Notes on Panel Estimation by IV

Thomas F. Crossley and Vincent Hildebrand *

May 2000

Abstract

Notes prepared for the Summer Programme in Data Analysis at the Institute for Social Research, York University, Toronto Canada, June 5 to 17, 2000

Contents

*Crossley: Department of Economics and Institute for Social Research, York University, Canada. Hildebrand: Department of Economics, York University, Canada and IRISS/CEPS-INSTEAD.

1 Introduction

Panel, or Longitudinal data allows us to address two important features of reasonable models:

Heterogeniety

Dynamics

The simplest model that incorporates the first of these is the linear "fixed effect" model.

$$y_{it} = x_{it}\beta + \gamma_i + \varepsilon_{it}$$

The usual problem is that γ_i is correlated with the x_{it}.There are is a large literature on the estimation of such models. Many students would be familar with the simple "within" (or ANACOVA?) estimator for such models. In this note, we emphasize that estimators (such as the "within" estimator) which are based on data transformations can be thought of as Instrumental Variables (IV) estimators. This insight in turn provides an accessible way to approach the more advanced literature in the area.

2 Difference Estimators as IV

A common way to deal with an model such as (1) with panel data is to transform the data to eliminate the unobserved but time invariant heterogeniety γ_i. For example, we might estimate the model (where the individual subscript i has been surpressed):

$$y_t - \overline{y} = (x_t - \overline{x})\beta + (\varepsilon_t - \overline{\varepsilon})$$

This is typically referred to as the "within" estimator. It turns out that this is just an application of IV. To see this, consider a matrix of individual dummies D and the associated projection matrix $P_D = D(D'D)^{-1}D$. The individual means of x and y are then $\overline{x} = P_D x$ and $\overline{y} = P_D y$, and the deviations from means are just $x - \overline{x} = M_D x$ and $y - \overline{y} = M_D y$. So, we can write the within estimator as:

$$b^w = (x'M_D M_D x)^{-1}(x'M_D M_D y) \tag{1}$$

Note that the idempotency of projection matrices means that:

$$b^w = (x'M_D x)^{-1}(x'M_D y) \tag{2}$$

The above is what one would get if one regressed the level y on the transformation $M_D x$. This illustrates that the transformation of y is redundant. (Note, however, that the transformation of x is necessary, even if one has transformed y).[1]

Finally, note that the linear prediction of x given $M_D x$ (the projection of x onto $M_D x$) is just $M_D x$.

$$M_D x(x'M_D M_D x)^{-1}x'M_D x = M_D x \tag{3}$$

Thus the regression of y on $M_D x$ is the regression of y on the prediction of x given $M_D x$, or equivalently, the instrumental variables estimate of (1)

[1] Oksanen () refers to the regression of y on $M_D x$ as "single difference regression" and the regression of $M_D y$ on $M_D x$ as "double difference regression".

where $M_D x$ is used as the instrument. Thus the within estimator is just an IV estimator.

This point is made in a number of places, including Moffit (1993). We have made it a round-about way, but hopefully in a way by which most students would find most of the steps somewhat familar. Once one recognizes that the the within estimator is an IV estimator, with the instrument generated by a transformation of x, then some of the more advanced literature is easy to approach. Recall that the 2SLS framework we started with accomodates multiple instruments. Some of the advanced literature concerns combining multiple instruments to estimate (1), where the additional instruments are derived from alternative transformations of x, or from other kinds of exclusion restrictions.

In addition, once one recognizes that many linear panel techniques are simply IV estimators, with instruments generated by transformations of x, it is possible to implement such techniques easily in canned econometrics packages such as STATA. The idea is simply to manipulate the data to generate the desired instruments, and then use a 2SLS routine. Examples are given below.

3 Additional Instruments.

3.1 Identifying Coefficients on Time Invariant Regressors (Hausman-Taylor).

One problem with the within estimator, and other estimators where the instruments are derived from transformations of x is that one cannot estimate the coefficients on time invariant regressors. Suppose that x contains two time varying regressors and one time invariant regressor. If one transforms the right hand side variables with the "within" transformations or any other transformation that removes the time invariant individual heterogeneity for (1), one generates two instruments - corresponding to the two time varying regressors. (The time invariant regressor does not generate an instrument because the transformation returns a vector of zeros.) Thus we can only estimate two coefficients - the coefficients on the time varying variables.

