Estimating Dynamic Models from Repeated Cross-Sections*

Marno Verbeek
Erasmus University Rotterdam

Francis Vella
European University Institute

Revision January 31, 2002

Econometric Institute Report EI 2002-05

Abstract

An important feature of panel data is that it allows the estimation of parameters characterizing dynamics from individual level data. Several authors argue that such parameters can also be identified from repeated cross-section data and present estimators to do so. This paper reviews the identification conditions underlying these estimators. As grouping data to obtain a pseudo-panel is an application of instrumental variables (IV), identification requires that standard IV conditions are met. This paper explicitly discuss the implications of these conditions for empirical analyses. We also propose a computationally attractive instrumental variables estimator that is consistent under a relatively weak set of conditions. A Monte Carlo study indicates that this estimator may work well in practice.

*Correspondence: Marno Verbeek, Erasmus University Rotterdam, Dept. of Financial Management and Econometric Institute, P.O.Box 1738, 3000 DR Rotterdam, The Netherlands. E-mail: M.Verbeek@fbk.eur.nl. Helpful comments and suggestions from two anonymous referees, and from seminar participants at Tilburg University, CORE, and the International Conference on the Analysis of Repeated Cross-Sectional Surveys (Nijmegen, 2000) are gratefully acknowledged. JEL-codes: C21, C23, C81.

1 Introduction

An important feature of panel data is its facilitation of the estimation of parameters capturing the dynamic relationships in individual data. Recently, Moffitt (1993), Collado (1998), Girma (2000) and McKenzie (2001) have argued that the same parameters can also be estimated via repeated cross-section (RCS) data. However, although these papers are concerned with essentially the same model, namely the first order autoregressive model with exogenous variables, the proposed estimators and their presentations are very different.[1] This makes it difficult to compare these procedures and their underlying assumptions. While there are differences across estimators the approach employed is either explicitly, or implicitly, instrumental variables (IV). The estimators first aggregate the individual data into cohorts comprising individuals with similar observed characteristics (e.g. year of birth). Using this pseudo panel the lagged dependent variable is then replaced by a predicted value from an auxiliary regression and the dynamic model is subsequently estimated via IV.

As the ability to estimate dynamics from RCS data is useful, and the proposed estimators are all quite different, it is important to understand how they are related and under which conditions one might be employed in favor of the others. Accordingly, this paper reviews the identification conditions for consistent estimation of a linear dynamic model from RCS data and puts the respective assumptions of these different procedures into perspective. One of our conclusions from doing so is that it appears that these estimators are either too simplistic or too elaborate to be useful in realistic settings. Moreover, we are able to propose fairly simple alternatives that are preferable under comparable conditions.

We begin our analysis with an examination of the estimator in Moffitt (1993). In this procedure the dynamic relationship is estimated via ordinary least squares (OLS) where the lagged dependent variable is replaced by a predicted value. We show that the resulting estimator is inconsistent unless strong, and often unrealistic, conditions are imposed on the exogenous variables. This is true even if the instruments used to predict the lagged dependent variable satisfy the standard IV requirements. We propose that this inconsistency can be overcome by instrumenting the other explanatory

[1] The models considered by Girma (2000) and McKenzie (2001) differ as they allow for heterogeneity of the parameters across cohorts.

variables.

When the instruments are correlated with the unobservables these estimators are inconsistent. However, if one is willing to assume that the correlation between the unobservables and the (time-invariant) instruments is time-invariant, one can correct for this by including the instruments, with time-invariant coefficients, in the model. When the instruments comprise a group of dummies corresponding to mutually exclusive cohorts, the resulting estimator is simply the standard within estimator applied to a dynamic model in terms of cohort averages rather than individual observations. While this estimator is often used in empirical work for both the dynamic and static model (see, for example, Dargay and Vythoulkas, 1999), it has received surprisingly little attention in the theoretical literature.

Although the conditions required for consistency of the proposed estimators are relatively weak, they are not trivially satisfied. Estimation via grouping is an IV technique and, accordingly, the grouping variables should satisfy the appropriate conditions, including a rank condition, for an IV estimator to be consistent. This requires that the chosen set of cohort dummies is appropriately instrumenting each of the explanatory variables in the model. For the within estimator at the cohort level, this requires that all explanatory variables have time-varying cohort effects which, in addition, are not collinear.

The literature on estimation from RCS data is characterized by a range of alternative asymptotics (compare Verbeek, 1996, McKenzie, 2001). Consistency in this paper will refer to large N and small T. That is, the number of individuals in each cross-section is assumed to be large, while the number of cross-sections is fixed. An additional important issue is how the number of cohorts is treated as the number of individuals increases. In many applications, cohorts are defined on the basis of year of birth, in which case it seems natural to treat their number as fixed (compare Moffitt, 1993, Girma, 2000, and McKenzie, 2001). Such an approach has two advantages. First, it simplifies the analysis. Second, and more importantly, it is difficult to think of appropriate sequences of instruments that grow with the number of individuals in a sample (compare McKenzie, 2001), unless the data generating process is formulated at the cohort level. This, however, seems unnatural. For example, it seems inappropriate to assume that the distribution of cohort-specific means is unaffected by the level of aggregation that a researcher employs in constructing cohorts.

The remainder of this paper is organized as follows. Section 2 presents the model of interest and shows that OLS using the pseudo panel with a predicted lagged dependent variable will generally lead to inconsistent estimators. Section 3 discusses some potentially suitable IV estimators, including those suggested by Moffitt, Collado and Girma, and critically reviews the necessary conditions for consistency. Sections 4 and 5 provide an illustrative example and a Monte Carlo experiment, respectively, while Section 6 concludes.

