

POL 606 Time Series Analysis
Week 6
Time Series Regression Models
March 3rd and 5th, 2008

Box-Jenkins techniques are generally separated from Time Series regression analysis.

The biggest differences between them is the treatment of moving average parameters and non-stationarity.

If we believe that a moving average parameter is present in the error term we should use Box-Jenkins.

If not, more traditional regression techniques can be used.

Why?

Because moving averages involve using lags of the error term and for standard regression techniques, the error term is something that is estimated. It is impossible, in one step, to estimate the error term and use it to estimate other parameters.

Traditional Time Series Analysis

Researchers using time series data in political science typically have used many of the same regression techniques as are employed to analyze cross-sectional data.

We can call this the *Econometric Approach*

Mostly single equation models such as the following:

$$Y_t = \beta_0 + \sum \beta_{1-k} X_{1-k,t-i} + \varepsilon_t$$

Where: Y_t is the dependent variable at time t .

$X_{t-i} = 1$ to k independent variables at time $t-i$.

$\beta_0 =$ constant.

$\beta_{1-k} =$ parameters associated with variables X_{1-k}

$\varepsilon_t =$ stochastic error term $\sim N(0, \sigma^2)$

For a model such as above the possible non-stationarity of the variables is ignored and OLS is employed to estimate the parameters.

The effects of the X's may be specified to occur simultaneously (i.e. at time t , or with a lag i).

Inferences about the parameters are made using t -statistics.

When doing diagnostics on such regression models, particular attention is given to the possibility that the errors are correlated (i.e. $\text{cov}(\varepsilon_t, \varepsilon_{t-1}) \neq 0$).

Review Gauss-Markov Assumptions

Assumptions of the classical OLS regression model.

Relationship between the X's and Y are linear and additive – constant slope,

The error term ε is normally distributed with:

Mean = 0

Constant variance, $E(\varepsilon_t, \varepsilon_{t-1}) = 0$ ← homoskedasticity assumption.

Error terms are not autocorrelated – that is, the expected value of covariance = 0:

$$\text{cov}(\varepsilon_t, \varepsilon_{t-1}) \neq 0$$

X is non-stochastic and is measured perfectly.

Assume a properly specified model.

No Multicollinearity – if 2 X variables are highly correlated the standard errors get bigger and R^2 becomes inflated.

Xs are uncorrelated to the error term.

These are our assumptions for any kind of regression.

But for time series, some of these assumptions become more problematic.

Especially, that the error terms are uncorrelated and that the Xs are uncorrelated with the error term.

If the Gauss-Markov assumptions hold, then OLS estimates will have desirable properties and be:

1. Unbiased – taking repeated samples gives a mean of samples equal to the population mean.

2. Efficient – variance of the estimates is smallest possible.

Why do we want efficiency?

Because we only get one sample so with minimum variance we will be closest to the population value.

3. Consistency – as sample size increases our estimate converges on the population value.

When all assumptions are met, OLS estimator is Best Linear Unbiased Estimate.

If all of these assumptions hold for our time series data, we can simply use OLS regression for them.

Why worry about autocorrelations?

Consequences:

Does not bias estimates of regression parameters, but does bias estimates of those parameters' standard errors *downwards*.

So: $t = \frac{\hat{\beta}}{SE_{\hat{\beta}}}$ gets larger and we are more likely to make a Type I error – falsely reject the null hypothesis.

What should we do?

Look at residuals and then try to model the error process.

We might find that there is an autoregressive process in the residuals.

We could diagnose this in an ACF and PACF of the residuals.

We can also find evidence of autocorrelation in the Durbin-Watson test statistic.

This is a test for 1st order autocorrelation.

$$d = \frac{\sum (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum \hat{\varepsilon}_t^2} = 2(1 - \rho)$$

where: $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$ $\rho = \text{Rho}$.

That is, ρ is the autoregressive parameter for the error term.

When $\rho=0$, $d=2$ and no autocorrelation exists.

If there is perfect positive autocorrelation $\rho=1$ and $d=0$.

