

Specification Tests: A Panel Data Example

C. Flinn

March 5, 1999

The general idea for the specification test (which is fundamentally based on the Rao-Blackwell Theorem) is the following. Let there exist two estimators for an unknown parameter θ - let this parameter be a scalar for present purposes. Denote the estimators by $\hat{\theta}$ and $\tilde{\theta}$. There are two possible true states of the world. Under the null hypothesis, we have that

$$\begin{aligned}H_0 : \text{plim}(\hat{\theta}) &= \theta \\ \text{plim}(\tilde{\theta}) &= \theta \\ \text{avar}(\hat{\theta}) &= \text{Cramer-Rao Lower Bound.}\end{aligned}$$

Under the alternative state of the world,

$$\begin{aligned}H_A : \text{plim}(\hat{\theta}) &\neq \theta \\ \text{plim}(\tilde{\theta}) &= \theta.\end{aligned}$$

Consider a simple panel data application

$$y_{it} = \alpha_0 + \alpha_1 x_{it} + \varepsilon_{it}; \quad i = 1, \dots, N; \quad t = 1, 2,$$

with

$$\varepsilon_{it} = \eta_i + \xi_{it}.$$

Assuming that ξ_{it} is i.i.d. with mean 0 and variance σ_ξ^2 , we define two estimators

of α_1 . The first estimator is the FGLS estimator. Stack the observations as follows:

$$y = \begin{bmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{N1} \\ y_{12} \\ y_{22} \\ \vdots \\ y_{N2} \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{N1} \\ 1 & x_{12} \\ 1 & x_{22} \\ \vdots & \vdots \\ 1 & x_{N2} \end{bmatrix}, \quad \alpha = \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{N1} \\ \varepsilon_{12} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{N2} \end{bmatrix}.$$

Now consider

$$\begin{aligned} H_0 : E(\eta_i | x_{i1}, x_{i2}) &= 0, \\ E(\eta_i^2 | x_{i1}, x_{i2}) &= \sigma_\eta^2, \forall i. \end{aligned}$$

In this case the error term ε_{it} is mean independent of the regressors, and we have

$$\begin{aligned} E(\varepsilon\varepsilon' | X) &= \begin{bmatrix} \sigma_\eta^2 + \sigma_\xi^2 & 0 & \dots & 0 & \sigma_\eta^2 & 0 & \dots & 0 \\ 0 & \sigma_\eta^2 + \sigma_\xi^2 & \ddots & \vdots & 0 & \sigma_\eta^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_\eta^2 + \sigma_\xi^2 & 0 & \dots & 0 & \sigma_\eta^2 \\ \sigma_\eta^2 & 0 & \dots & 0 & \sigma_\eta^2 + \sigma_\xi^2 & 0 & \dots & 0 \\ 0 & \sigma_\eta^2 & \ddots & \vdots & 0 & \sigma_\eta^2 + \sigma_\xi^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_\eta^2 & 0 & \dots & 0 & \sigma_\eta^2 + \sigma_\xi^2 \end{bmatrix} \\ &= \Omega \equiv \Sigma \otimes I_N. \end{aligned}$$

where

$$\Sigma = \begin{bmatrix} \sigma_\eta^2 + \sigma_\xi^2 & \sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\eta^2 + \sigma_\xi^2 \end{bmatrix}.$$

which implies that OLS is a consistent estimator of α , and furthermore that the Feasible GLS estimator is consistent and asymptotically efficient. Given the OLS residuals, form an consistent estimator of the matrix Σ by estimating the elements $\sigma_\eta^2 + \sigma_\xi^2$ and σ_η^2 . With your estimated $\hat{\Omega}$, the feasible GLS estimator is

$$\hat{\alpha} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y.$$

The asymptotic covariance matrix of the estimator is

$$(X'\hat{\Omega}^{-1}X)^{-1}.$$

Next consider the alternative which specifies

$$H_A : E(\eta_i|x_{i1}, x_{i2}) \neq 0.$$

In this case, the FGLS estimator (and the OLS estimator on which it is based) are inconsistent. We can define another estimator for α_1 (only) as follows. Since

$$\Delta y_i = \alpha_1 \Delta x_i + \Delta \varepsilon_i,$$

where $\Delta z_i \equiv z_{i2} - z_{i1}$, we have $E(\Delta \varepsilon_i | \Delta x_i) = 0$ all i . Thus the OLS estimator of α_1 in the “difference” equation is unbiased and consistent under either H_0 or H_A . This estimator is

$$\tilde{\alpha}_1 = \frac{\sum_{i=1}^N \Delta x_i \Delta y_i}{\sum_{i=1}^N (\Delta x_i)^2}.$$