2 Model and Data

The observed data consist of T independent cross-sections at different points in time with each being a random sample of some underlying population. As individuals in different periods are not the same people, we follow the convention in this literature and index the variables by a double subscript. The t denotes the cross-section, while $i(t)$, $1, ..., N_t$ indexes the individuals in cross-section t. The model is:

$$y_{i(t),t} = \alpha y_{i(t),t-1} + x'_{i(t),t}\beta + \varepsilon_{i(t),t},\ \ t = 1, ..., T;\ \ i(t) = 1, ..., N_t, \qquad (1)$$

where the K-dimensional vector $x_{i(t),t}$ may include an intercept, and time-invariant and time-varying variables. Here, $y_{i(t),t-1}$ refers to the value of y at $t-1$ for an individual observed in cross-section t noting that it is unobserved in the available data. We assume the error term has the following properties:

$$E\{\varepsilon_{i(t),t}x_{i(t),t}\} = 0,\ t = 1, ..., T. \qquad (2)$$

As the lagged values for y are not observed for any individual, we construct an estimate by using information on the y-values of other individuals observed at $t-1$. To do so, let $z_{i(t)}$ denote a set of time-invariant variables, including an intercept term.[2] Now, consider the orthogonal projection in cross-section t of $y_{i(t),t}$ upon $z_{i(t)}$,

$$E^*\{y_{i(t),t}|z_{i(t)}\} = z_{i(t)}\delta_t,\ \ t = 1, ..., T, \qquad (3)$$

where E^* denotes the orthogonal projection (for a given t). This "reduced form" can be estimated consistently (for $N_t \to \infty$) by regressing $y_{i(t),t}$ on $z_{i(t)}$

[2] Actually, the requirement that $z_{i(t)}$ is time-invariant is unnecessary. Time-varying variables that can be backcasted with reasonable accuracy may also be included (for example, number of children). For the sake of notation, we abstract from this.

to give $\hat{\delta}_t$. Following Moffitt (1993), one obtains an estimate of $y_{i(t),t-1}$ as the predicted value from this regression, substituting the appropriate z values for the individuals in cross-section t. That is,

$$\hat{y}_{i(t),t-1} = z_{i(t)}\hat{\delta}_{t-1}, \tag{4}$$

noting that $\hat{\delta}_{t-1}$ is estimated from data on different individuals than those indexed by $i(t)$. For later reference, we also define the population equivalent of this as $y^*_{i(t),t-1} = z_{i(t)}\delta_{t-1}$. In the sequel, we shall assume that asymptotics are such that $\text{plim}\,\hat{\delta}_{t-1} = \delta_{t-1}$. In many circumstances it is convenient to think of $z_{i(t)}$ as a vector of dummy variables, corresponding to mutually exclusive cohorts (see Deaton, 1985, Collado, 1998, Girma, 2000). For example, with RCS data, it is common to use year-of-birth cohorts (see, e.g., Blundell, Browning, Meghir, 1994, or Alessie, Devereux and Weber, 1997). In this case, the orthogonal projection in (3) corresponds to the conditional expectation and (4) corresponds to taking sample averages within person i's cohort. Note that a difference between the approaches of Moffitt (1993) and Girma (2000) is that rather than using the average within a cohort, Girma (2000) suggest using the values of one or more arbitrarily chosen persons within the cohort. We discuss the significance, and merits, of this variation below.

Now, insert these predicted values into the original model to get:

$$y_{i(t),t} = \alpha\hat{y}_{i(t),t-1} + x'_{i(t),t}\beta + \varepsilon^*_{i(t),t},\ t = 1,...,T;\ i(t) = 1,...,N_t, \tag{5}$$

where

$$\varepsilon^*_{i(t),t} = \varepsilon_{i(t),t} + \alpha(y_{i(t),t-1} - \hat{y}_{i(t),t-1}). \tag{6}$$

No matter how $\hat{y}_{i(t),t-1}$ is generated, its inclusion implies that one of the explanatory variables is measured with error, although the measurement error will be (asymptotically) uncorrelated with the predicted value.[3] Given assumption (2) and the result that prediction error and predictor are orthogonal, consistency of OLS applied to this equation requires that $\varepsilon_{i(t),t}$ is uncorrelated[4] with $y^*_{i(t),t-1}$ as well as that the prediction error $y_{i(t),t-1} - y^*_{i(t),t-1}$ is uncorrelated with $x_{i(t),t}$. However, both assumptions may be problematic. While the assumption that $\varepsilon_{i(t),t}$ is uncorrelated with $y^*_{i(t),t-1}$ can be defended by the usual IV assumption that

$$E\{\varepsilon_{i(t),t}z_{i(t)}\} = 0,\ t = 1,...,T, \tag{7}$$

[3] Unlike the standard textbook measurement error examples.

[4] Recall that $\hat{y}_{i(t),t-1}$ is the sample counterpart of $y^*_{i(t),t-1}$.

it excludes the possibility that there are "cohort effects" in the unobservables. While this may appear unreasonable, and too restrictive for empirical analyses, this assumption is made in Moffitt (1993) and Girma (2000).[5] Second, it is inappropriate to argue that $x_{i(t),t}$ is uncorrelated with the "prediction error" $y_{i(t),t-1} - y^*_{i(t),t-1}$. Consider, for example, where high x-values in one period on average correspond with high x-values in the next period. If the β coefficients are positive this will generally imply that a high value for $x_{i(t),t-1}$, which is unobservable, will result in an underprediction of $y_{i(t),t-1}$. On the other hand, $x_{i(t),t-1}$ is positively correlated with $x_{i(t),t}$. Consequently, this will produce a positive correlation between $\varepsilon^*_{i(t),t}$ and $x_{i(t),t}$, resulting in an inconsistent estimator for β. This inconsistency carries over to α unless $y^*_{i(t),t-1}$ is uncorrelated with $x_{i(t),t}$. As a result, the estimator suggested by Moffitt (1993), based on applying OLS to (5), is typically inconsistent unless there are either no time-varying exogenous regressors or the time-varying exogenous variables do not exhibit any autocorrelation.