If there is perfect negative autocorrelation $\rho = -1$ and $d=4$.

Which makes one ask, if Durbin was so smart why can't 0 be neutral?

Values close to 0 or 4 indicate problems of 1st order autocorrelation.

You can use these numbers as a guide or find a table of critical values the test statistic in a Durbin's table.

Test for Positive 1st order autocorrelation:

If less than lower bound \rightarrow reject null hypothesis of no 1st order autocorrelation.

If above \rightarrow fail to reject null.

If in between \rightarrow inconclusive.

Test for negative 1st order autocorrelation:

if $4 - d_L < d \rightarrow$ reject the null.

if $d < 4 - d_U \rightarrow$ fail to reject the null.

if $4 - d_U \leq d \leq 4 - d_L \rightarrow$ inconclusive.

Durbin-Watson statistics is not valid if the model includes a lagged endogenous variable:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 X_t + \varepsilon_t$$

For this we would need to use Durbin's H.

So one way to think about autocorrelation is to think about ρ in the following:

$$\varepsilon_t = \rho \varepsilon_{t-1} + v_t$$

Where ρ captures the relationship between temporally adjacent errors, and v_t is a well behaved error term, $\sim N(0, \sigma^2)$.

This relationship between the errors is treated as a nuisance to be corrected.

The alternative possibility – that the correlation among the residuals comes from a misspecified model – is not considered.

Correction employed is usually Generalized Least Squares.

Generalized Least Squares (GLS)

Involves multiplying both sides of the model by the “quasi-differencing” operator $(1 - \rho L)$.

Suppose we perform a Durbin-Watson test on:

$$y_t = a + bX_t + \varepsilon_t \quad \text{Call this Equation 1.}$$

and conclude that the error process is 1st order autoregressive.

What do we do to fix this?

Assume first that we know ρ .

In GLS we do some kind of transformation.

Here, lag equation back 1 period and multiply by ρ .

Remember that $-1 \leq \rho \leq 1$.

$$\begin{aligned} \rho y_{t-1} &= \rho(a + bX_{t-1} + \varepsilon_{t-1}) \\ \rho y_{t-1} &= \rho a + \rho bX_{t-1} + \rho \varepsilon_{t-1} \end{aligned} \quad \text{Call this Equation 2.}$$

Equation 3 = Equation 2 – Equation 1.

$$y_t - \rho y_{t-1} = (a - \rho a) + (bX_t - \rho bX_{t-1}) + (\varepsilon_t - \rho \varepsilon_{t-1})$$

or:

$$y_t - \rho y_{t-1} = a(1 - \rho) + b(X_t - \rho X_{t-1}) + (\varepsilon_t - \rho \varepsilon_{t-1})$$

Call $(\varepsilon_t - \rho \varepsilon_{t-1})$ v_t which is now a well behaved error term.

Since we know what ρ , X , and Y are we can transform the variables and calculate a and b .

This is called the “**Cochrane Orcutt Transformation.**”

The problem is that we don’t usually know ρ .

So must estimate it.

This is called: *Pseudo GLS* or *Feasible GLS*.

Start with the original equation.

Save residuals.

Regress the residuals on itself lagged one period:

$$\hat{\varepsilon}_t = \hat{\rho}\hat{\varepsilon}_{t-1} + v_t$$

This gives an estimate of ρ .

We can transform Y and X and go back to the beginning.

The value of $\hat{\rho}$ will quickly converge on the proper value of ρ .

There are other Pseudo-GLS procedures.

1. Hildreth-Lu procedure is a “Grid Search” procedure.
Tries every possible value of ρ and chooses the value that minimizes the RSS.

2. Prace-Winston

3. Beach Mackinon.

We could assume that $\rho=1.0$ by first differencing the residuals. This is pretty simple but won't work if ρ is far from 1.

Notes: transformation causes us to lose our first case.

also, R^2 of the GLS procedure is based on transformed data – not original data.

1. do PGLS and get \hat{a} and \hat{b} .

2. $\hat{y}_t = \hat{a} + \hat{b}X_t$ using original data and PGLS \hat{a} and \hat{b} .

