Exercise Sheet 2: Specification Tests for Panel Data Models

Review the Concepts and Proofs

1. Why would one test for an unobserved individual specific effect?
2. Explain Wooldridge's test of an unobserved individual specific effect and derive its asymptotic distribution.
3. Describe the Breusch-Pagan LM test. What is the consequence of obtaining a test statistic of $LM = 8.3$?
4. Why would one test for autocorrelation in a linear panel data model?
5. Describe the test of first-order autocorrelation in v_{it}.
6. Describe the test of first-order autocorrelation in u_{it}.
7. Prove the asymptotic distribution of the classical Hausman test.
8. How can you obtain a Hausman test that is robust to arbitrary autocorrelation or heteroscedasticity in $\mathbf{v}_i$?
9. Why is it advisable to include a large set of individual-specific controls whenever you apply the RE estimator?

Paper-pen exercises

1. Consider the error-components model

$$\mathbf{y}_i = \mathbf{x}_i\boldsymbol{\beta} + \mathbf{v}_i, \qquad \mathbf{v}_i = \boldsymbol{\iota}_T c_i + \mathbf{u}_i. \tag{1}$$

Suppose that, conditional on $\mathbf{x}_i$, $\mathbf{v}_i \sim \text{Normal}(\mathbf{0}, \boldsymbol{\Omega})$ where

$$\boldsymbol{\Omega} = \sigma_c^2 \boldsymbol{\iota}_T \boldsymbol{\iota}_T' + \sigma_u^2 \mathbf{I}_T.$$

In the following, derive step by step the Breusch-Pagan LM test of the null hypothesis $\sigma_c^2 = 0$. Use (without proof) the fact that the information matrix is block diagonal between $\boldsymbol{\beta}$ and the variance parameters $\boldsymbol{\theta} = (\sigma_u^2, \sigma_c^2)'$ which allows you to compute score and Hessian solely for $\boldsymbol{\theta}$.

(a) Show that for random sampling of $\mathbf{y}_i$ and $\mathbf{x}_i$, $i = 1, \ldots, N$, the conditional log-likelihood function is

$$\ell_i = \frac{1-T}{2} \log \sigma_u^2 - \frac{1}{2} \log(T\sigma_c^2 + \sigma_u^2) - \frac{(T\sigma_c^2 + \sigma_u^2)^{-1}}{2} \mathbf{v}_i' \mathbf{J}_T \mathbf{v}_i - \frac{(\sigma_u^2)^{-1}}{2} \mathbf{v}_i' \mathbf{Q}_T \mathbf{v}_i.$$

Note that the conditional log density of the multivariate normal distribution is, up to an irrelevant constant, $\log f(\mathbf{v}_i) = -\frac{1}{2} \log |\boldsymbol{\Omega}| - \frac{1}{2} \mathbf{v}_i' \boldsymbol{\Omega}^{-1} \mathbf{v}_i$ and recall that $\boldsymbol{\Omega}^{-1} = \frac{\phi^2}{\sigma_u^2} \mathbf{J}_T + \frac{1}{\sigma_u^2} \mathbf{Q}_T$. Hint: use the rule $|\mathbf{A}'\mathbf{A} + \mathbf{I}_n| = |\mathbf{A}\mathbf{A}' + I_m|$, where $\mathbf{A}$ is a $m \times n$ matrix, and the rule $|c\mathbf{M}| = c^T |\mathbf{M}|$, where c is a scalar and $\mathbf{M}$ is a $T \times T$ matrix.

(b) Find the score $\mathbf{s}_i(\boldsymbol{\theta})$ and Hessian $\mathbf{H}_i(\boldsymbol{\theta})$ of ℓ_i with respect to the parameter vector $\boldsymbol{\theta} = (\sigma_u^2, \sigma_c^2)'$. Evaluate Score and Hessian under the null hypothesis $\sigma_c^2 = 0$.

(c) Find the conditional expectation $\mathbf{A}(\mathbf{x}_i, \boldsymbol{\theta}) \equiv -\operatorname{E}[\mathbf{H}_i(\boldsymbol{\theta})|\mathbf{x}_i]$ under the null hypothesis $\sigma_c^2 = 0$.

(d) Find $\frac{1}{N} \sum_{i=1}^N \tilde{\mathbf{s}}_i = \frac{1}{N} \sum_{i=1}^N \mathbf{s}_i(\tilde{\boldsymbol{\theta}})$ and $\frac{1}{N} \sum_{i=1}^N \tilde{\mathbf{A}}_i = \frac{1}{N} \sum_{i=1}^N \mathbf{A}(\mathbf{x}_i, \tilde{\boldsymbol{\theta}})$, where $\tilde{\boldsymbol{\theta}}$ is the CML estimator under the null hypothesis $\sigma_c^2 = 0$. Without proof use $\tilde{\sigma}_u^2 = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \tilde{v}_{it}^2$, where $\tilde{v}_{it}$ are the pooled OLS residuals.

