University of Illinois
Fall 2016

Department of Economics
Roger Koenker

Economics 536
Lecture 7

Introduction to Specification Testing in Dynamic Econometric Models

In this lecture I want to briefly describe some techniques for evaluating dynamic econometric models like the models for gasoline demand you have been estimating. Until now, we have implicitly assumed that these models satisfied the classical assumptions of the Gaussian linear model. In particular, we have assumed that the error sequences $\{u_t\}$ were iid and approximately Gaussian, thus justifying the application of elementary least squares methods of estimation.

Testing for Autocorrelation
We might begin by recalling some basic facts about autocorrelation. In classical regression with fixed regressors,

$$y_i = x_i'\beta + u_i$$

we know that if the vector, $u = (u_i) \sim \mathcal{N}(0, \sigma^2 I)$ then

$$\hat{\beta} = (X'X)^{-1}X'y \sim \mathcal{N}(\beta, \sigma^2(X'X)^{-1})$$

but when the errors are autocorrelated, for example, $u \sim \mathcal{N}(0, \Omega)$, then

$$\hat{\beta} \sim \mathcal{N}(\beta, (X'X)^{-1}X'\Omega X(X'X)^{-1})$$

and therefore the conventional estimates of standard errors from ordinary least squares regression may badly misrepresent the true precision of $\hat{\beta}$. Of course, in this case it is preferable to use

$$\check{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \sim \mathcal{N}(0, (X'\Omega^{-1}X)^{-1})$$

which, in effect, restores the model to the original iid structure and thereby achieves optimality. DiNardo and Johnston give an example, a standard one, showing that the precision of $\check{\beta}$ can be considerably greater than that of $\hat{\beta}$, even for modest amounts of autocorrelation.

One can estimate deirectly the OLS covariance matrix using a variant of the proposal of Newey and West (1987)

$$\hat{\Omega} = \hat{\Gamma}_0 + \sum_{j=1}^{n}(1 - j/(p+1))(\hat{\Gamma}_j + \hat{\Gamma}_j')$$

where

$$\hat{\Gamma}_j = n^{-1} \sum_{t=j+1}^{n} u_t u_{t-j} x_t x_{t-j}'$$

When dynamic models are specified with lagged endogenous explanatory variables, autocorrelation raises more serious problems. Even in the simplest examples, there are bias as well as efficiency costs to ignoring the autocorrelation. To see this consider the model

$$y_t = \beta_0 y_{t-1} + \beta_1 x_t + u_t$$

where

$$u_t = \rho u_{t-1} + \varepsilon_t.$$

with $\{\varepsilon_t\}$ assumed to be iid $\mathcal{N}(0, \sigma^2)$. Consistent estimation of β requires orthogonality of u_t and the "explanatory variables" (y_{t-1}, x_t), but note that

$$\begin{aligned} Ey_{t-1}u_t &= E(\beta_0 y_{t-2} + \beta_1 x_{t-1} + u_{t-1})(\rho u_{t-1} + \varepsilon_t) \\ &= \rho\sigma^2 + \beta_0 \rho E(y_{t-2}u_{t-1}) \end{aligned}$$

But, by stationarity, a concept introduced in the next lecture, $Ey_{t-1}u_t = Ey_{t-2}u_{t-1}$ so

$$Ey_{t-1}u_t = \rho\sigma_u^2/(1 - \beta_0\rho)$$

Not surprisingly, given the serious consequences of this bias effect, there is a large literature on testing for autocorrelation in this context. The most straight forward approach is that of Breusch and Godfrey which derives from work of Durbin. While the details are usually reserved for Ec575, the Breusch and Godfrey test is easily described.