3. Get $R^2 = 1 - \frac{\sum (y_t - \hat{y}_t)^2}{\sum (y_t - \bar{y})^2}$

There is no focus in GLS on moving average effects.

for this we would do diagnostics suggested by Box-Jenkins modeling such as ACF and PACF.

If we found the error process was not AR1 but was an MA process we should go back and use Box-Jenkins, estimating the noise model for the errors.

Note: In general, finding a violation of the autocorrelation assumption means that we have misspecified – maybe there are other ways to deal with the problem.

We don't always have to respecify the noise model.

Political scientists using GLS usually don't recognize that they have, in effect, respecified their original model in autoregressive distributed lag form and imposed a common-factor restriction $(1 - \rho L)$.

This means that we are saying that the effects of all of the independent variables decay at the same rate.

Hendry (1995) emphasizes that this restriction should be justified empirically rather than assumed.

Most time series analyses in political science haven't done this.

The major point: autocorrelated residuals do **not** necessarily imply autocorrelated errors – so this approach is subject to misspecification.

Distributed Lag Model

$$y_t = a + b_1 X_t + b_2 X_{t-1} + b_3 X_{t-2} + \varepsilon_t$$

Effect of X on Y is spread over 3 periods or some number of finite periods.

Total effect of X = $b_1 + b_2 + b_3$

or: $b = \sum_{i=1}^k b_i$

$\frac{b_i}{\sum_{i=1}^k b_i}$ gives us the proportion of the affect appearing at time i .

Problems with this model:

1. With several X's and several lag we are estimating many parameters. We lose degrees of freedom and patience!
2. Problem of multicollinearity.

This will inflate standard errors.

We can test for multicollinearity using Klein test, Farran-Glaube Test or others.

Koyck Lag Scheme

Regression coefficients with lag weights to show the exponential decay.

$$b_k = b_0 \lambda^k \quad \text{where } 0 < \lambda < 1 \quad \lambda = \text{lambd}$$

The larger the λ , the slower the rate of decay.

$$\text{Equation 1: } y_t = a + b_0 \lambda^0 X_t + b_0 \lambda^1 X_{t-1} + b_0 \lambda^2 X_{t-2} + \dots + \varepsilon_t$$

There are an infinite number of lags on X on the right-hand-side.

How do we deal with this?

Koyck Transformation which is similar to the Cochrane-Orcutt transformation.

Take Equation 1 and lag back one period and multiply by λ .

Gives Equation 2:

$$\lambda y_{t-1} = \lambda a + b_0 \lambda X_{t-1} + b_0 \lambda^2 X_{t-2} + b_0 \lambda^3 X_{t-3} + \dots + \lambda \varepsilon_{t-1}$$

Equation 1 – Equation 2:

$$y_t - \lambda y_{t-1} = a - \lambda a + b_0 \lambda^0 X_t + \varepsilon_t - \lambda \varepsilon_{t-1}$$

Which can be expressed at:

$$y_t = a(1 - \lambda) + \lambda y_{t-1} + b_0 \lambda X_t + \varepsilon_t - \lambda \varepsilon_{t-1}$$

Now we only need to estimate a , λ , and b_0 .

Lagged Endogenous (Dependent) Variables

If serial correlation exists, OLS gives biased but consistent estimates – as sample grows, the value of the parameter will converge on its natural value.

Why?

$$y_t = a + b_1 y_{t-1} + \varepsilon_t \quad (1)$$

where, $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$ (2)

Note that y appears on both sides of equation (1) which becomes problematic.

Both ρ and b are bounded between -1 and +1.

