Unobservable individual effects in unbalanced panel data

Robert Gardner*

Department of Agricultural Economics, Michigan State University, 401 Agricultural Hall, East Lansing, MI 48824-1039, USA

Received 29 July 1997; accepted 22 September 1997

Abstract

A modification to the [Hausman, Taylor, 1981. Panel data and unobservable individual effects. Econometrica 49, 1377–1399.] instrumental variables panel data model is warranted with unbalanced panel data. This avoids biasedness when unobservable individual effects are related to explanatory variables without the drawbacks of a fixed effects specification.

Keywords: Panel Data; Unobservable Individual Effects

JEL classification: C33; C51

1. Introduction

A benefit of combining time series and cross sectional data is the ability to account for unobservable individual effects. As shown in the panel data model below Eq. (1), these unobservable effects, \( \mu_i \), are part of the error component, \( u_{it} \). Obviously, there are implications in reducing biasedness should a relationship exist between the explanatory variables and this unobservable effect.

\[
y_{it} = \alpha + X_{it} \beta + Z_i \gamma + u_{it},
\]

where

\[
u_{it} = \mu_i + v_{it}.
\]

A common method of handling this biasedness is to simply remove the individual effects from the error term through demeaning of the data (see Mundlak, 1961). This fixed effects or within-groups estimation procedure is not fully efficient due to lost information contained in the eliminated means. Variation across individuals is ignored and parameter estimation of the time invariant variables, \( Z_i \), is impossible as these variables are eliminated in the demeaning process. There are also implications in panel data estimation of firm efficiency levels (e.g. Schmidt and Sickles, 1984).

Hausman and Taylor (1981) offer a remedy to these problems by providing an estimator that is, in
essence, a hybrid of the Fixed Effects and Generalized Least Squares (Random Effects) models. They suggest a two stage generalized least squares model with instrument set: $[Q_Y X_1, Q_Y X_2, P_Y X_1, Z_1]$; where $Q_Y$ transforms a vector of observations into a vector of deviations from group means and $P_Y$ produces a vector of group means. The variables $X_1, X_2, Z_1,$ and $Z_2$ are derived from variables $x_{it}$ and $z_{it}$ from Eq. (1). $X_1$ and $Z_1$ are not correlated with the individual firm effects while $X_2$ and $Z_2$ are. This technique allows further information to be had, without incurring a biased regression, from the appropriate data means, $X_1,$ and time invariant regressors, $Z_1,$ that are added to the model as instruments.

2. The modified model

In the balanced case, $\text{cov}(u_{it}|X_{it}, Z_{it}) = \Omega = \sigma^2_{\mu}I_T + T\sigma^2_{\mu}P_Y$ is the familiar block-diagonal matrix elicited from Eq. (1). The Gauss–Markov estimator, $\theta,$ is the minimum variance matrix–weighted average of the within and between group estimators as shown in Eq. (2).

$$\theta = \sqrt{\frac{\sigma^2_{\mu}}{\sigma^2_{\mu} + T\sigma^2_{\mu}}}. \quad (2)$$

With a little manipulation,

$$\Omega = \sigma^2_{\mu} \left( Q_Y + \frac{1}{\theta^2} P_Y \right)$$

becomes: $\Omega^{-1/2} = \frac{1}{\sigma_{\mu}} (Q_Y + \theta P_Y). \quad (3)$

As Hausman (1978) indicates, transforming Eq. (1) by $\Omega^{-1/2},$ needed to institute the generalized two staged least squares model, is equivalent to simple differencing of the observations:

$$\Omega^{-1/2} y = \Omega^{-1/2} X_{it} \beta + \Omega^{-1/2} Z_{it} \gamma + \Omega^{-1/2} u_{it}, \quad (4)$$

which, by substitution with Eq. (3), becomes:

$$(Q_Y + \theta P_Y) y_{it} = (y_{it} - \bar{y}) + \theta \bar{y} = y_{it} - (1 - \theta) \bar{y}, \quad (5)$$

with $X_{it}$ and $u_{it}$ being transformed similarly and $Z_{it}$ becoming $\theta Z_{it}$ because of its time invariance.

Again, a two stage generalized least squares model is sought; however, the instrument set should be $\{Q_Y X_1, Q_Y X_2, \theta P_Y X_1, \theta Z_1\};$ i.e. $\theta$-weighted means are the instruments needed to reproduce the $\theta$-weighted means of the regressors. With balanced panel data, $\theta$ is constant and omission of this weighting is inconsequential.

In the unbalanced case, $\theta$ must be estimated differently. For clarity, $Q_Y$ and $P_Y$ are expressed in matrix notation (Eqs. (6) and (7), respectively) to best reflect the unbalanced nature of the data set. Note that since $P_Y Q_Y = 0, P_Y Q_Y = Q_Y P_Y = 0.$

$$Q_Y = \begin{bmatrix} Q_1 & 0 & \cdots & 0 \\ 0 & Q_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Q_n \end{bmatrix} \text{ where: } Q_Y y_i = \begin{bmatrix} y_{i1} - \bar{y}_i \\ \vdots \\ y_{ni} - \bar{y}_i \end{bmatrix}. \quad (6)$$
\[
\begin{bmatrix}
P_1 & 0 & \cdots & 0 \\
0 & P_2 & \cdots & 0 \\
\vdots & & \ddots & \vdots \\
0 & 0 & \cdots & P_n
\end{bmatrix}
\text{where: } P_i y_i = \begin{bmatrix}
\bar{y}_i \\
\vdots \\
\bar{y}_{m_i}
\end{bmatrix}.
\]
(7)

\[\Omega = \begin{bmatrix}
\Omega_1 & \cdots & 0 \\
\Omega_2 & \ddots & \vdots \\
0 & \cdots & \Omega_n
\end{bmatrix}
\text{where: } \Omega_i = \sigma^2_v I_i + T_i \sigma^2_{\mu} P_i.
\]
(8)

Consequently, the Gauss–Markov estimator, \( \theta \) from Eq. (2) will now be expressed as \( \theta_i \), again varying between firms depending on the individual number of observations.

\[\theta_i = \sqrt{\frac{\sigma^2_v}{\sigma^2_v + T_i \sigma^2_{\mu}}}.
\]
(9)

Therefore, the model will be transformed on a firm by firm basis:

\[\Omega^{-1/2} y_{it} = y_{it} - (1 - \theta_i) \bar{y}_i = (y_{it} - \bar{y}) + \theta_i \bar{y}_i,
\]
(10)

with \( X_t \) and \( u_{it} \) transformed similarly and \( Z_t \) becoming \( \theta_i Z_t \) because of its time invariance. Note that \( Q \Sigma P \theta_i \) still equals 0.

As in the balanced case, a two stage generalized least squares estimator is sought but with the appropriate set of instruments being \( \{Q \Sigma X1, Q \Sigma X2, \theta_i P \Sigma X1, \theta_i Z1\} \); i.e. \( \theta_i \)-weighted means are used for instruments to reproduce the \( \theta_i \)-weighted means of the regressors. In the unbalanced panel case, where \( \theta_i \) is not constant across firms, this firm specific weighting is necessary to elicit an accurate model.

3. Summary

In this paper, I have proposed a modification to the Hausman–Taylor (1981) Two Stage Least Square Model that allows efficient estimation when an endogeneity problem exists without the drawbacks associated with a fixed effects model. This modification is necessary in order to institute the technique with unbalanced panel data.

Acknowledgements

Peter Schmidt, Department of Economics at Michigan State University, made significant comments though responsibility for any error remains with the author.
References