3 Instrumental variables estimators

To overcome the problem of correlation between the regressors and the error term in (5) one may employ IV. Note that now we need instruments for $x_{i(t),t}$ even though these variables are exogenous in the original model. Let $w_{i(t),t}$ denote an R-dimensional vector of potential instruments that can be used to estimate (5). Standard IV conditions require that the instruments are uncorrelated with the equation's error term. This requires that[6]

$$E\{\varepsilon_{i(t),t} w_{i(t),t}\} = 0, \; t = 1, ..., T \tag{8}$$

$$E\{(y_{i(t),t-1} - y^*_{i(t),t-1}) w_{i(t),t}\} = 0, \; t = 1, ..., T \tag{9}$$

The first condition says that the instruments should be exogenous, which is similar to condition (7). The second condition, (9), implies that the prediction error is orthogonal to the instruments. That is, the instruments $w_{i(t),t}$ should not be able to predict any of the variation in $y_{i(t),t-1}$ left unexplained by $z_{i(t)}$. Consequently, a natural choice is $w_{i(t),t} = z_{i(t)}$, such that both (8)

[5] The majority of the results in McKenzie (2001) is also based on this assumption.

[6] We interpret these conditions as being required for each period t. On the pooled sample, this implies that the instruments in $w_{i(t),t}$ may be interacted with time dummies. Alternatively, the conditions can be imposed to hold over the pooled sample only. In this case, consistency is to be interpreted for T (or both T and N_t) going to infinity.

and (9) are automatically satisfied under condition (7). Given this result, and to simplify the discussion, we assume that $w_{i(t),t} = z_{i(t)}$. Thus, (9) is automatically satisfied and conditions (7) and (8) are identical.

It is well known that taking group averages is equivalent to instrumental variables estimation with the group dummies as instruments (Angrist, 1991, Moffitt, 1993). Thus, when the instruments $z_{i(t)}$ are a set of cohort dummies, estimation of (5) by instrumental variables is identical to applying OLS to the original model where all variables are replaced by their (time-specific) cohort sample averages.

However, imposing (7) may not be sufficient to identify the model parameters. To see this, consider the orthogonal projection of (1) upon the set of instruments:

$$E^*\{y_{i(t),t}|z_{i(t)}\} = \alpha E^*\{y_{i(t),t-1}|z_{i(t)}\} + E^*\{x_{i(t),t}|z_{i(t)}\}'\beta + E^*\{\varepsilon_{i(t),t}|z_{i(t)}\}. \quad (10)$$

While (7) implies that the last term in this expression is zero, it is clear that in order to identify α and β, the first two terms should not exhibit perfect collinearity. Consequently, a rank condition needs to be also satisfied. This can be formulated as:

$$E\begin{pmatrix} z_{i(1)}x'_{i(1),1} & z_{i(1)}y_{i(1),0} \\ z_{i(2)}x'_{i(2),2} & z_{i(2)}y_{i(2),1} \\ \vdots & \vdots \\ z_{i(T)}x'_{i(T),T} & z_{i(T)}y_{i(T),T-1} \end{pmatrix} = \begin{pmatrix} \Sigma_1 \\ \Sigma_2 \\ \vdots \\ \Sigma_T \end{pmatrix} = \Sigma \text{ has rank } K+1. \quad (11)$$

This condition says that $TR \times (K+1)$ cross-moment matrix Σ has full column rank. It requires that the instruments capture variation in $y_{i(t),t-1}$ independent of the variation in $x_{i(t),t}$. It is important to note that the rank condition for identification in (11) is in terms of the population moments. As a result, this order condition cannot be directly verified from the data and estimation error may thereby provide spurious identification.[7]

The pairwise quasi-differencing approach of Girma (2000) deviates from the above estimation strategy in two respects, although it essentially makes the same assumptions. First, the lagged value of y is not approximated by

[7] For example, if the instruments are cohort dummies, it is required that the population cohort means exhibit sufficient variation. In practice, the cohort sample averages will vary across cohorts due to sampling error even if all cohorts have the same mean and the model is not identified.

the lagged cohort average but by an arbitrarily selected observation from the cohort. Second, the instruments are not the cohort dummies, but individual, or averaged, observations from the cohort. As a result, Girma's approach employs a noisy approximation to the unobserved lagged values as well as noisy instruments. Although, under appropriate assumptions, this noise will cancel out asymptotically, there does not seem to be any gain in doing so. One can easily see this by noting that for an arbitrarily individual j that is in the same cohort as individual i,

$$E^*\{x_{i(t),t}|x_{j(t),t}, z_{i(t)}\} = E^*\{x_{i(t),t}|z_{i(t)}\}. \tag{12}$$

The only reason why individual j's observation provides information about individual i is because they are in the same cohort and this is already captured by the cohort dummies. A similar argument holds for the approximation of $y_{i(t),t-1}$.

The availability of appropriate instruments satisfying condition (7) may be rather limited. Thus it is natural to ask whether we can somewhat relax the requirements on the instruments. One possibility is to explicitly capture the cohort effects by including cohort fixed effects, as in Deaton (1985) and Collado (1998). This is done by including $z_{i(t)}$ as additional regressors in (5) but with *time-invariant* coefficients.[8] Thus, we allow for "cohort effects" by replacing assumption (7) with

$$E^*\{\varepsilon_{i(t),t}|z_{i(t)}\} = z'_{i(t)}\lambda \tag{13}$$

so that we can write[9]

$$y_{i(t),t} = \alpha\hat{y}_{i(t),t-1} + x'_{i(t),t}\beta + z'_{i(t)}\lambda + \eta_{i(t),t}, \tag{14}$$

where

$$E\{\eta_{i(t),t}z_{i(t)}\} = 0, \; t = 1, ..., T \tag{15}$$

replaces condition (7). This allows one to relax assumption (2) to

$$E\{\eta_{i(t),t}x_{i(t),t}\} = 0, \; t = 1, ..., T. \tag{16}$$

[8]The vector $z_{i(t)}$ not necessarily consists of cohort dummies. Alternative functional forms, that are either more flexible or more parsimonious, may be employed (see Moffitt, 1993). In the discussion we shall abstract from this possibility.