The asymptotic variance of this estimator is

$$\text{avar}(\tilde{\alpha}_1) = \frac{N^{-1} \sum_{i=1}^N (\Delta y_i - \tilde{\alpha}_1 \Delta x_i)^2}{\sum_{i=1}^N (\Delta x_i)^2}. \quad (0.1)$$

Note: The coefficient associated with any measured variable that is invariant over time for all individuals in the sample cannot be estimated when we difference the data. In this case, since the “constant” is by definition constant, differencing eliminates this term. In comparing the estimators $\hat{\theta}$ and $\tilde{\theta}$, we will have to confine our attention to those elements of the parameter vector θ which appear in both estimators. In this case, that means the slope coefficient α_1 . For the fixed effects estimator of the slope coefficient to be estimable implies that we are assuming that $x_{i2} \neq x_{i1}$ for some i in our sample.

Let's return to the general setup. If under H_0 both $\hat{\theta}$ and $\tilde{\theta}$ have limiting normal distributions, then the statistic

$$\frac{(\hat{\theta} - \tilde{\theta})^2}{\text{avar}(\hat{\theta} - \tilde{\theta})} \sim \chi_{(1)}^2.$$

We have to work out the term in the denominator. Now

$$\text{avar}(\hat{\theta} - \tilde{\theta}) = \text{avar}(\hat{\theta}) + \text{avar}(\tilde{\theta}) - 2 \text{acov}(\hat{\theta}, \tilde{\theta}).$$

and the last term is seemingly not straightforward to work out. This is where the assumption of asymptotic efficiency of θ under the null comes in. Consider a new estimator

$$\theta^* = \hat{\theta} + \frac{[\text{avar}(\hat{\theta}) - \text{acov}(\hat{\theta}, \tilde{\theta})]}{\text{avar}(\hat{\theta} - \tilde{\theta})}(\tilde{\theta} - \hat{\theta}).$$

The asymptotic variance of this new estimator is

$$\begin{aligned} \text{avar}(\theta^*) &= \text{avar}(\hat{\theta}) + \left(\frac{[\text{avar}(\hat{\theta}) - \text{acov}(\hat{\theta}, \tilde{\theta})]}{\text{avar}(\hat{\theta} - \tilde{\theta})} \right)^2 \text{avar}(\tilde{\theta} - \hat{\theta}) \\ &\quad + 2 \text{acov}(\hat{\theta}, \frac{[\text{avar}(\hat{\theta}) - \text{acov}(\hat{\theta}, \tilde{\theta})]}{\text{avar}(\hat{\theta} - \tilde{\theta})}(\tilde{\theta} - \hat{\theta})) \\ &= \text{avar}(\hat{\theta}) + \left(\frac{[\text{avar}(\hat{\theta}) - \text{acov}(\hat{\theta}, \tilde{\theta})]}{\text{avar}(\hat{\theta} - \tilde{\theta})} \right)^2 \text{avar}(\hat{\theta} - \tilde{\theta}) \\ &\quad - 2 \text{acov}(\hat{\theta}, \frac{[\text{avar}(\hat{\theta}) - \text{acov}(\hat{\theta}, \tilde{\theta})]}{\text{avar}(\hat{\theta} - \tilde{\theta})}(\hat{\theta} - \tilde{\theta})) \\ &= \text{avar}(\hat{\theta}) + \left(\frac{[\text{avar}(\hat{\theta}) - \text{acov}(\hat{\theta}, \tilde{\theta})]}{\text{avar}(\hat{\theta} - \tilde{\theta})} \right)^2 \text{avar}(\hat{\theta} - \tilde{\theta}) \\ &\quad - 2 \frac{[\text{avar}(\hat{\theta}) - \text{acov}(\hat{\theta}, \tilde{\theta})]}{\text{avar}(\hat{\theta} - \tilde{\theta})} E(\hat{\theta}[\hat{\theta} - \tilde{\theta}]) \\ &= \text{avar}(\hat{\theta}) + \left(\frac{[\text{avar}(\hat{\theta}) - \text{acov}(\hat{\theta}, \tilde{\theta})]}{\text{avar}(\hat{\theta} - \tilde{\theta})} \right)^2 \text{avar}(\hat{\theta} - \tilde{\theta}) \\ &\quad - 2 \frac{[\text{avar}(\hat{\theta}) - \text{acov}(\hat{\theta}, \tilde{\theta})]}{\text{avar}(\hat{\theta} - \tilde{\theta})} E(\hat{\theta}[(\hat{\theta} - \theta) - (\tilde{\theta} - \theta)]) \\ &= \text{avar}(\hat{\theta}) + \frac{[\text{avar}(\hat{\theta}) - \text{acov}(\hat{\theta}, \tilde{\theta})]^2}{\text{avar}(\hat{\theta} - \tilde{\theta})} \\ &\quad - 2 \frac{[\text{avar}(\hat{\theta}) - \text{acov}(\hat{\theta}, \tilde{\theta})]}{\text{avar}(\hat{\theta} - \tilde{\theta})} [\text{avar}(\hat{\theta}) - \text{acov}(\hat{\theta}, \tilde{\theta})] \\ &= \text{avar}(\hat{\theta}) - \frac{[\text{avar}(\hat{\theta}) - \text{acov}(\hat{\theta}, \tilde{\theta})]^2}{\text{avar}(\hat{\theta} - \tilde{\theta})}. \end{aligned}$$

