clever choice of theory can simplify and make estimation possible (Almon lags, for example).

ARIMA modeling.

- AR-I-MA: "auto regressive - integrated - moving average" modeling.
- A time series y_t is 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.

ARMA means No i "Integration Component".

- ARMA model, with p lagged y's and q lagged errors (e_t):

$$y_t = \rho_1 y_{t-1} + ... + \rho_p y_{t-p} + e_t + \tau_1 e_{t-1} + ... + \tau_q e_{t-q} \qquad (18)$$

- How many non-zero ρ_j coefficients are need? How many τ_j are needed? That's the magical, mysterious field of ARIMA modeling for you. Judgment, graphs, tests.
- *Note that "autoregressive" in this context has a different meaning!*
  - The AR part concerns the lagged y's.
  - The MA part is the lagged unobserved error.

More ARIMA notation

- Lag operator notation,

$$\begin{aligned} y_{t-1} &= L(y_t) \quad (19)\\ y_{t-2} &= L^2(y_t) \\ y_{t-3} &= L^3(y_t) \quad (20) \end{aligned}$$

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

$$y_t - \rho_1 L(y_t) + ... + \rho_p L^{\rho}(y_t) = e_t + \tau_1 L(e_t) + ... + \tau_q L^q(e_t) \quad (21)$$

- Which is the same as:

$$y_t(1 - \rho_1 L_t + \rho_p L^{\rho}) \quad = \quad \epsilon_t(1 + \tau_1 L + ... + \tau_q L^q) \quad (22)$$

- Observed input variables x_t can be introduced. This is called an ARIMAX model:

$$y_t = \rho_1 y_{t-1} + ... + \rho_p y_{t-p} + \beta_0 x_t + \tau_0 \epsilon_t + \tau_1 \epsilon_t + ... + \tau_q \epsilon_{t-q} \quad (23)$$