[9]The model presented here is a special case of McKenzie (2001)'s general model.

Under these conditions, one would estimate (14) by IV using $z_{i(t)}$, interacted with time dummies, as instruments. We shall refer to this as the *augmented IV estimator* noting that a time-varying λ would make the model unidentified. To achieve identification, we need to assume that $E^*\{y_{i(t),t-1}|z_{i(t)}\}$ and $E^*\{x_{i(t),t}|z_{i(t)}\}$ exhibit time-variation and are not collinear. This imposes additional restrictions upon the moment matrices Σ_t in (11). In particular, we need to strengthen (11) to:

$$\begin{pmatrix} \Sigma_2 - \Sigma_1 \\ \vdots \\ \Sigma_T - \Sigma_1 \end{pmatrix} \text{ has rank } K+1. \tag{17}$$

If Σ_t does not vary sufficiently over the cross-sections, the variation of the instruments is collinear with $z_{i(t)}$ and estimation will break down. A direct implication of this is that one needs at least three cross-sections ($t = 0, 1, 2$) to identify the model under these assumptions.

Computation of this augmented IV estimator is remarkably simple if $z_{i(t)}$ is a set of cohort dummies. First, one simply aggregates the data into cohort averages, which gives

$$\hat{y}_{i(t),t} = \alpha \hat{y}_{i(t),t-1} + \hat{x}'_{i(t),t}\beta + z'_{i(t)}\lambda + \hat{\eta}_{i(t),t}, \tag{18}$$

where the hats denote predicted values from a period-by-period regression on $z_{i(t)}$, that is, they denote (sample) cohort averages. Second, applying OLS to (18) corresponds to the standard within estimator for $(\alpha, \beta')'$ based upon treating the cohort-level data as a panel. The usual problem with estimating dynamic panel data models (see Nickell, 1981)[10], does not arise because under assumption (15) the error term, which is a within cohort average of individual error terms that are uncorrelated with $w_{i(t)}$, is asymptotically zero.[11] However, it remains whether suitable instruments can be found that satisfy the above conditions.

The estimators presented in Moffitt (1993), Girma (2000) or McKenzie (2001) do not allow for condition (15) to be relaxed. This means that time-varying cohort effects in the unobservables are not allowed. Starting from the case where $z_{i(t)}$ is a set of cohort dummies, Collado (1998) presents estimators

[10] With genuine panel data, the within estimator in the dynamic model has a substantial bias for small and moderate values of T.

[11] Recall that, asymptotically, the number of cohorts is fixed and the number of individuals goes to infinity.

that allow for a time-varying correlation between $y^*_{i(t),t-1}$ and $\eta_{i(t),t}$.[12] Following standard dynamic panel data IV procedures, the cohort fixed effects λ are first eliminated by first-differencing, to give

$$\begin{aligned} \hat{y}_{i(t),t} - \hat{y}_{i(t-1),t-1} &= \alpha(\hat{y}_{i(t),t-1} - \hat{y}_{i(t-1),t-2}) \\ &\quad + (\hat{x}_{i(t),t} - \hat{x}_{i(t-1),t-1})'\beta + (\hat{\eta}_{i(t),t} - \hat{\eta}_{i(t-1),t-1}), \end{aligned} \tag{19}$$

where, because of the definition of $z_{i(t),t}$ as cohort dummies, $\hat{y}_{i(t-1),t-1} = \hat{y}_{i(t),t-1}$ and similarly for the other variables and lags.[13] In other words, in (19) we have a standard first-differenced dynamic panel data model with the unit of observation being a cohort. The population counterpart is given by

$$\begin{aligned} y^*_{i(t),t} - y^*_{i(t-1),t-1} &= \alpha(y^*_{i(t),t-1} - y^*_{i(t-1),t-2}) \\ &\quad + (x^*_{i(t),t} - x^*_{i(t),t-1})'\beta + (\eta^*_{i(t),t} - \eta^*_{i(t-1),t-1}). \end{aligned} \tag{20}$$

If condition (15) is not valid, the cohort means $\eta^*_{i(t),t}$ are asymptotically nonzero and correlated with $z_{i(t)}$ in a time-varying way. Furthermore, $y^*_{i(t),t-1}$ and $\eta^*_{i(t-1),t-1}$ are correlated by construction, and OLS applied to (19) is inconsistent. Consequently, equation (19) must also be estimated by instrumental variables techniques. For the genuine panel data model, Arellano and Bond (1991) provide a list of instruments which is exploited by Collado (1998). The essential condition for identification, however, is that $y^*_{i(t-1),t-2}$, or one of its lags, is a valid instrument for $y^*_{i(t),t-1} - y^*_{i(t-1),t-2}$. This, as before, requires time variation in $y^*_{i(t),t-1}$, but also requires that[14]

$$E\{(\eta^*_{i(t),t} - \eta^*_{i(t-1),t-1})y^*_{i(t-1),t-2}\} = 0, \tag{21}$$

which is trivially satisfied if there are no cohort effects in $\eta_{i(t),t}$ (as with the augmented IV estimator).[15] In general, however, it is less obvious that this condition is satisfied.[16] If

$$E^*\{\eta_{i(t),t}|z_{i(t)}\} = z'_{i(t)}\lambda_t, \; t = 1, ..., T \tag{22}$$

[12] We shall restrict attention to the estimator that is claimed to be consistent for fixed T.

[13] Collado (1998) writes $y_{c,t}$, where c indexes cohorts.