Hausman and Taylor (1991) suggested that one could estimate the coefficients on (some of) the time invariant variables if one could find additional instruments. In particular, suppose you were prepared to assume that one of the time varying variables (say, x_1) was a valid instrument. That is, it is uncorrelated with γ_i or excluded from () - so we can think of these instruments as being generated by exclusion restricitons. Then we would have three instruments (the transformation of x_1 and x_2, and the level of x_1) and could estimate three parameters.

There is a large literature following Hausman and Taylor (1991) in which various authors refine their original suggestion. Many of the refinements concern the optimal way to combine the different estimators. The most important paper is Arrelano and Boyer ().

3.2 The Chamberlain Framework

Chamberlain (1982, 1984) proposed a useful way of thinking about the fixed effects model, as follows. For $T = 2$ we can write:

$$y_{i1} = x_{i1}\beta + \gamma_i + \varepsilon_{i1}$$

and

$$y_{i2} = x_{i2}\beta + \gamma_i + \varepsilon_{i2}$$

That is, we can treat the repeated observations on y as a system of equations. The usual problem is that γ_i is correlated with the x_{it}. Chamberlain suggested capturing this correlation with the linear projection:

$$\gamma_i = \lambda_1 x_{i1} + \lambda_2 x_{i2} + ... + \lambda_1 x_{iT} + \upsilon_i$$

With $T = 2$:

$$\gamma_i = \lambda_1 x_{i1} + \lambda_2 x_{i2} + \upsilon_i$$

By substition:

$$y_{i1} = x_{i1}\Pi_{11} + x_{i2}\Pi_{12} + \nu_i + \varepsilon_{i1}$$

and

$$y_{i2} = x_{i2}\Pi_{21} + x_{i2}\Pi_{22} + \nu_i + \varepsilon_{i1}$$

Where

$$\Pi_{11} = \beta + \lambda_1$$
$$\Pi_{12} = \lambda_2$$
$$\Pi_{21} = \lambda_1$$
$$\Pi_{22} = \beta + \lambda_2$$

Since v_i is by construction uncorrelated with the x_{it}, this effectively tranforms the fixed effect model into a random effects model which can be consistently estimated by OLS (SUR, but RHS variables all the same).

β can be recovered by indirect least squares, but note that β is overidentified:

$$\beta = \Pi_{11} - \Pi_{21} = \Pi_{22} - \Pi_{12}$$

so that an optimal estimate of β can be obtained via a "mininum distance" (or "minimum Chi-square") step:

$$\widehat{\beta} = \underset{\beta, \lambda_1, \lambda_2}{\arg\min} \begin{pmatrix} \widehat{\Pi_{11}} - \beta - \lambda_1 \\ \widehat{\Pi_{12}} - \lambda_2 \\ \widehat{\Pi_{21}} - \lambda_1 \\ \widehat{\Pi_{22}} - \beta - \lambda_2 \end{pmatrix}' \widehat{V}\left(\widehat{\Pi}\right)^{-1} \begin{pmatrix} \widehat{\Pi_{11}} - \beta - \lambda_1 \\ \widehat{\Pi_{12}} - \lambda_2 \\ \widehat{\Pi_{21}} - \lambda_1 \\ \widehat{\Pi_{22}} - \beta - \lambda_2 \end{pmatrix}$$

The minimized value of this quadratic form is Chi-square distributed and provides an "omnibus" test of the fixed effects specification. For applications to estimating the returns to educations and the union wage premium see Angrist and Newey (1991) and Jakubson (1991), respectively.

Another advantage of this framework is that it extends easily to some limited dependent variable models ("index models") where:

$$y_{i1}^* = x_{i1}\beta + \gamma_i + \varepsilon_{i1}$$

and

$$\begin{array}{lll} y = 1 & \text{if} & y^* > 0 \\ y = 0 & \text{if} & y^* \leq 0 \end{array}$$

(probit or logit model) or

$$\begin{array}{lll} y = y^* & \text{if} & y^* > 0 \\ y = 0 & \text{if} & y^* \leq 0 \end{array}$$

(tobit model). As in the linear case one first estimates a random effects "reduced form" probit or tobit (for consistency of random effects probit and tobit see ......) and then recovers the "structural parameters by minimum distance. For further details see Chamberlain, 1984 and Bover and Arellano, 1997.

