Lecture 14: LAGS AND DYNAMICS

Modern time-series analysis does not treat autocorrelation as necessarily a property of the "errors", but as an indication of potentially interesting economic dynamics.

Definition: The lag operator L is defined such that $LX_t = X_{t-1}$and in general, $L^S X_t = X_{t-S}$.

Definition: $D(L)$ is a polynomial of order S in L such that

$$D(L)X_t = \delta_0 X_t + \delta_1 LX_t + \delta_2 L^2 X_t + ... + \delta_S L^S X_t$$
$$= \delta_0 X_t + \delta_1 X_{t-1} + \delta_2 X_{t-2} + ... + \delta_S X_{t-S}.$$

The above polynomial is of order S. It can be of infinite order:

$$D(L)X_t = \sum_{i=0}^{\infty} \delta_i L^i X_{t-i}.$$

Suppose $y_t = \alpha + D(L)X_t + \varepsilon_t$, where $D(L)$ is of order S.

Interpretation: If X_t is fixed at $\bar{x}$, then this implies that Ey_t is fixed at $\bar{y} = \alpha + D(1)\bar{x}$ where

$$D(1) = \sum_{i=0}^{S} \delta_i.$$

Consider a change in x to a new constant level. This implies a change in $\bar{y}$, but this occurs slowly.

The effect at zero lag is δ_0. δ_0 is referred to as the "impact multiplier", and it represents the immediate effect of a change in $\bar{x}$. At lag 1, the effect is δ_0, etc.

The total effect is given by $\sum_{i=0}^{S} \delta_i$ which is equal to $D(1)$.

Note that $\delta_i / \sum_{i=0}^{S} \delta_i$ gives the proportion of total impact occurring at lag i.

Definition: The mean lag is given by

$$\frac{\sum \delta_i i}{\sum \delta_i} = \frac{D'(1)}{D(1)}$$

where $D'(1)$ is the derivative of $D(L)$ with respect to L evaluated at $L = 1$.

We usually want to restict the δ_i and not estimate $(S+1)$ free coefficients if S is large. To this end, we generally put some kind of pattern on the lags.

Example: Adaptive expectations

Suppose $y_t = \alpha + \beta X^*_{t+1} + \varepsilon_t$ where X^*_{t+1} is the expected X in period $t+1$.

Suppose

$X^*_{t+1} - X^*_t = (1-\lambda)(X_t - X^*_t)$ or

$X^*_{t+1} = (1-\lambda)X_t + \lambda X^*_t$.

Interpretation?

Iterating gives

$X^*_{t+1} = (1-\lambda)(X_t + \lambda X_{t-1} + \lambda^2 X_{t-2} + ...)$.

Substituting in the expression for y_t yields

$$y_t = \alpha + \beta(1-\lambda)(X_t + \lambda X_{t-1} + ...) + \varepsilon_t.$$

This is $D(L)$ with $S = \infty$ and $\delta_i = \beta(1-\lambda)\lambda^i$.

Use the trick by Koyck.

$$y_t = \alpha + \beta(1-\lambda)(X_t + \lambda X_{t-1} + ...) + \varepsilon_t$$

$$\lambda y_{t-1} = \lambda\alpha + \lambda\beta(1-\lambda)(X_{t-1} + \lambda X_{t-2} + ...) + \lambda\varepsilon_{t-1}$$

Subtracting λy_{t-1} from y_t gives

$$y_t = \alpha(1-\lambda) + \beta(1-\lambda)X_t + \lambda y_{t-1} + \varepsilon_t - \lambda\varepsilon_{t-1}.$$

The above equation is not quite in the standard framework, because $u_t = \varepsilon_t - \lambda\varepsilon_{t-1}$ is correlated with y_{t-1}. Note that a lagged dependent variable with autocorrelation generally results in the LS estimators being inconsistent. (*Why*?)

When are adaptive expectations rational?

When X is generated by $X_t = X_{t-1} + \eta_t - \lambda\eta_{t-1}$ where η_t are independent and identically distributed i.e., X_t is IMA(1). (*Prove this by substitution*.)

Other distributed lags:

1. *Polynomial Distributed Lags* (*Almon*)

In this case, coefficients (i.e., $\delta i's$) are given by a polynomial in the lag length. For example,

$\delta_i + a_0 + a_0 i + a_2 i^2$.

This is used for finite-length lag distribution. This imposes linear restrictions.

2. *Smoothness Priors* (*Shiller*)

The coefficients are given by

$\delta_i = a_0 a_0 i + a_2 i^2 + \eta_i$

3. *Rational Distributed Lags* (*Jorgenson*)

In this case, $D(L) = B(L)/A(L)$ where $D(L)$ is of infinite order (with restrictions), and $B(L)$ and $A(L)$ are low order polynomials.

Consider the model $y_t = \alpha + D(L)X_t + \varepsilon_t$.