[14] The equality sign is to be interpreted as "asymptotically equal".

[15] Note that any time-invariant cohort effects would have been eliminated by the first-difference transformation.

[16] Recall that we ignore the measurement error problem in Collado (1998) by assuming that $\hat{\delta}_t$, and the other "reduced form" parameters, asymptotically converge to their true values. In Collado's approach (following Deaton, 1985) the number of cohorts is assumed

we thus need that, for finite T,

$$E(z'_{i(t)}(\lambda_t - \lambda_{t-1}) \cdot z'_{i(t)}\delta_{t-2}\} = 0, \tag{23}$$

which, given the definition of $z_{i(t)}$, requires that the two vectors $\lambda_t - \lambda_{t-1}$ and δ_{t-2} are orthogonal. Collado (1998) imposes this condition while assuming that the dimension of z, and thus the dimensions of λ_t and δ_t, increase with sample size through imposing that there is no autocorrelation in the cohort effects for any given cohort. Thus, this allows for time-varying cohort effects in the unobservables provided they are uncorrelated over time.

It is hard to think of cases where (21) is satisfied while (13) is not, unless one is willing to make sampling assumptions at the cohort level.[17] If instruments can be chosen such that (13) is satisfied (as is done in Girma, 2000, and McKenzie, 2001) there are no time-varying cohort effects, and the augmented instrumental variables estimator proposed above is not only computationally attractive but also more efficient than Collado (1998)'s cohort-level GMM estimator. The reason is simple: unnecessarily instrumenting $\hat{y}_{i(t),t-1} - \hat{y}_{i(t-1),t-2}$ will lead to a loss in efficiency compared to the augmented IV estimator. As a result, estimating a dynamic model from cohort level data can be computationally much simpler than from genuine panel data.

4 A simple example

To illustrate and clarify the conditions for consistency of the respective estimators, as discussed above, we consider a simple example, where the model of interest contains only one exogenous variable. The next section presents a Monte Carlo study based upon this example. Given that the definition of a cohort is left to the researcher, it is important to start from a data generating process at the individual level rather than the cohort level. It is convenient

to grow with sample size. This implies that the number of instruments is increasing with N_t and it is no longer obvious that $\hat{\delta}_t - \delta_t$ is asymptotically zero (note that the definition of δ_t changes with sample size). Keeping the number of instruments fixed, as we shall do, the estimation error in $\hat{\delta}_t$ results in a small sample bias, which may or may not be negligible, depending upon the number of observations within a cohort. The Monte Carlo study in Section 5 will address this issue. See also Verbeek and Nijman (1992, 1993) for a discussion of this issue in the static model.

[17] In which case λ_t and δ_t are treated as random variables.

to think of the population being a large panel data set, from which different individuals are sampled each period. Assume that the data generating process is given by[18]

$$y_{it} = \alpha y_{i,t-1} + \beta x_{it} + \varepsilon_{it}, \qquad 0 < \alpha < 1, \tag{24}$$

where the error term has the usual error components structure

$$\varepsilon_{it} = \theta_i + v_{it}. \tag{25}$$

It is assumed that v_{it} is uncorrelated over time. Let z_i denote a set of time-invariant variables that are used as instruments. The main question is what conditions need to be imposed upon the relationships between z_i and the other variables in the model to guarantee that one or more of the estimators are consistent.

The first step is to approximate the lagged dependent variable by a linear projection upon z_i, denoted $E^*\{y_{i,t-1}|z_i\}$. Using recursive substitution, one can easily derive that

$$\begin{aligned} E^*\{y_{i,t-1}|z_i\} &= \beta \sum_{j=0}^{t-2} \alpha^j E^*\{x_{i,t-j-1}|z_i\} + \left(\sum_{j=0}^{t-2} \alpha^j\right) E^*\{\theta_i|z_i\} \\ &\quad + \alpha^{t-1} E^*\{y_{io}|z_i\} + \sum_{j=0}^{t-2} \alpha^j E^*\{v_{i,t-j-1}|z_i\}. \end{aligned} \tag{26}$$

In a first case, case A, we assume that the starting value y_{io} is exogenous and independent of θ_i.[19] In a second case, case B, we assume that the process is in equilibrium or - equivalently - that it started in an infinite past. For case B, the expression for $E^*\{y_{i,t-1}|z_i\}$ simplifies to

$$\begin{aligned} E^*\{y_{i,t-1}|z_i\} &= \beta \sum_{j=0}^{\infty} \alpha^j E^*\{x_{i,t-j-1}|z_i\} \\ &\quad + \frac{1}{1-\alpha} E^*\{\theta_i|z_i\} + \sum_{j=0}^{\infty} \alpha^j E^*\{v_{i,t-j-1}|z_i\}. \end{aligned} \tag{27}$$

[18] Because this and the following section discuss data generating mechanisms, rather than the sample available for estimation, it is appropriate and preferable to use the standard panel data notation.

[19] The starting date need not coincide with the beginning of the sample period.

Note that in the first case the instruments z_i may be predicting the lagged dependent variable partly through the initial value y_{io}.

Let us consider the respective estimators. First, the OLS estimator requires that the instruments are uncorrelated with the equation's error terms. That is

$$E^*\{\theta_i|z_i\} = 0 \quad (28)$$
$$E^*\{v_{it}|z_i\} = 0. \quad (29)$$

Further, it is required that the prediction error is uncorrelated with x_{it}, which requires that $x_{i,t-j-1} - E^*\{x_{i,t-j-1}|z_i\}$ is uncorrelated with x_{it} $(j = 0, 1, 2, ..)$. For a time-varying x_{it}-variable, this imposes the strong restriction of the absence of autocorrelation relative to the cohort-specific mean.

