University of Illinois
Fall 1998

Department of Economics
Roger Koenker

Economics 472
Lecture 11

Specification Tests in Simultaneous Equations Models

In this lecture we will consider the problem of testing the validity of certain hypotheses concerning the exogeneity of variables in simultaneous equation models. These tests are often referred to as "Hausman" tests because they were reconsidered in a systematic way by Hausman (1978).

Consider the single equation model,

$$\begin{aligned} y &= Y\gamma + X_1\beta_1 + X_2\beta_2 + u \\ &= Z\delta + u \end{aligned}$$

where we will refer to the variables Y as "included endogenous," X_1 as "included exogenous," and X_2 as "dubious exogenous." Under the null hypothesis we maintain that X_2 is exogenous, while under the alternative hypothesis we regard X_2 as endogenous. We can express this concisely as,

$$H_0: \; X_2 \perp u \quad \text{vs} \quad H_1: \; X_2 \not\perp u$$

Under the null hypothesis we have the estimator

$$\hat{\delta} = \text{ argmin } \hat{u}' P_X \hat{u} = (Z' P_X Z)^{-1} Z' P_X y$$

where $P_X = X(X'X)^{-1}X$ and $X = [X_1 \vdots X_2 \vdots X_3]$ denotes the full set of exogenous variables, under H_0. Under H_1, the columns of X_2 are no longer valid instrumental variables so we have the restricted set of exogenous variables $\tilde{X} = [X_1 \vdots X_3]$, and the corresponding estimator

$$\tilde{\delta} = \text{ argmin } \hat{u}' P_{\tilde{X}} \hat{u} = (Z' P_{\tilde{X}} Z)^{-1} Z' P_{\tilde{X}} y$$

The only difference between the two estimators is that they use different sets of instrumental variables in the first stage of two stage least squares.

How are we to test H_0 vs. H_1? The basic idea of the Hausman test is quite simple. If H_0 is true both $\hat{\delta}$ and $\tilde{\delta}$ are consistent estimators of δ, that is they both converge almost surely to δ. However, under $H_1, \hat{\delta}$ is biased while $\tilde{\delta}$ is still consistent. Thus, if we could determine the distribution of

$$\Delta = \hat{\delta} - \tilde{\delta}$$

under H_0 we could exploit the fact that it should have mean zero (in large samples) under H_0 and nonzero mean under H_1. We now fill in some details of this argument.

Under H_0, we have

$$(\hat{\delta} - \delta) \rightsquigarrow \mathcal{N}(0, \sigma^2(Z'P_X Z)^{-1})$$

while

$$(\tilde{\delta} - \delta) \rightsquigarrow \mathcal{N}(0, \sigma^2(Z'P_{\tilde{X}} Z)^{-1}).$$

Now, what about the distribution of Δ? We would expect that $\hat{\delta}, \tilde{\delta}$ would be (asymptotically) jointly normal and we would have

$$\text{Var}\ (\hat{\delta} - \tilde{\delta}) = \text{ Var}\ (\hat{\delta}) + \text{ Var}\ (\tilde{\delta}) - 2\ \text{Cov}\ (\hat{\delta}, \tilde{\delta})$$

where the first terms on the right hand side are given above. What about the Covariance term? We now make a rather surprising claim. We will provide two different proofs of this result since it is (a.) quite easy, and (b.) quite fundamental.

Lemma: Under H_0, $\text{Var}\ (\hat{\delta} - \tilde{\delta}) = \text{ Var}\ (\tilde{\delta}) - \text{ Var}\ (\hat{\delta})$.

proof 1: This proof is quite general and is a "proof by contradiction." It may be thought of as an asymptotic version of Basu's (1955) theorem which is discussed further in 476. By the theorem on the optimality of two stage least squares in the previous lecture we know that $\text{Var}\ (\hat{\delta})$ is as small as possible among linear estimators. Consider

$$\begin{aligned} \text{Var}\ (\check{\delta}) &= \text{Var}((1-\lambda)\hat{\delta} + \lambda\tilde{\delta}) = V(\hat{\delta} + \lambda(\hat{\delta} - \tilde{\delta})) \\ &= V(\hat{\delta}) + \lambda^2 V(\tilde{\delta} - \hat{\delta}) + 2\lambda \text{Cov}(\hat{\delta} - \tilde{\delta}, \hat{\delta}). \end{aligned}$$

Note that for λ sufficiently small the covariance term is larger than the $\lambda^2 V(\tilde{\delta} - \hat{\delta})$ term, hence *for λ sufficiently small*, the variance of $\check{\delta}$ could be made smaller than $\text{Var}(\hat{\delta})$ if $\text{Cov}(\hat{\delta} - \tilde{\delta}, \hat{\delta})$ were non-zero. (Note that λ could be negative in the case that the covariance term were positive.) But this would contradict the efficiency of $\hat{\delta}$, so it follows that

$$\text{Cov}(\hat{\delta} - \tilde{\delta}, \hat{\delta}) = 0$$

and this immediately implies

$$\text{Cov}(\tilde{\delta}, \hat{\delta}) = \text{Var}(\hat{\delta})$$

which yields the result stated in the Lemma.

