Hausman Taylor model [1]

1 Fixed and Random effects - Review

1.1 Fixed effects (FE)

In the fixed effects model, we follow panel with subgroups $i = 1, ..., n$ throughout time $t = 1, .., T$.

$$y_{it} = \alpha_i + x_{it}^{'}\beta + \epsilon_{it}, \tag{1}$$

which we can easily estimate by time averaging the y_{it} and x_{it}, where α_i disappears and regress the model

$$(y_{it} - \bar{y}_i) = (x_{it} - \bar{x}_i)^{'}\beta + (\epsilon_{it} - \bar{\epsilon}_i). \tag{2}$$

to get $\hat{\beta}$, the estimates of β To find $\hat{\alpha}_i$, the estimates of α_i, simply use equation (4):

$$\hat{\alpha}_i = y_{it} - x_{it}^{'}\hat{\beta}, \tag{3}$$

The estimated variance of $\hat{\beta}$ is

$$\widehat{var}(\hat{\beta}) = s^2(\mathbf{X}' M_D \mathbf{X})^{-1} \tag{4}$$

where M_D is the residual-maker matrix (w.r.t. time).

All time-invariant characteristics are washed out of the model.

1.2 Random effects (FE)

As in the fixed effect model, we have

$$y_{it} = \alpha_i + x_{it}^{'}\beta + \epsilon_{it}, \tag{5}$$

but now $\alpha_i = \alpha + \mu_i$, which means our model becomes

$$y_{it} = \alpha + x_{it}^{'}\beta + (\mu_i + \epsilon_{it}). \tag{6}$$

An advantage of RE is you can put time-invariant explanatory variables. Estimation can be achieve by Feasible GLS, with

[1] Adapted from Class Lectures notes, Greene [2] [3] and [1]

$$\hat{\beta}_{GLS} = (X_*'X_*)^{-1}(X_*'Y_*) = ((PX)'(PX))^{-1}((PX)'(PY)) = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y). \qquad (\;$$

For the RE model,

$$P = \Omega^{-1/2} = [I_n \otimes \Sigma_T^{-1/2}]$$

where

$$\Sigma_T^{-1/2} = \frac{1}{\sigma_\epsilon}[I_T - \frac{\theta}{T} i_T i_T']$$

and

$$\theta = 1 - \frac{\sigma_\epsilon}{\sqrt{\sigma_\epsilon^2 + T\sigma_\mu^2}}.$$

The feasible estimation is computed by replacing σ_ϵ by

$$\hat{\sigma}_\epsilon^2 = \frac{\sum_{i=1}^n \sum_{t=1}^T (e_{it} - \bar{e}_i)^2}{nT - n - k}$$

calculated by regressing the “deviations from the mean” in each group; that is regressing the dependent variable $(y_{it} - \bar{y}_i)$ on $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\beta + (\epsilon_{it} - \bar{\epsilon}_i)$. The estimate for σ_μ is found using the estimate of the variance in the “mean” regression

$$\bar{y}_i = \alpha + \beta'\bar{\mathbf{x}}_i + (\mu_i + \bar{\epsilon}_i)$$

and the estimate is given by

$$\hat{\sigma}_\mu^2 = \frac{e_{**}'e_{**}}{n - k} - \frac{\hat{\sigma}_\epsilon^2}{T}$$

2 Hausman and Taylor model

Is it possible that some individual-specific unobservable effects are correlated with some other explanatory variables? Yes! If so, we need to take that into account in the RE model. Hausman and Taylor (1981) proposed the following model.[2]

$$y_{it} = \mathbf{x}'_{1it}\beta_1 + \mathbf{x}'_{2it}\beta_2 + \mathbf{z}'_{1i}\alpha_1 + \mathbf{z}'_{2i}\alpha_2 + \epsilon_{it} + u_i,$$

where

$\mathbf{x}_{1it}$ is K_1 variables that are time varying and uncorrelated with u_i,
$\mathbf{z}_{1i}$ is L_1 variables that are time-invariant and uncorrelated with u_i,
$\mathbf{x}_{2it}$ is K_2 variables that are time varying and are correlated with u_i,
$\mathbf{z}_{2i}$ is L_2 variables that are time-invariant and are correlated with u_i