It is useful to think about where the "overidentification" of β comes from. Note that in the case of $T = 2$ we have 2 equations (() and ()), one parameter of interest (β) and one independent instrument, $M_D x$. (For $T = 2$, $(x_2 - \overline{x}) = \frac{1}{2}(x_2 - x_1) = -(x_1 - \overline{x})$). Thus we have two "moment" conditions:

$$\begin{array}{l} E\left[x' M_D(\gamma_i + \varepsilon_{i1})\right] = 0 \\ E\left[x' M_D(\gamma_i + \varepsilon_{i2})\right] = 0 \end{array}$$

which lead to the two estimates of β. We could equivalently think of pooling the data in a single equation

$$y_{it} = x_{it}\beta + \gamma_i + \varepsilon_{it}$$

and "picking out" the first and second period in terms by multiplying through by time dummies (t_1, t_2). Thus we can think of the two moment equations as being generated a single equation and two instruments:

$$\begin{array}{l} E\left[x' M_D t_1(\gamma_i + \varepsilon_{it})\right] = 0 \\ E\left[x' M_D t_2(\gamma_i + \varepsilon_{it})\right] = 0 \end{array}$$

where the two instruments are generated by interacting $M_D x$ with the time dummies (t_1, t_2). Notice that functions of t are uncorrelated with γ_i and so are valid instruments. As we shall discuss below in section (), this turns out to be very useful when $M_D x$ is not available (that is, when we do not have panel data).

4 QuasiPanels from Repeated Cross Sections.

Panel data is not available for every application, nor for every country. Starting with Browning, Deaton and Irish (1985) have used "quasipanel" data constructed from repeated cross sections to estimate models such as (1). Browning, Deaton and Irish took cohort year means of their data and then ran a regression of y on x and cohort dummies using this cohort data. This is equivalent to a within estimator on aggregate (cohort) data. It also has an IV interpretation, as is discussed by Moffit (1993) and others.

Browning, Deaton and Irish wanted to estimate 'Frisch' or intertemporal life cycle consistent labour supply and commodity demand functions:

$$h_{it} = \alpha_t + \beta \ln w_{it} - \beta \ln r_i$$

(dropping cross price effects) and similarly for consumption. Good panel data was unavailable (especially for consumption). They employed stacked cross sections of the Family Expenditure Survey (UK, 'FES') for the seven years from 1970 to 1976. They identified birth year cohorts. Individuals that were 20 in 1970 were 21 in 1971 and so. The key insight was that although individuals could not be followed through time, birth cohorts could. BDI defined 5 year birth cohorts, split by manual/nonmanual occupation.

Since each cross section is an independent (random) sample of the population, cohort characteristics can be consistently estimated in each year. The key assumpiton is constant cohort composition.

BDI took within cohort x year means. They had 8 birth year bands, x manual/nonmanual, and so 16 cohorts, which they followed for 7 years giving 102 "cells" or cohort-year observations. Most cells contained more than 200 observations. Taking means through (1):

$$\overline{h_{ct}} = \alpha_t + \beta\overline{\ln w_{ct}} - \beta \ln r_c$$

and then first differencing (or taking deviations from means):

$$\Delta\overline{h_{ct}} = \Delta\alpha_t + \beta\overline{\Delta \ln w_{ct}}$$

To implement this population relationship they regressed sample 'cell' means of hours on sample 'cell'means of log wages plus cohort dummies plus time dummies. Equivalently one could implement a "within" estimator (with time dummies) on the cohort data.

4.1 Quasipanel techniques are IV estimators

The BDI procedure is an IV estimor. The cohort x year 'cell' means are the predicted values from a regression of wages (hours) on cohort dummies, time dummies, and a full set of interactions. This is the 'auxilliary' regression.

$$\ln w_{it} = D_c\Pi_1 + D_t\Pi_2 + (D_c \times D_t)\Pi_3 + e_{it}$$

The regression of interest has cohort and time effects:

$$h_{it} = \alpha_t + \beta \ln w_{it} - \beta(\ln r_c + (\ln r_i - \ln r_c))$$

So the 'excluded' variables are the cohort x time interactions. Effectively, BDI use time x birth year x occupation interactions as instruments for wages.