ε_t and v_t are random white noise:

$$E(\varepsilon_t) = E(v_t)$$

$$E[\varepsilon_{t-\tau}v_t] = E[\varepsilon_tv_{t-\tau}] = 0 \text{ for all } \tau > 1.$$

Lagging equation (1) and multiplying both sides by ρ gives:

$$\rho y_{t-1} = b_1\rho y_{t-2} + \rho\varepsilon_{t-1} \quad (3)$$

Collecting terms gives:

$$\rho y_{t-1} - b_1\rho y_{t-2} = \rho\varepsilon_{t-1} \quad (4)$$

Now express equation (1) in terms of equations (4) and (2):

First, put (2) into (1):

$$y_t = a + b_1y_{t-1} + \rho\varepsilon_{t-1} + v_t \quad (5)$$

Then, put (4) into (5):

$$y_t = a + b_1y_{t-1} + \rho y_{t-1} - b_1\rho y_{t-2} + v_t \quad (6)$$

Which is equivalent to:

$$y_t = a + (b_1 + \rho)y_{t-1} - b_1\rho y_{t-2} + v_t \quad (7)$$

Compare (7) to (1).

(1) is what we think we are estimating, but (7) is what we are actually estimating.

That is, the coefficient on y_{t-1} includes ρ , not just b .

The variable y_{t-2} has been left out of equation (1).

Therefore, using OLS to estimate (1) will give us omitted variable bias and a bias of ρ on the coefficient of the lagged endogenous variable.

If $\rho=0$, there is no serial correlation and b will not be biased.

Given that we are talking about time series data, it is very unlikely that the residuals are i.i.d. and that no serial correlation exists.

Consequences of using OLS to estimate this model:

The variance will be incorrect – inefficiency.

The coefficients will be biased.

Residuals cannot be used to correct the problems because they are inherently biased – due to being the result of a mis-estimated equation.

Cannot use Durbin-Watson statistic to diagnose the problem because the coefficients are mis-estimated.

So, especially when we have a lagged dependent variable we should check for autocorrelation.

Three tests for serial correlation with lagged endogenous variable.

Durbin's H

Durbin-Watson statistic is no longer valid.

$$h = \hat{\rho} \sqrt{\frac{n}{1 - n \text{var}(\hat{b})}}$$

Where n is the sample size, $\hat{\rho} = 1 - \frac{d}{2}$ using d from a Durbin-Watson statistic, and the $\text{var}(\hat{b})$ is the variance on the coefficient on the lagged endogenous variable.

In a large sample, Durbin's H is distributed as a standard normal so that at the .05 level we can reject the null hypothesis when absolute H is greater than 1.96.

2 potential problems:

It is only a test for 1st order autocorrelation.

It won't work if $1 - n \text{var}(\hat{b}) < 0$.

If so, try M-test.

M-test

1. Do an OLS regression of equation of interest.
 2. Save residuals, $\hat{\varepsilon}_t$.
 3. Do auxiliary regression of $\hat{\varepsilon}_t$ on all the independent variables in the original regression: y_{t-1}, x_t as well as $\hat{\varepsilon}_{t-1}$.
- Look at the t-value for the coefficient on the lagged error terms $\hat{\varepsilon}_{t-1}$.

If statistically significant reject the null hypothesis of no auto-correlation.

Breusch-Godfrey Test – a.k.a. LM Test

1. Do OLS.
2. Save residuals, $\hat{\varepsilon}_t$.
3. Regress $\hat{\varepsilon}_t$ on $y_{t-1}, x_t, \hat{\varepsilon}_{t-1}, \hat{\varepsilon}_{t-2}, \dots, \hat{\varepsilon}_{t-k}$.
4. Calculate $N \cdot R^2$

Where N is the number of observations.

This is distributed as a χ^2 with degrees of freedom = k..

So, if we conclude that we have auto correlated errors and we want to use a lagged endogenous variable, what do we do?

We cannot directly do pseudo GLS.

We use Instrumental Variable Estimation.

Take:

$$y_t = a + b_1 y_{t-1} + b_2 X_t + \varepsilon_t$$

We want to fix the correlation between y_{t-1} and ε_t .

Find some variable that correlates highly with y_{t-1} but not ε_t .

This is much easier in time series analysis than in cross-sectional analysis since we can use X_{t-1} .

This will give some multicollinearity but will give an unbiased estimate of ρ .

After that, we can do the Cochrane Orcutt Transformation.

See Hamilton pp.215-17 and p.227. Also Gujarati pp. 604-5.