The assumptions are

$$E[u_i \mid \mathbf{x}_{1it}, \mathbf{z}_{1i}] = 0 \text{ though } E[u_i \mid \mathbf{x}_{2it}, \mathbf{z}_{2i}] \neq 0,$$
$$\text{Var}[u_i \mid \mathbf{x}_{1it}, \mathbf{z}_{1i}, \mathbf{x}_{2it}, \mathbf{z}_{2i}] = \sigma_u^2,$$
$$\text{Cov}[\varepsilon_{it}, u_i \mid \mathbf{x}_{1it}, \mathbf{z}_{1i}, \mathbf{x}_{2it}, \mathbf{z}_{2i}] = 0,$$
$$\text{Var}[\varepsilon_{it} + u_i \mid \mathbf{x}_{1it}, \mathbf{z}_{1i}, \mathbf{x}_{2it}, \mathbf{z}_{2i}] = \sigma^2 = \sigma_\varepsilon^2 + \sigma_u^2,$$
$$\text{Corr}[\varepsilon_{it} + u_i, \varepsilon_{is} + u_i \mid \mathbf{x}_{1it}, \mathbf{z}_{1i}, \mathbf{x}_{2it}, \mathbf{z}_{2i}] = \rho = \sigma_u^2/\sigma^2$$

- OLS and GLS not convergent - some variables are correlated with random effects

- Obtain consistent estimates of β_1 and β_2 using differences from the "temporal" mean - LSDV method

$$(y_{it} - \bar{y}_i) = (x_{1it} - \bar{x}_{1i})'\beta_1 + (x_{2it} - \bar{x}_{2i})'\beta_2 + (\epsilon_{it} - \bar{\epsilon}_i). \quad (7)$$

- We need instruments...
  - $x_{1it} - \bar{x}_{1i}$ and $x_{2it} - \bar{x}_{2i}$ act as instrument that produce unbiased estimates of the β's
  - We do not need instruments for $\mathbf{z}_{1i}$ as it is uncorrelated with u_i
  - $\bar{x}_{1i}$ is a valid instrument for $\mathbf{z}_{2i}$ (Hausman and Taylor)

3 HT - Step-by-Step Estimation

1. Obtain consistent estimates of β_1 and β_2 using differences from the "temporal" mean - LSDV method

$$(y_{it} - \bar{y}_i) = (x_{1it} - \bar{x}_{1i})'\beta_1 + (x_{2it} - \bar{x}_{2i})'\beta_2 + (\epsilon_{it} - \bar{\epsilon}_i). \quad (8)$$

[2] Adapted from Greene [2] and [3] and Hausman and Taylor [1]

2. (a) From step 1, use the residuals to compute the "intra-group" temporal mean of the residuals, $\bar{e}_i = \frac{\sum_{t=1}^{T} e_{it}}{T}$, and stack them into vector $\bar{e}' = \left((\overbrace{\bar{e}_1, \bar{e}_1, ..., \bar{e}_1}^{T}), ..., (\bar{e}_n, \bar{e}_n, ..., \bar{e}_n), \right)$

   (b) Do a regression of $\mathbf{z}_{2i}$, the invariant effects correlated with u_i, on $\mathbf{z}_{1i}$ and $\mathbf{x}_{1it}$.

   (c) Use the predicted values $\hat{\mathbf{z}}_{2i}$ from (b) in the big matrix $\mathbf{Z} = (\mathbf{Z}_1^*, \hat{\mathbf{Z}}_2^*)$, where matrices $\mathbf{Z}_k$ are formed using the $\mathbf{z}_{ki}$ for each group i.

   (d) Regress vector $\bar{e}$ on $\mathbf{Z}$ to get estimates of $(\hat{\alpha}_1, \hat{\alpha}_2)$.

   (e) Note: we just did a 2SLS regression...

3. Estimate of σ_ϵ^2: Use the estimate from the LSDV regression in Step 1
   Estimate of σ_u^2: As in the RE model, use the estimate of σ^{*2} from the 2SLS regression in Step 2. Since

$$\sigma^{*2} = \sigma_u^2 + \frac{\sigma_\epsilon^2}{T}$$

   then an estimate of σ_u is

$$\sigma_u^2 = \sigma^{*2} - \frac{\sigma_\epsilon^2}{T}$$

4. We need weights to compute the FGLS. Let $\hat{\theta} = \sqrt{\frac{\hat{\sigma}_\epsilon^2}{\hat{\sigma}_\epsilon^2 + T\hat{\sigma}_u^2}}$, then, for each group i, let

$$W^* = [\mathbf{x}_{it1}, \mathbf{x}_{it2}, \mathbf{z}_{i1}, \mathbf{z}_{i2}] - \hat{\theta}[\mathbf{x}_{it1}, \mathbf{x}_{it2}, \mathbf{z}_{i1}, \mathbf{z}_{i2}] \quad (9)$$

$$y^* = \mathbf{y}_{it} - \hat{\theta}\mathbf{y}_{it} \quad (10)$$

$$\mathbf{v}'_{it} = [(\mathbf{x}_{1it} - \mathbf{x}_{1i})', (\mathbf{x}_{2it} - \mathbf{x}_{2i})', \mathbf{z}'_{1i}, \bar{\mathbf{x}}'_{1i}] \quad (11)$$

   be the new weighted data and V the matrix of instruments, then do a 2SLS regression of y^* on W^* with instruments V:

   (a) Regress W^* on V, then generate the predicted values $\hat{W}^*$.