The connection between quasipanel techniques and IV is drawn out formally by Moffit (J.Econometric, 1993).

- $x'M_D$ is unavailable as an instrument.
- functions of time are valid instruments, but may want to allow for time effects
- some time invariant variables may be valid
- interactions may be suitable instruments
- in general 'grouping' estimators are IV estimators.

4.2 Measurement error problem

We have sample estimates of the cohort means in each year, so that the cohort means are measured with error.

$$\widehat{h_{ct}} = \overline{h_{ct}} + \varepsilon_{ct}$$
$$\widehat{\ln w_{ct}} = \overline{\ln w_{ct}} + \eta_{ct}$$

$$\overline{h_{ct}} = \alpha_t + \beta \overline{\ln w_{ct}} - \beta \ln r_c$$

$$\widehat{h_{ct}} = \alpha_t + \beta \widehat{\ln w_{ct}} - \beta \ln r_c - \beta \eta_{ct} - \varepsilon_{ct}$$

Differencing does not correct this.

$$\Delta\widehat{h_{ct}} = \Delta\alpha_t + \beta(\widehat{\ln w_{ct}} - \widehat{\ln w_{ct-1}}) - \beta(\eta_{ct} - \eta_{ct-1}) - (\varepsilon_{ct} - \varepsilon_{ct-1})$$

There are three possible solutions:

1. Once could instrument $(\widehat{\ln w_{ct}} - \widehat{\ln w_{ct-1}})$ with $(\widehat{\ln w_{ct-2}} - \widehat{\ln w_{ct-3}})$.

2. Deaton (J.Econometrics, 1985) notes that one can implement a measurement error model. To see this, recall (from any standard econometrics text book) the case of a simple linear regression ($y_i = x_i^*\beta + \varepsilon_i$) with the idependant variable measured with classical measurement error ($x_i = x_i^* + u_i$). OLS estimates of β will be inconsistent.

$$plim \frac{\sum xy}{\sum x^2} = \beta \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2}$$

   But in the quasi-panel case the variables are mean and measurement error is sampling variability of the mean. So $\sigma_u^2 = (\frac{\sigma_x^2}{n})_{ct}$. We can estimate this from the microdata (just as we estimate the means) and "correct" our estimate of β.(See also Verbeek and Nijman, J.Econometrics 93).

3. The most common solution is to do nothing, except to make sure that n_{ct} is large. As the second solution makes clear, the importance of the sampling variability decreases with cell size. Note however that if $n_{ct} \times C \times T$ is fixed, increasing n_{ct} - by widening the year bands for example - must decrease C or T . Thus there is an effective tradeoff between the precision of the means and the number of effective observations. A rule of thumb that is sometimes given (See Browning and Lusardi?) is that cells should

be greater than 100 but there is little gain to increases beyond 400. (A related issue: Asymptotics in CT?, C? T?

4.3 Advantages and problems of Quasipanels.

Advantages

- aggregation may average out some measurement error.
- fresh smaples mean no problem with panel attrition.
- sometimes its all we have (cross sections are cheap/ some kinds of information collection are only feasible in cross section)
- sometimes we need really long panels.

Problems

- cohort composition changes (death/immigration and emmigration/household formation).
- Loss of individual level variation.

The latter is equivalent to saying that the "instruments are weak". (Example).

4.4 Richer Models.

- Limited Dependent Variables: A variant of the Chamberlain/ Bover-Arellano approach is feasible under certain conditions (Collado, Investigaciones Economicas, 1988).
- Dynamics, Lagged Dependent Variables: (Collado, J.Econometrics, 1997).

- Dynamics, Markov Models, Duration Models: Moffit (J.Econometrics, 1993) presents a Markov model of labour force participation. Heisz and Walsh, (Statcan 2000) present survivor function estimates of job duration.

  Especially for the third of the above one can allow very limited or no heterogeniety. (Heterogeniety and Dynamics are always hard, and without true panel data the deck is really stacked).

4.5 Additional Comments

- common applications are life cycle models of labour supply, consumption, savings and wealth, and immigrant assimilation.
- age-cohort-time problem.
- respeated short panels.
- micro data is essential (measurement error model, and also, $\overline{\ln w_{ct}} \neq \ln \overline{w_{ct}}$!)

5 Miscellaneous Comments.

Strict Versus Weak Exogeniety.