Let $\hat{u}_t = y_t - x_t'\hat{\beta}, \;\; t = 1, \cdots, n$ denote the residuals from a least squares regression in which the vector x_t may include lagged endogenous variables. Suppose that we wish to test the hypothesis

$$H_o: \;\; \rho_1 = \rho_2 = \cdots = \rho_s = 0$$

in the potential autocorrelation model

$$u_t = \rho_1 u_{t-1} + \cdots + \rho_s u_{t-s}.$$

Consider the auxiliary regression equation

$$\hat{u}_t = \rho_1 \hat{u}_{t-1} + \cdots + \rho_s \hat{u}_{t-s} + x_t' \gamma + v_t$$

and the associated test statistic

$$T_n = nR^2$$

based on the conventional R^2 of the auxiliary regression. Under $H_0, T_n \sim \chi^2_s$, providing a rather general testing strategy for this important class of models. Note that it is crucial that the variables x_t appear in the auxiliary regression, even though in some superficially similar circumstances this is found to be unnecessary.

Digression on R^2 asymptotics

The connection between R^2 and F is an important aspect of trying to interpret nR^2 as a reasonable test statistic. For a general linear hypothesis, say $R\beta = r$, in the regression setting, let $\mathcal{S}_\omega$ and $\mathcal{S}_\Omega$ denote the restricted and unrestricted sums of squared residuals, and

$$F = \frac{(\mathcal{S}_\omega - \mathcal{S}_\Omega)/q}{\mathcal{S}_\Omega/(n-p)}$$

and we can define an R^2 of ω relative to Ω as,

$$R^2 = 1 - \frac{\mathcal{S}_\Omega}{\mathcal{S}_\omega}$$

thus

$$\frac{n-p}{q} \; \frac{R^2}{1-R^2} = \frac{(\mathcal{S}_\omega - \mathcal{S}_\Omega)/q}{\mathcal{S}_\Omega/(n-p)}$$

Under the null hypothesis we have suggested that nR^2 can be used as a asymptotically valid test statistic which has χ^2_q behavior under the null, where $q =$ rank (R). What is the connection of this to the foregoing algebra which relates the finite sample value of R^2 to the appropriate F statistic for testing H_0? Note first that under H_0 $\mathcal{S}_\Omega/(n-p) \to \sigma^2$ and $\mathcal{S}_\omega/(n-p) \to \sigma^2$ so for n large,

$$nR^2 = \frac{\mathcal{S}_\omega - \mathcal{S}_\Omega}{\mathcal{S}_\omega/n} \approx \frac{\mathcal{S}_\omega - \mathcal{S}_\Omega}{\mathcal{S}_\Omega/n}$$

and therefore nR^2 is approximately equal to the numerator χ^2_q of the F statistic. Now to make the connection between χ^2_q and F we need only note that $\chi^2_q/q \sim F_{q,\infty}$. Alternatively, we may observe that under $H_0, R^2 \to 0$ so

$$\frac{n-p}{q}\ \frac{R^2}{1-R^2} \to \frac{n}{q}R^2$$

The next obvious question is: what do we do if the Breusch-Godfrey test rejects H_0? There are two general approaches which I will describe briefly.

Nonlinear Models

Dynamic models of the type we have been discussing can always be written in the form

$$y_t = D(L)x_t + u_t$$

where $D(L) = B(L)/A(L)$ is called a rational lag polynomial. As an example, take the simple model

$$(*) \qquad y_t = \alpha y_{t-1} + \beta x_t + u_t$$

with $u_t = \rho u_t + \varepsilon_t$ and $\{\varepsilon_t\}$ iid $\mathcal{N}(0, \sigma^2)$. Subtracting $\rho y_{t-1} = \rho\alpha y_{t-2} + \rho\beta x_{t-1} + \rho u_t$ from (*) we have

$$y_t = (\rho + \alpha)y_{t-1} - \rho\alpha y_{t-2} + \beta x_t - \beta\rho x_{t-1} + \varepsilon_t.$$

Since this version of the model has a nice (iid) error structure, it can be consistently estimated by ordinary least squares. Note, however, that we now have 4 parameters not 3, as in the original formulation of the model. Since these 4 parameters are simple functions of the original 3 parameters, we can impose the implied constraints and estimate the model by nonlinear least squares.

