An introduction to Hausman-Taylor model

Xiang Ao

January 27, 2009

1 Hausman-Taylor model

Random effects and fixed effects models are used widely in econometrics for panel data. Many economists tend to like fixed-effect model better since it eliminates all the commonality within an individual (or a firm, etc), therefore the unobserved individual heterogeneity is controlled for. However, in a fixed-effect model, any covariates that are constant within individual cannot be included in the estimation. A random-effect model can have a time-invariate variable in the regression; however, it assumes orthogonality between error term and the individual effects, which is often not true.

Hausman-Taylor(1981) (HT) model uses a "mixed" structure to handle this situation that we need to include a time-invariate variable and model unobserved individual heterogeneity. It's "mixed" in the sense that it is between fixed effect and random effect; or a mixture of both.

Consider a standard panel data model with time-invariate variables in it:

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

where Z_i are cross-sectional time-invariant variables. HT splits the covariates into two sets: $X = [X_1; X_2]$ and $Z = [Z_1; Z_2]$ where X_1 and Z_1 are exogenous and X_2 and Z_2 are endogenous, in the sense that they are correlated with μ_i but not ϵ_{it}. Then we have

$$y_{it} = X_{1it}'\beta_1 + X_{2it}'\beta_2 + Z_{1i}\gamma_1 + Z_{2i}\gamma_2 + \mu_i + \epsilon_{it},$$

First we do a "within" transformation, which is basically deduct all the variables in the regression from its group mean (individual mean). In that case, obviously Z's would be removed. Therefore we are left with

$$\tilde{y}_{it} = \tilde{X}_{1it}\beta_1 + \tilde{X}_{2it}\beta_2 + \tilde{\epsilon}_{it}$$

where $\tilde{y}_{it}$ is the "within" transformed y_{it}, etc.

From this equation we can estimate the "within" estimator of β_1 and β_2; call them $\hat{\beta}_{1w}$ and $\hat{\beta}_{2w}$.

Then we obtain the "within" residual:

$$\tilde{d}_{it} = \tilde{y}_{it} - \tilde{X}_{1it}\hat{\beta}_{1w} - \tilde{X}_{2it}\hat{\beta}_{2w}$$

The variance of the idiosyncratic error term, σ_ϵ^2 can be estimated:

$$\hat{\sigma}_\epsilon^2 = \frac{RSS}{N - n}$$

where RSS is the residual sum of squares from the within regression.

Now regress $\tilde{d}_{it}$ on Z_1 and Z_2, using X_1 and Z_1 as instruments. We get $\hat{\gamma}_{1IV}$ and $\hat{\gamma}_{2IV}$, which are consistent estimates of γ_1 and γ_2.

$$\gamma_{IV} = (Z' P_A Z)^{-1} Z' P_A \hat{d}$$

where $P_A = A(A'A)^{-1}A'$ and $A = [X_1, Z_1]$ is a set of instruments.

With $\hat{\gamma}_{1IV}$, $\hat{\gamma}_{2IV}$ and $\hat{\sigma}_\epsilon^2$, we can estimate $\hat{\sigma}_\mu^2$ (procedure of doing this omitted).

Then define $\hat{\theta}_i$ as

$$\hat{\theta}_i = 1 - (\frac{\hat{\sigma}_\epsilon^2}{\hat{\sigma}_\epsilon^2 + T_i \hat{\sigma}_\mu^2})^{1/2}$$

A random effect transformation can be done on each of the variables:

$$w_{it}* = w_{it} - \hat{\theta}_i \bar{w}_i$$

where $\bar{w}_i$ is the within-panel mean. That is, each of y, X and Z are transformed in this way. We have now

$$y_{it}* = X_{it} * \beta + Z_{it} * \gamma + \epsilon_{it}*$$

Then the HT estimator can be obtained by IV regression of $y_{it}*$ on $X_{it}*$ and $Z_{it}*$, with instruments $\tilde{X}_{it}$, $\bar{X}_{1i}$ and Z_{1i}.

2 Implemented in Stata

In Stata, HT model is implemented as xthtaylor for version 10.1.