Extensions to Nonlinear Models.

Extension to Dynamic Models.

7 References

Amemiya, T. and T. MacCurdy, (1986), "Instrumental variables estimation of an error components model," *Econometrica*, **54** , 869-880.

Angrist and Newey, (1991), "" *Journal of Business and Economic Statistics*, .

Arellano, M. and O. Bover, (1995), "Another look at the instrumental variable estimation of error-components models," *Journal of Econometrics*, **68**, 29-51.

Baltagi, B. and S. Khanti-Akom, (1990), On efficent estimation with panel data: An empirical comparison of instrumental variables estimator, " *Journal of Applied Econometrics*, **5**, 401-406.

Breusch, T.S., Mizon, G.E. and P. Schmidt (1989), "Efficient Estimation using Panel Data," *Econometrica*, **57**(3), 695-700.

Browning, M., Deaton, A., and Irish, M., 1985, "A Profitable Approach to Labour Supply and Commodity Demands over the Life-Cycle," *Econometrica*, **53**, 503- 544.

Collado, D., 1992, "Estimating Dynamic Models from Time Series of Independent Cross-Sections," *Journal of Econometric*, **82**, 37-62.

Cornwell, C., and P. Rupert 1988, "Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variables Estimators," *Journal of Applied Econometrics*, **3**, 149-155.

Deaton, A., 1985, "Panel Data from Times Series of Cross-sections," *Journal of Econometrics*, **30**, 109-126.

Grilishes, Z., and Hausman, J., "Errors in Variables in Panel Data," *Journal of Econometrics*, **31**, 93-118.

Hausman, J., and Taylor, W., 1981, "Panel Data and Unobservable Individual Effects," *Econometrica*, **49**, 1377-1398.

Jakubson, (1991), "" *Review of Economic Studies*

Meghir, C., and Whitehouse, E., 1996, "The Evolution of Wage in the United Kingdom: Evidence from Micro Data," *Journal of Labor Economics*, **14**, 1-25.

Moffitt, R., 1993, "Identification and Estimation of Dynamic Models with a Time Series of Repeated Cross-Sections," *Journal of Econometrics*, **59**, 99-123.

Solon, G., 1989, "The Value of Panel Data in Economic Research," in *Panel Surveys*, eds. Daniel Kasprzyk, Greg Duncan, Graham Kalton, and M.P. Singh, John Wiley & Sons, 486-469.

Verbeek, M., and Nijman, T., 1992, "Can Cohort Data be Treated as Genuine Panel Data?," *Empirical Economics*, **17**, 9-23.

Verbeek, M., "Pseudo Panel Data", 1992, in *The Econometric of Panel Data-Handbook of Theory and Application*, edited by Matyas, L., and Sevestre, P., Kluwer Academic Publishers, 303-315.

Verbeek, M., and Nijman, T., 1993, "Minimum MSE estimation of a regression model with fixed effects from a series of cross-sections," *Journal of Econometrics*, **59**, 125-136.

B The Within Estimator is an IV estimator (no matrices).

$$\beta^{within} = \frac{\sum_i \sum_t (x_{it} - \overline{x_i})(y_{it} - \overline{y_i})}{\sum_i \sum_t (x_{it} - \overline{x_i})^2}$$

$$= \frac{\sum_i \sum_t y_{it}(x_{it} - \overline{x_i}) - \sum_i \sum_t \overline{y_i}(x_{it} - \overline{x_i})}{\sum_i \sum_t x_{it}(x_{it} - \overline{x_i}) - \sum_i \sum_t \overline{x_i}(x_{it} - \overline{x_i})}$$

$$= \frac{\sum_i \sum_t y_{it}(x_{it} - \overline{x_i}) - \sum_i \overline{y_i} \sum_t (x_{it} - \overline{x_i})}{\sum_i \sum_t x_{it}(x_{it} - \overline{x_i}) - \sum_i \overline{x_i} \sum_t (x_{it} - \overline{x_i})}$$

$$= \frac{\sum_i \sum_t y_{it}(x_{it} - \overline{x_i}) - 0}{\sum_i \sum_t x_{it}(x_{it} - \overline{x_i}) - 0}$$

$$= \frac{\sum_i \sum_t y_{it} z_{it}}{\sum_i \sum_t x_{it} z_{it}} = \beta^{IV}$$