A related approach, which we may also illustrate with this simple model, is to write the original model in the form

$$y_t = \sum_{j=0}^{\infty} \delta_j x_{t-j} + v_t.$$

This form may appear impractical since we have an infinite number of δ_j's, but as we have seen in the second lecture these δ's may be expressed in terms of a finite number of α's and β's. Harvey (1989) discusses several versions of this in some simple models and the resulting nonlinear least squares estimation strategy.

Instrumental Variable Estimation

An alternative strategy for estimating models of this type relies on instrumental variables. Recall that our fundamental problem, the bias resulting from autocorrelation in dynamic models, was attributed to the lack of orthogonality between errors and lagged endogenous variables. An obvious strategy for dealing with this problem is to identify suitable instrumental variables (IV's) which have the properties:

- orthogonality with u, i.e., $EZ'u = 0$
- Mutual association with X, i.e., $E\hat{X}'X$ positive definite, where $\hat{X} = Z(Z'Z)^{-1}Z'X$ is the projection of X onto the column space of Z.

The immediate question is where would such variables, Z, come from? In the case of dynamic models of the type we have been discussing it is quite straightforward to answer this question. They may be chosen to be lagged exogenous variables. But this question changes a worry that there might be too few IV's into a new worry that there might be too many. This question does not have a good analytical answer and remains a question of active research concern. Nevertheless, empirical research has suggested some rules of thumb which can be used in applications. Generally, we would suggest that the number of IV's be kept to some modest number above the number of existing explanatory variables, say, $p \leq q \leq 2p$ where p is the parametric dimension of the original model and q is the number of IV's. We will return to this question later in the course.

ARCH in Brief

Frequently, we observe (particularly in financial data) time-varying heteroscedasticity. In the early 80's Engel coined the term autoregressive conditional heteroscedasticity ARCH to refer to model in which

$$V(u_t | \text{ past }) = h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_p u_{t-p}^2$$

In subsequent work this has been generalized in several directions notably to GARCH (1,1),

$$V(u_t | \text{ past }) = h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \gamma_1 h_{t-1}$$

These are simple examples of a broad class of nonlinear time series models.

In the ARCH model we have unconditional expectations,

1.) $Eu_t = Eh_t \varepsilon_t = 0$

2.) $Vu_t = \alpha_0 / A(1)$ where $A(L) = (1 + \alpha_1 L + \cdots + \alpha_p L^p)$

As long as the lag polynomial $A(L)$ is stable, i.e., if the roots of $A(z) = 0$ lie outside the unit circle, then we get a gradual oscillation of h_t around the unconditional variance. In integrated ARCH, i.e., roots *on* the unit circle, we get long swings away from the initial h_t.

Testing for ARCH

Naive LM tests can be implemented just like LM-AR tests. Regress $\hat{u}_t^2$ on $\{\hat{u}_{t-1,\cdots,}^2 \ \hat{u}_{t-q}^2 \ \text{ and } \ x_t\}$ compute nR^2 compare to χ_q^2 or nR^2/q compare to $F_{q,n-p}$. Note that in this form the test may be regarded as a joint test for ARCH and heteroscedasticity of the form usually tested by the Breusch-Pagan, and related tests. One could obviously consider refining the hypothesis under consideration in light of the results obtained for this expanded version of the test.

The LM Principle

Let $\ell(\tilde{\theta})$ denote log likelihood evaluated at mle under $H_0: \ \theta \in \Theta_0 \subset \Theta_1$ we are interested in testing H_0 vs $H_1 : \theta \in \Theta_1$. One way to do this is to ask how does ℓ change as we move the restricted estimator $\tilde{\theta} \in \Theta_0$ toward the unrestricted $\hat{\theta} \in \Theta_1$. To explore this we need to say something about nonlinear optimization, but this would take us too far away from the main topic. Sufficient to say that the LM-test is based on the magnitude of the gradient of ℓ at $\hat{\theta} \in \Theta_0$ in the direction of $\hat{\theta}_1 \in \Theta_1$. Again, there are questions to be answered about what to do when ARCH effects are found to be present in the model. Joint estimation of ARCH and regression shift effects is the preferred solution when it is computationally feasible. Various iterative solutions are obviously available as alternatives.