The standard IV estimator, using z_i as instruments for x_{it} as well, also imposes conditions (28)-(29). In addition, the rank condition requires that $E^*\{x_{it}|z_i\}$ is not collinear with $E^*\{y_{i,t-1}|z_i\}$. For case A it is sufficient that $E^*\{x_{it}|z_i\} \neq 0$ (and $\beta \neq 0$) or that $E^*\{y_{io}|z_i\} \neq 0$ (and $\alpha \neq 0$). For case B we need that $E^*\{x_{it}|z_i\}$ varies with t. For case B it is thus required that the time-invariant variables in z_i have a time-varying relationship with the exogenous variables in the model. For case A this is not required as the fixed starting date of the process induces variation over time even if the relationship between z_i and x_{it} is time-invariant.[20]

For the augmented IV estimator, which extends the standard IV estimator by including z_i in the model, condition (28) is no longer required. Implicitly, this allows for cohort-specific means in the processes for y_{it} and x_{it}. Both case A and case B now require that $E^*\{x_{it}|z_i\}$ varies with t (and $\beta \neq 0$). For case A, this may again be replaced by $E^*\{y_{io}|z_i\} \neq 0$. It should be stressed that, given the inclusion of z_i in the model, the need for time-variation in $E^*\{x_{it}|z_i\}$ and $E^*\{y_{i,t-1}|z_i\}$ is much stronger.

Finally, consider a simple variant of the estimator proposed by Collado (1998), which restricts attention to discrete-valued z_i. While this estimator does not require conditions (28)-(29), it does require that

$$E^*\{v_{it}|z_i\} - E^*\{v_{i,t-1}|z_i\}$$

is uncorrelated with $E^*\{v_{i,t-j}|z_i\}$ for $j = 2, 3, ...$ Note that the case with a non-zero time-invariant $E^*\{v_{it}|z_i\}$ is already captured by the augmented

[20] If α is small or t is large, the impact of the initial values is small and time-variation in $E^*\{x_{it}|w_i\}$ is recommended (as in case B).

IV estimator through the inclusion of z_i. The only relevant extension thus occurs with a time-varying $E^*\{v_{it}|z_i\}$. However, the only stationary process that is consistent with this condition is characterized by equi-correlation and implies that $E^*\{v_{it}|z_i\}$ is zero (given the presence of the cohort-specific effect $E\{\theta_i|z_i\}$). As a result, the relaxed conditions seem to be no weaker than those for the augmented IV estimator and the additional stage of instrumentation in Collado (1998) appears unnecessary.

5 A Monte Carlo study

We now present the results of a Monte Carlo experiment that investigate how large the cohort-specific variation in individual data need to be for the instrumental variables estimators to work satisfactorily with reasonable sample sizes. Cohort-specific effects may be present in a number of places: in the starting value of the process for y, in the process that generates the x-variables, in the individual-specific heterogeneity θ_i and in the idiosyncratic error terms v_{it}. We shall assume that v_{it} has no cohort effects. This assumption is stronger than that made by Collado, but is substantially weaker than what is assumed in Moffitt (1993), Girma (2000) and in the simulation study of McKenzie (2001). Note that the theoretical section of McKenzie presents a general model that incorporates the one we simulate here. However, he provides no simulation results for such a model.

Both Collado (1998) and Girma (2000) present a Monte Carlo study to illustrate how well their respective estimators perform with realistic sample sizes. Both studies find that for the estimators to work well it is necessary that the amount of variation in the data that is due to cohort effects is substantial. For example, in some instances it needs to be up to 75% of the total variation. There are some additional features of the design in Collado (1998) that are worth noting. First, there are no exogenous variables in the simulation. Second, she starts by assuming an autoregressive model at the cohort level and then generates individual observations as the cohort observations plus a random noise term. We feel that there are some shortcomings associated with such an approach. First, we prefer to think of the data generating process as something that operates at the individual level, where cohort effects are affecting some of the individual variables, rather than at the cohort level, individual observations being random deviations from the cohort means. Choosing cohorts is something done by the researcher when he/ she starts analyzing the data, not something determined by nature. In particular, when choosing cohorts one has to worry about similar issues as when choosing instruments. Second, the cohort-level AR(1) model is inconsistent with the individual data being generated from an AR(1) model (even though Collado suggests otherwise). In fact, her data generating mechanism implies that individual observations are generated by an AR(1) model plus white noise, which – as is well-known in the time series literature, see Hamilton (1994, Sect. 4.7) – does not correspond to an AR(1) process.

The data in our Monte Carlo experiment are generated as follows. First, we generate a vector of cohort dummies z_i. The exogenous variable is generated as

$$x_{it} = \theta x_{i,t-1} + \kappa_{1t}' z_i + \xi_{it}, \tag{30}$$

where $\xi_{it} \sim NID(0, \sigma_\xi^2)$ and κ_{1t} are vectors of cohort effects with elements $NID(0, \sigma_1^2)$.[21] We shall assume that 50% of the variation in these cohort effects is time-invariant. The total error variances and θ are chosen in such a way that the R^2 of the equation is 75%, (not taking into account the cohort effect), while the total unconditional variance of x_{it} is set to one, without loss of generality. An important element is the relative importance of the cohort effects in (30), which we shall vary between 25 and 50%. Note that in the absence of cohort effects in x_{it}, none of the estimators is expected to have good properties.

Second, we generate a starting value for the y process as

$$y_{i0} = \kappa_2' z_i + v_{io},$$

and any subsequent observations are generated as

$$y_{it} = \alpha y_{i,t-1} + \beta x_{it} + \kappa_3' z_i + v_{it},$$

where $v_{it} \sim NID(0, \sigma_v^2)$, $v_{i0} \sim NID(0, \sigma_{v0}^2)$. Both κ_2 and κ_3 are vectors of cohort effects with elements $NID(0, \sigma_2^2)$ and $NID(0, \sigma_3^2)$, respectively. Note that Girma (2000) and McKenzie (2001) assume $\kappa_3 = 0$ in their simulation studies. Without loss of generality, parameter values are chosen such that the total error variance (including the cohort effects) is normalized to one. For the initial period, it is normalized to $1/(1-\alpha^2)$. The relative importance of the cohort effects varies between 0%, 25% and 50%. The first 10 periods of data are discarded.

