Econometric Methods for Panel Data

Based on the books by BALTAGI: *Econometric Analysis of Panel Data* and by HSIAO: *Analysis of Panel Data*

Robert M. Kunst
robert.kunst@univie.ac.at

University of Vienna
and
Institute for Advanced Studies Vienna

April 11, 2016

Outline

Aspects of endogeneity in panels

Often, some covariates may be considered endogenous, i.e. dependent on the response (notorious example: crime rate may depend on strength of police force). Then, one may consider *instrumental variables* for the endogenous regressors. Alternatively, one may be concerned about individual effects that depend on covariates (Hausman and Taylor).

Instrumental variables may be viewed as exogenous regressors in a larger, virtual multivariate model with simultaneous equations. In the regression, there are four categories of variables: Y endogenous regressors, X_1 exogenous regressors, X_2 instruments, y dependent response,

$$y = \mathbf{Z}\delta + \mathbf{u}, \quad \mathbf{u} = \mathbf{Z}_\mu \mu + \nu,$$

with $\mathbf{Z} = (\mathbf{Y}, \mathbf{X}_1)$.

Within two-stage least squares

Because Y is correlated with u, OLS or LSDV will not be consistent. Purging from individual means by $\mathbf{Q}$ yields

$$\mathbf{Q}y = \mathbf{QZ}\delta + \mathbf{Qu} = \mathbf{QZ}\delta + \mathbf{Q}\nu.$$

Denoting $\tilde{y} = \mathbf{Q}y, \tilde{Z} = \mathbf{Q}Z, \tilde{X} = \mathbf{Q}X$, this is

$$\tilde{y} = \tilde{\mathbf{Z}}\delta + \tilde{\nu}.$$

Then, the within 2SLS estimator

$$\hat{\delta}_{W2SLS} = (\tilde{\mathbf{Z}}'\mathbf{P}_{\tilde{X}}\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}'\mathbf{P}_{\tilde{X}}\tilde{y},$$

where $\mathbf{P}_{\tilde{X}} = \tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'$ projects on the exogenous $\tilde{X}$, will be consistent for δ if $\dim(X) \geq \dim(Z)$.

Random-effects 2SLS

If the true variances of the error components are known, one may also consider the RE estimator, applying the GLS transformation

$$\mathbf{\Omega}^{-1/2} = \sigma_1^{-1}\mathbf{P} + \sigma_\nu^{-1}\mathbf{Q}$$

to all variables y, Z, X to yield y^*, Z^*, X^*, remembering $\sigma_1^2 = T\sigma_\mu^2 + \sigma_\nu^2$. The corresponding 2SLS estimator is *generalized two-stage*,

$$\hat{\delta}_{G2SLS} = (\mathbf{Z}^{*\prime}\mathbf{P}_{X^*}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*\prime}\mathbf{P}_{X^*}y^*.$$

An alternative is replacing X^* by $(\mathbf{Q}X, \mathbf{P}X)$, which yields the asymptotically equivalent EC2SLS estimator. Estimates for the variances can be obtained from the within 2SLS, which is consistent even for the RE model. It can be shown that the choice of instruments $(\mathbf{Q}X, \mathbf{P}X)$ or $\mathbf{X}^*$ is optimal.

System estimation

When there are really M simultaneous equations available, efficiency gains can be obtained from three-stage procedures that exploit the error correlation across equations. Formally, the 'efficient' 3SLS estimator or system estimator is written as

$$\hat{\delta}_{E3SLS} = (\mathbf{Z}^{*\prime}\mathbf{P}_{X^*}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*\prime}\mathbf{P}_{X^*}y^*,$$

where y^*, Z^*, X^* now evolve from applying the large matrix $(MNT \times MNT)$

$$\mathbf{\Omega}^{-1/2} = \mathbf{\Sigma}_1^{-1/2} \otimes \mathbf{P} + \mathbf{\Sigma}_\nu^{-1/2} \otimes \mathbf{Q}$$

to y, Z, X. This requires pre-estimation of $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_\nu$ from G2SLS and calculating Cholesky factors.

The model of Hausman and Taylor

Hausman and Taylor (HT) consider the regression model

$$y_{it} = X_{it}'\beta + Z_i'\gamma + \mu_i + \nu_{it},$$

with Z_i time-invariant, and both $X = [X_1, X_2]$ and $Z = [Z_1, Z_2]$, with 1-portions exogenous and 2-portions correlated with the effects but uncorrelated with ν. $\dim(X_1) \geq \dim(Z_2)$ is an important identification condition.

The RE estimator is inconsistent here, as X_2 and Z_2 violate the exogeneity condition for regressors. The FE estimator is consistent for β but it cannot estimate γ.

The Hausman-Taylor estimator

HT suggest IV estimation on an RE-transformed regressions

$$y^* = \mathbf{X}^* \beta + \mathbf{Z}^* \gamma + u^*,$$

with instruments comprising $(\mathbf{Q}X, \mathbf{P}X_1, Z_1)$. This again requires pre-estimation of variances of variance components, using some consistent first-stage estimator, for example FE for β followed by some two-stage between estimator for γ.

A variant by Amemiya and McCurdy uses time-constant replicates of all T observations of X_1 instead of $\mathbf{P}X_1$ and is regarded as more efficient.