Multiplying this model by $A(L)$ yields

$A(L)y_t = \alpha^* + B(L)X_t + v_t$.

Note that stability requires that the roots of $A(L)$ be greater than 1 in absolute value.

Example: Suppose $A(L) = 1 + \alpha L$.

$A(L) = 0 \Rightarrow L = -1/\alpha$ which is the root of this polynomial.

Stabiilty implies that $1/\alpha > 1$, i.e., $\alpha < 1$.

Since

$$y_t = \alpha + D(L)X_t + \varepsilon_t = \alpha + (B(L)/A(L))X_t + \varepsilon_t,$$

$\alpha = \alpha^*/A(L)$ and $\varepsilon_t = v_t/A(L)$.

Definition: The form of the model with no y's on the regressor side is referred to as the "transfer function" form.

The above discussion shows the correspondence between models with lagged dependent variables and distributed lags on independent variables.

DYNAMICS AND AUTOCORRELATION

Consider the model $y_t = \alpha + \beta X_t + u_t$ where $u_t = \rho u_{t-1} + \varepsilon_t (AR(1))$.

Transforming the model by subtracting λy_{t-1} from y_t yields

$$y_t = \alpha(1-\rho) + \rho y_{t-1} + \beta X_t - \beta\rho X_{t-1} + \varepsilon_t$$

$$= \gamma_0 + \gamma_1 y_{t-1} + \gamma_2 X_t + \gamma_3 X_{t-1} + \varepsilon_t.$$

This is a dynamic linear regression with independent and identically distributed errors.

However, note the nonlinear restriction $\gamma_0\gamma_2 + \gamma_3 = 0$.

Thus simple regression with $AR(1)$ errors is a restricted version of the dynamic model. This nonlinear restriction can be tested.

The Durbin-Watson statistic can indicate dynamic misspecification (apparent autocorrelation). Of course, the D-W is inappropriate in the dynamic model - the h-statistic should be used to detect further autocorrelation.

ESTIMATION OF DYNAMIC MODELS

Consider the model $y = Z\beta + \varepsilon$ where Z includes lagged y and X.

The LS esimator in this case is
$\hat{\beta} = (Z'Z)^{-1}Z'y = \beta + (Z'Z)^{-1}Z'\varepsilon$.

Assume that plim$(Z'Z/T) = Q$ where Q is a positive definite matrix and plim$(Z'\varepsilon/T) = 0$. Then $\hat{\beta}$ is consistent.

If also

$$\frac{Z'\varepsilon}{\sqrt{T}} \xrightarrow{D} N(0, \sigma^2 Q),$$

then

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{D} N(0, \sigma^2 Q^{-1}).$$

However, in dynamic models with autocorrelation,

$$plim(Z'\varepsilon/T) \neq 0. \; (Why?)$$

In this case, an easy way to get consistent estimates is to use the method of instrumental variables.

Instrumental Variables (*IV*):

To get consistent estimates, instead of usign matrix Z, use a $T \times K$ matrix of *instruments* W with $p\lim(W'u/T) = 0$ where u is the vector of errors. The instrumental variables estimator $\hat{\beta}_{IV}$ is $(W'Z)^{-1}W'y$ where W is the matrix of instruments.

To obtain this, multiply $y = Z\beta + u$ by W and write $W'y = W'Z\beta + W'u$.

Note that $W'Z$ is $K \times K$. If it is nonsingular, then $\hat{\beta}_{IV} = (W'Z)^{-1}W'y$.

Proposition: If plim$(W'Z/T) = Q$ where Q is a positive definite matrix and plim$(W'u/T) = 0$, then $\hat{\beta}_{IV}$ is consistent.

Proof:

Note that $\hat{\beta}_{IV} = (W'Z)^{-1}W'y = \beta + (W'Z)^{-1}W'u$.

Thus $\hat{\beta}_{IV} = \beta + (W'Z/T)^{-1}(w'u/T)$. So,

plim$(\hat{\beta}_{IV} - \beta) = Q^{-1}$plim$(W'u/T) = 0$.

Hence $\hat{\beta}_{IV}$ is consistent. ■

Proposition: The asymptotic variance of $\hat{\beta}_{IV}$ is

$$(W'Z)^{-1}W'VW(W'Z)^{-1},$$

where $V = V(u)$.

Notes:

1. If Z is nonstochastic (i.e., $Z = X$), then X is a good choice of instruments.

2. If $Z = [Y_{-1}X]$ where Y_{-1} represents the vector of lagged y values, then use X and more columns as instruments. For example, use in W *all* X variables uncorrelated with errors *and* additional variables as needed so that W has rank K. Usually, additional lagged values of X are used. Predicted values of Y_{-1} based on lagged X and other variables can also be used.

3. $\hat{\beta}_{IV}$ is *not* usually efficient. The ML estimator is more difficult to calculate but it is better.

4. The ML estimator requires specification of the form of autocorrelation. IV estimation gives consistent estimates without specification of the form of autocorrelation.