   (b) Regress y^* on the predicted values $\hat{W}^*$ to get $(\hat{\beta}', \hat{\alpha}')'$

5. To get the variance of $(\hat{\beta}', \hat{\alpha}')'$, one should not use the residuals of of the 2SLS regression, because it is not convergent. See Greene Ch.8 eq (8.8)

4 HT - other topics

4.1 How to choose which variables are correlated with u_i

1. Specification Testing in Panel Data With Instrumental Variable - Gilbert E. Metcalf (NBER)

Empirical Example for Hausman and Taylor Estimator for Panel Data

Taken from Cornwell, Christopher and Peter Rupert (1988). "Efficient Estimation With Panel Data: An Empirical Comparison of Instrumental Variables" Journal of Applied Econometrics, Vol. 3, No. 2 (April 1988), pp. 149-155.

General Problem

Capturing the real returning of school (on wage) is not an easy task. There are unobserved aspects of **ability** that are not observed, therefore we would like to run a Random Effects estimator model for Panel Data. Nevertheless, there is a strong correlation between the observed person-specific aspects, in this case years of education, and the unobserved ability.

$$ln(wage_{i,t}) = x'_{1i,t}\beta_1 + x'_{2i,t}\beta_2 + z'_{1i,t}\gamma_1 + z'_{2i,t}\gamma_2 + u_i + \epsilon_{i,t}$$

$$x_{1i,t} = \begin{bmatrix} WKS_{i,t} \\ SOUTH_{i,t} \\ SMSA_{i,t} \\ MS_{i,t} \end{bmatrix}, x_{2i,t} = \begin{bmatrix} EXP_{i,t} \\ EXP^2_{i,t} \\ OCC_{i,t} \\ IND_{i,t} \\ UNION_{i,t} \end{bmatrix}, z_{1i,t} = \begin{bmatrix} FEM_i \\ BLK_i \end{bmatrix}, z_{2i,t} = \begin{bmatrix} EDU_i \end{bmatrix}$$

Table 1: Characteristics of Variables.

$X_{i,t}/Z_{i,t}$	**UNcorrelated with** u_i	**Correlated with** u_i
Time-Variant	WKS, SOUTH, SMSA,MS	EXP, EXP2, OCC, IND, UNION
Time-Invariant	FEM, BLK	EDU

List of Variables

1. EXP: Work experience
2. WKS: Weeks worked

3. OCC: Occupation, 1 if blue collar
4. IND: Works in manufacturing industry
5. SOUTH: Resides in south
6. SMSA: Resides in a city (SMSA)
7. MS: Married status
8. FEM: Female
9. UNION: Dummy showing if wage was set by a union in contract
10. EDU: Years of education
11. BLK: Individual is black
12. WAGE: Wage

Procedure

.

1. Fixed Effects Estimator with individual and time dummy variables.
2. Don't use EXP variable for the first step (fixed effects)
3. Add an intercept for the final random effects regression and the instrumental variable regression.

Results

1. Schooling matters more than it was originally observed.
2. If coefficients for within are close to the HT, it means that the use of the instrument variable is legitimate.

Table 2: Regression Results.

VARS	Pooled	FE	RE	Hausman and Taylor IV
INTERCEPT	5.2511236*		5.2210241*	2.9259224
WKS	0.0042160890*	0.00083594602	0.0063332680*	0.0025558388
SOUTH	-0.055637368*	-0.0018611924	-0.056689494*	0.056586813
SMSA	0.15166712*	-0.042469153*	0.16781227*	-0.072097855
MS	0.048448508*	-0.029725839	0.088100813*	-0.43434849*
EXP	0.040104650*	0.11320827*	0.035021236*	0.099274709*
EXP2	-0.00067337705*	-0.00041835132*	-0.00060920201*	-0.00043191471
OCC	-0.14000934*	-0.021476498	-0.15645174*	-0.022457208
IND	0.046788640*	0.019210122	0.054151530*	0.00035122990
UNION	0.092626749*	0.032784860*	0.10007716*	-0.033840603
FEM	-0.36778522*		-0.33967358*	-0.56341823*
BLK	-0.16693763*		-0.16158707*	-0.22374152
EDU	0.056704208*		0.053246951*	0.18481326

References

[1] HAUSMAN, Jerry A. and William E. Taylor, 1981, *Panel Data and Unobservable Individual Effect*, Econometrica, 49(6), 1377-1398.

[2] GREENE, William, 2012, *Econometric Analysis* 7^{th}*edition*, Prentice Hall.

[3] GREENE, William, 2012, *B55.9912: Econometric Analysis of Panel Data - Class Notes*, NYU Stern Business School, online at http://people.stern.nyu.edu/wgreene/Econometrics/PanelDataNotes.htm.