[21] The starting value of the x_{it} process is fixed at zero.

Within this Monte Carlo set-up, there are three sources of cohort effects: in the process for the exogenous variable, in the starting value y_{i0} and in the time-invariant unobserved heterogeneity θ_i. In our experiments we shall vary the proportion of cohort-specific variation in these three sources from 0%, 25% to 50%. The values of α and β are fixed at 0.5. We consider situations with $T = 5$ and $T = 10$, where $N = 2000$ or $N = 5000$ individuals.[22] The number of cohorts is varied from 20 to 100.

For the case with five time periods available for estimation, average estimates over 1000 replications are presented in Table 1 for $N = 2000$, and in Table 2 for $N = 5000$. Standard errors (in parentheses) are computed as sample standard deviations of the Monte Carlo estimates. The corresponding results for $T = 10$ are presented in Tables 3 and 4, respectively. The tables present results for the IV estimator, determined as OLS at the cohort aggregates, and the augmented IV estimator (AIV), based on the within estimator on cohort averages.

An examination of the simulation results leads to a number of interesting conclusions. First, the magnitude of the cohort effects in the starting value of y_{it} has only a small impact on the estimators. This is expected as the starting value dates back 10 periods from the beginning of the estimation sample. For the case with $T = 5$ estimation periods, the magnitude of cohort effect in the starting value of y_{it} has a bigger impact, both in the reduction of the bias and upon the standard errors of the estimators.

Second, when many cohorts are used, with a limited number of observations each, all estimators have a substantial bias. This bias is more pronounced for the coefficient on the exogenous variable. This bias reduces somewhat when the cohort effects are more important, but is still non negligible. This conclusion is consistent with current practice of using cohorts of at least 100 individuals and is also in line with the conclusion of Verbeek and Nijman (1992) that 100-200 individuals per cohort are recommended in order to ignore biases due to measurement error in the cohort averages (which, in our terminology, corresponds to estimation error in the reduced form parameters). Somewhat surprisingly, having more time periods available for estimation in some cases increases the bias that occurs from the small cohort sizes. The augmented IV estimator typically suffers somewhat more from this bias than does the IV estimator. For cases where the IV estimator is inconsistent, cohort sizes are hardly related to the amount of bias that is

[22] Note that $T = 5$ requires 6 cross-sections ($t = 0, 1, ..., T$) while $T = 10$ requires 11.

				$T=5$; $N=2000$; $\alpha=0.5$; $\beta=0.5$							
cohort effects (%)				$C=20$				$C=100$			
				IV		AIV		IV		AIV	
x	$y0$	y									
25	0	0	α	0.444	(0.032)	0.439	(0.033)	0.316	(0.030)	0.296	(0.033)
			β	0.534	(0.021)	0.592	(0.082)	0.614	(0.020)	0.714	(0.060)
25	25	0	α	0.464	(0.024)	0.452	(0.029)	0.367	(0.024)	0.330	(0.029)
			β	0.523	(0.017)	0.587	(0.080)	0.584	(0.017)	0.707	(0.060)
25	50	0	α	0.474	(0.019)	0.460	(0.025)	0.399	(0.020)	0.357	(0.027)
			β	0.517	(0.014)	0.582	(0.078)	0.566	(0.015)	0.702	(0.059)
50	25	0	α	0.480	(0.018)	0.474	(0.021)	0.417	(0.019)	0.397	(0.022)
			β	0.513	(0.012)	0.539	(0.058)	0.551	(0.013)	0.631	(0.053)
50	50	0	α	0.484	(0.015)	0.477	(0.019)	0.433	(0.016)	0.409	(0.020)
			β	0.510	(0.011)	0.537	(0.057)	0.542	(0.011)	0.626	(0.052)
25	25	25	α	0.869	(0.073)	0.474	(0.020)	0.820	(0.034)	0.398	(0.021)
			β	0.281	(0.084)	0.576	(0.071)	0.319	(0.036)	0.694	(0.053)
25	50	25	α	0.830	(0.076)	0.477	(0.018)	0.792	(0.034)	0.412	(0.020)
			β	0.304	(0.088)	0.575	(0.069)	0.335	(0.037)	0.692	(0.053)
50	25	25	α	0.805	(0.074)	0.483	(0.016)	0.774	(0.034)	0.432	(0.017)
			β	0.317	(0.074)	0.534	(0.051)	0.338	(0.032)	0.618	(0.047)
25	25	50	α	0.961	(0.063)	0.484	(0.015)	0.940	(0.028)	0.436	(0.016)
			β	0.228	(0.090)	0.572	(0.061)	0.248	(0.038)	0.687	(0.046)
25	50	50	α	0.929	(0.067)	0.486	(0.013)	0.915	(0.029)	0.444	(0.014)
			β	0.246	(0.096)	0.571	(0.059)	0.263	(0.040)	0.686	(0.045)
50	25	50	α	0.907	(0.071)	0.489	(0.012)	0.897	(0.031)	0.454	(0.013)
			β	0.258	(0.081)	0.530	(0.044)	0.266	(0.034)	0.609	(0.040)
50	50	50	α	0.882	(0.073)	0.491	(0.011)	0.876	(0.032)	0.460	(0.012)
			β	0.272	(0.084)	0.530	(0.043)	0.278	(0.035)	0.607	(0.039)