(e) Show that the LM statistic

$$LM = N \left(\frac{1}{N} \sum_{i=1}^N \tilde{\mathbf{s}}_i \right)' \left(\frac{1}{N} \sum_{i=1}^N \tilde{\mathbf{A}}_i \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^N \tilde{\mathbf{s}}_i \right)$$

can be simplified to

$$LM = \frac{NT}{2(T-1)} \left(\frac{\frac{1}{N} \sum_{i=1}^N \tilde{\mathbf{v}}_i' \mathbf{J}_T \tilde{\mathbf{v}}_i}{\tilde{\sigma}_u^2} - 1 \right)^2.$$

(f) Show that the LM statistic can be reformulated to

$$LM = \frac{NT(T-1)}{2} \left(\frac{\ddot{\sigma}_u^2}{\tilde{\sigma}_u^2} - 1 \right)^2,$$

where $\ddot{\sigma}_u^2 \equiv \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T} (\tilde{v}_{it} - \bar{\tilde{v}}_i)^2$. Interpret the statistic.

2. Consider the error-components model

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + v_{it}, \qquad v_{it} = c_i + u_{it}$$

with $\mathrm{E}(c_i) = 0$, $\mathrm{E}(u_{it}) = 0$, and $\mathrm{E}(c_i u_{it}) = 0$.

   (a) Show without using an expectation operator that for $T = 2$, the within-transformed disturbances $\ddot{u}_{it}$ have first-order autocorrelation $\mathrm{Corr}(\ddot{u}_{i2}, \ddot{u}_{i1}) = -1$.

   (b) Show that in general the within-transformed disturbances $\ddot{u}_{it}$ have first-order autocovariance $\mathrm{E}(\ddot{u}_{it}\ddot{u}_{it-1}) = -\sigma_u^2/T$ and first-order autocorrelation $\mathrm{Corr}(\ddot{u}_{it}, \ddot{u}_{it-1}) = (1-T)^{-1}$ if u_{it} is white noise with variance σ_u^2.

3. Consider the simple model

$$y_{it} = \alpha + x_t\beta + v_{it},$$

where x_t is a scalar regressor that varies solely with t (e.g., a time dummy or an aggregate control variable). Show that the FE and RE estimators of β are numerically identical. Hint: apply POLS to the within-transformed equation (which yields the FE estimator) and to the quasi-demeaned equation (which yields the RE estimator). In the latter case, use an arbitrary value $\lambda = 1 - \phi$ for the quasi demeaning and show that the two estimators are identical for any λ.

4. Consider the structural equation

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + v_{it}, \qquad v_{it} = c_i + u_{it}, \tag{2}$$

where $\mathbf{x}_{it}$ is a $1 \times K$ vector which includes only regressors that vary across i and t, and the augmented equation

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i\boldsymbol{\delta} + r_{it}, \tag{3}$$

where $\bar{\mathbf{x}}_i$ is a $1 \times K$ vector of time averages. For what follows assume that RE.3 holds and the variance components of the RE estimator are estimated using the Swamy-Arora approach.

(a) Denote the RE estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ in (3) by $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\delta}}$. Show that $\tilde{\boldsymbol{\beta}}$ is identical to the FE estimator of $\boldsymbol{\beta}$ in (2), and $\tilde{\boldsymbol{\delta}}$ is identical to the difference of the between and FE estimators of $\boldsymbol{\beta}$ in (2): $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{RE}$ and $\tilde{\boldsymbol{\delta}} = \hat{\boldsymbol{\beta}}_B - \hat{\boldsymbol{\beta}}_{FE}$. Hint: inversion of the partitioned matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

yields

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{D} & -\mathbf{D}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{D} & \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{D}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \end{bmatrix},$$

where $\mathbf{D} = (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}$.

(b) Find the Wald statistic to test the null hypothesis $\boldsymbol{\delta} = \mathbf{0}$ based on the classical RE estimator of the augmented equation (3).

(c) Show that the classical Hausman statistic applied to equation (2) is identical to the Wald statistic computed above.

Empirical exercises

1. Reconsider Acemoglu et al. (2008) who analyze the effect of income on democracy. Consult problem set 1 for explanations. Load their data set `AJRY_2008_data.dta`.

   (a) Estimate the specification

   $$dem_{it} = \beta_1 inc_{i,t-1} + \mu_t + c_i + u_{it} \tag{4}$$

   by RE. Test for existence of unobserved country-specific effects.

   (b) Re-estimate the model with POLS. Check whether $v_{it} = c_i + u_{it}$ is autocorrelated. Interpret the result.

   (c) Re-estimate the model with FE. Check whether u_{it} is autocorrelated. Interpret the result.

   (d) Compare FE and RE by means of a robust Hausman test.