Proof 2: A more direct proof goes like this:

$$\Delta = [(Z'P_X Z)^{-1} X' P_X - (Z'P_{\tilde{X}} Z)^{-1} Z' P_{\tilde{X}}] u \equiv (H - G)u$$

so $E\Delta\Delta' = \sigma^2[HH' + GG' - HG' - G'H]$ where $HH' = (Z'P_X Z)^{-1}$ $GG' = (Z'P_{\tilde{X}} Z)^{-1}$ and

$$HG' = (Z'P_X Z)^{-1} Z' P_X P_{\tilde{X}} Z (Z' P_{\tilde{X}} Z)^{-1}.$$

But $P_X P_{\tilde{X}} = P_{\tilde{X}}$ (explain why?) so $E\Delta\Delta' = GG' - HH'$ which completes the proof.

A worthwhile exercise: Explain why the matrix $GG' - HH'$ is positive semi-definite?

Implementation of the test

We have shown that, under H_0,

$$\hat{\Delta} \rightsquigarrow \mathcal{N}(0, \mathrm{Var}(\tilde{\delta}) - \mathrm{Var}(\hat{\delta}))$$

while under $H_1, \hat{\Delta}$ has a nonzero mean and the some covariance matrix. Thus a natural test statistic for H_0, would seem to be

$$T = \hat{\Delta}'(\mathrm{Var}(\tilde{\delta}) - \mathrm{Var}(\hat{\delta}))^{-1}\hat{\Delta}.$$

This appears to be quite easy to implement since the two covariance matrixes are typically returned as a by-product of estimating $\hat{\delta}$ and $\tilde{\delta}$. Unfortunately, the situation is not quite so simple. The major potential problem, at least in some practical settings is that the matrix

$$\Omega = \mathrm{Var}(\tilde{\delta}) - \mathrm{Var}(\hat{\delta})$$

is singular and therefore the inverse required for computing T doesn't exist. What is to be done in this case? Fortunately, this problem has a relatively simple solution: we need to replace Ω^{-1} by its *generalized* inverse. The degree of freedom of the statistic T in this case is simply the rank of the matrix Ω. We illustrate this approach in the following example.

Consider the simple supply and demand model of part 2 of Problem Set 4 ,

$$\begin{aligned} Q_t &= \alpha_0 + \alpha_1 p_t + \ \alpha_2 p_{t-1} + \alpha_3 z_t + u_t \\ p_t &= \beta_0 + \beta_1 Q_t + \beta_2 w_t + v_t \end{aligned}$$

and suppose we wish to test the legitimacy of treating p_{t-1} as exogenous in the first equation. We will *maintain* the exogeneity of $W_1 = [z_t, w_t, z_{t-1}, w_{t-1}]$ and ask whether expanding the set of instruments to $W_0 = [W_1 \vdots p_{t-1}]$ is fine or foolish.

We begin by estimating the first equation by 2SLS for both W_0 and W_1 and forming, $\hat{q} =$ $0.024, -0.032, 0.035, -0.00151)$. We then compute $D'D = (W'P_1W)^{-1} - (W'P_0W)^{-1}$ which we find has one non-zero eigenvalue of $\lambda = .002603$ and corresponding eigenvector $v = (0.451, 0.594, -0.664, 0.028)$. Note that these numbers come from an earlier version of the data, so results may be expected to differ for the new data. Using the second result from the next section we may compute $\hat{q}'(D'D)^{-1}\hat{q} =$ $.112359$, which upon dividing by the estimate of σ^2 from 2SLS under H_0 of 2.06325 yields a test statistic of .0544, which is a suspiciously small value for a χ^2 on one degree freedom, but must be interpreted as supporting our hypothesis that p_{t-1} can be treated as exogenous. It is interesting to note, see Hausman and Taylor(1981), that repeating the computation, but using only the coefficients of p_t and p_{t-1} would yield identical results. It would not, however, allow us to escape the computation of the g-inverse. The required g-inverse would now be rank 2 rather than 4.

A Note on G-inverses

This is a large topic, a good introduction, and to all topics on linear algebra related to statistics, is Rao(1973).

Definition. Let A be a $n \times p$ matrix of rank r. A g-inverse of A is a $p \times n$ matrix A^-, such that $x = A^-y$ is a solution to $Ax = y$ for any y in the space spanned by the columns of A. A fundamental property of g-inverses is the following:

$$AA^{-}A = A.$$

There are many such g-inverses and their distinctions need not concern us since all are equivalent for the purposes of this lecture. We will illustrate the computation of g-inverses in two very special cases. Case (2.) is used in the computation reported above. Case (1) should look familiar.

1. If A is a $n \times p$ real matrix with $\text{rank}(A) = p \leq p \leq n$, then $A^{-} = (A\prime A)^{-1}A'$ is a g-inverse of A.

2. If A is a $n \times p$ real matrix with $\text{rank}(A) = r \leq p \leq n$, with non-zero eigenvalues $\lambda_1, \ldots, \lambda_r$ and corresponding eigenvectors $v_1, \ldots, v_r$ of $A'A$, then $A^{-} = \lambda_1^{-1}v_1v_1' + \ldots + \lambda_r^{-1}v_rv_r'$ is a g-inverse of $A'A$.

References

Basu, D. (1955) On Statistics Independent of a Complete Sufficient Statistic, *Sankhya* 15, 377-380.

Hausman, J. (1978) Specification Tests in Econometrics, *Econometrica.* 46, 1241-71.

Hausman, J. and W. Taylor (1981), A generalized specification test, *Economics Letters*, 8, 239-45.

Rao, C.R. (1973) *Linear Statistical Inference and Its Applications* (Wiley).