Under the null $\hat{\theta}$ is efficient and thus has variance at least as small as any other

consistent estimator. Thus it must be the case that $\text{avar}(\hat{\theta}) - \text{acov}(\hat{\theta}, \tilde{\theta}) = 0$. In other words, the asymptotic covariance between any other consistent estimator θ^* and an efficient estimator $\hat{\theta}$ is equal to $\text{avar}(\hat{\theta})$. With this result, we have

$$\begin{aligned} \text{avar}(\hat{\theta} - \tilde{\theta}) &= \text{avar}(\hat{\theta}) + \text{avar}(\tilde{\theta}) - 2\text{acov}(\hat{\theta}, \tilde{\theta}) \\ &= \text{avar}(\hat{\theta}) + \text{avar}(\tilde{\theta}) - 2\text{avar}(\hat{\theta}) \\ &= \text{avar}(\tilde{\theta}) - \text{avar}(\hat{\theta}). \end{aligned}$$

Thus the test statistic has the form

$$\frac{(\hat{\theta} - \tilde{\theta})^2}{\text{avar}(\tilde{\theta}) - \text{avar}(\hat{\theta})} \sim \chi_{(1)}^2.$$

Returning to our application, we first need to define the asymptotic variance of our estimator of $\hat{\alpha}_1$, which is

$$\text{avar}(\hat{\alpha}_1) = [(X'\hat{\Omega}^{-1}X)^{-1}]_{2,2},$$

which indicates that it is the element in the second row and second column of the 2 by 2 matrix $(X'\hat{\Omega}^{-1}X)^{-1}$. Then under the null

$$\frac{(\hat{\alpha}_1 - \tilde{\alpha}_1)^2}{\text{avar}(\tilde{\alpha}_1) - \text{avar}(\hat{\alpha}_1)} \sim \chi_{(1)}^2.$$

As we noted in class, in a finite sample even under the null it may be the case that our estimated $\text{avar}(\tilde{\alpha}_1) - \text{avar}(\hat{\alpha}_1)$ is not positive. In this case the test can not be carried out, at least in this form.

When we generalize the testing situation to a vector-valued parameter θ with k elements, the test is given by

$$(\hat{\theta} - \tilde{\theta})'(\hat{\Sigma}_{\tilde{\theta}} - \hat{\Sigma}_{\hat{\theta}})^{-1}(\hat{\theta} - \tilde{\theta}) \sim \chi_{(k)}^2.$$

That the matrix $\hat{\Sigma}_{\tilde{\theta}} - \hat{\Sigma}_{\hat{\theta}}$ may not be positive definite in any finite sample is a practical problem for large k .