Table 1: Average estimates and standard errors, $T=5$, $N=2000$

				$T=5$; $N=5000$; $\alpha=0.5$; $\beta=0.5$							
cohort effects (%)				$C=20$				$C=100$			
				IV		AIV		IV		AIV	
x	$y0$	y									
25	0	0	α	0.476	(0.020)	0.474	(0.021)	0.404	(0.020)	0.397	(0.020)
			β	0.515	(0.013)	0.544	(0.053)	0.560	(0.013)	0.641	(0.046)
25	25	0	α	0.485	(0.015)	0.479	(0.018)	0.436	(0.015)	0.417	(0.018)
			β	0.510	(0.010)	0.540	(0.052)	0.541	(0.011)	0.633	(0.045)
25	50	0	α	0.489	(0.012)	0.479	(0.018)	0.453	(0.012)	0.432	(0.016)
			β	0.507	(0.009)	0.538	(0.051)	0.530	(0.009)	0.628	(0.045)
50	25	0	α	0.491	(0.011)	0.489	(0.013)	0.463	(0.011)	0.454	(0.013)
			β	0.505	(0.008)	0.517	(0.038)	0.523	(0.008)	0.567	(0.035)
50	50	0	α	0.493	(0.010)	0.491	(0.012)	0.471	(0.010)	0.460	(0.012)
			β	0.504	(0.007)	0.516	(0.037)	0.519	(0.007)	0.564	(0.034)
25	25	25	α	0.881	(0.073)	0.489	(0.012)	0.870	(0.032)	0.453	(0.013)
			β	0.273	(0.084)	0.535	(0.045)	0.282	(0.036)	0.621	(0.040)
25	50	25	α	0.839	(0.075)	0.490	(0.011)	0.834	(0.033)	0.460	(0.011)
			β	0.297	(0.089)	0.534	(0.045)	0.304	(0.037)	0.618	(0.039)
50	25	25	α	0.813	(0.074)	0.492	(0.010)	0.809	(0.033)	0.470	(0.010)
			β	0.312	(0.074)	0.514	(0.033)	0.316	(0.032)	0.559	(0.031)
25	25	50	α	0.968	(0.062)	0.493	(0.009)	0.970	(0.026)	0.471	(0.009)
			β	0.223	(0.089)	0.533	(0.039)	0.223	(0.037)	0.614	(0.034)
25	50	50	α	0.934	(0.067)	0.494	(0.008)	0.940	(0.028)	0.475	(0.008)
			β	0.242	(0.096)	0.532	(0.038)	0.241	(0.039)	0.613	(0.033)
50	25	50	α	0.912	(0.071)	0.495	(0.008)	0.918	(0.031)	0.451	(0.008)
			β	0.254	(0.081)	0.513	(0.028)	0.250	(0.034)	0.554	(0.026)
50	50	50	α	0.886	(0.072)	0.496	(0.007)	0.895	(0.031)	0.483	(0.007)
			β	0.269	(0.084)	0.513	(0.027)	0.264	(0.035)	0.553	(0.025)

Table 2: Average estimates and standard errors, $T=5$, $N=5000$

cohort effects (%)				$T = 10$; $N = 2000$; $\alpha = 0.5$; $\beta = 0.5$							
				$C = 20$				$C = 100$			
x	$y0$	y		IV		AIV		IV		AIV	
25	0	0	α	0.431	(0.026)	0.428	(0.027)	0.289	(0.023)	0.278	(0.026)
			β	0.555	(0.021)	0.586	(0.042)	0.669	(0.019)	0.728	(0.033)
25	25	0	α	0.447	(0.021)	0.441	(0.024)	0.322	(0.021)	0.306	(0.024)
			β	0.543	(0.018)	0.579	(0.041)	0.644	(0.018)	0.717	(0.033)
25	50	0	α	0.456	(0.019)	0.450	(0.021)	0.347	(0.019)	0.329	(0.022)
			β	0.536	(0.016)	0.574	(0.039)	0.625	(0.016)	0.709	(0.033)
50	25	0	α	0.472	(0.015)	0.469	(0.016)	0.392	(0.016)	0.381	(0.018)
			β	0.523	(0.012)	0.536	(0.028)	0.587	(0.013)	0.631	(0.027)
50	50	0	α	0.476	(0.013)	0.472	(0.014)	0.404	(0.015)	0.392	(0.017)
			β	0.520	(0.011)	0.534	(0.028)	0.578	(0.012)	0.626	(0.026)
25	25	25	α	0.894	(0.053)	0.466	(0.017)	0.841	(0.025)	0.374	(0.018)
			β	0.195	(0.066)	0.565	(0.035)	0.245	(0.030)	0.694	(0.029)
25	50	25	α	0.875	(0.056)	0.470	(0.015)	0.829	(0.026)	0.387	(0.017)
			β	0.210	(0.070)	0.563	(0.035)	0.255	(0.031)	0.670	(0.029)
50	25	25	α	0.840	(0.060)	0.480	(0.012)	0.808	(0.027)	0.418	(0.014)
			β	0.234	(0.065)	0.530	(0.025)	0.262	(0.028)	0.612	(0.023)
25	25	50	α	0.957	(0.040)	0.479	(0.012)	0.933	(0.018)	0.416	(0.014)
			β	0.147	(0.061)	0.558	(0.030)	0.175	(0.026)	0.679	(0.025)
25	50	50	α	0.944	(0.043)	0.481	(0.011)	0.924	(0.019)	0.424	(0.013)
			β	0.157	(0.065)	0.557	(0.029)	0.183	(0.028)	0.677	(0.025)
50	25	50	α	0.919	(0.050)	0.487	(0.009)	0.906	(0.021)	0.444	(0.011)
			β	0.174	(0.061)	0.525	(0.021)	0.186	(0.026)	0.599	(0.020)
50	50	50	α	0.908	(0.051)	0.488	(0.009)	0.898	(0.022)	0.448	(0.010)
			β	0.183	(0.064)	0.524	(0.020)	0.192	(0.027)	0.597	(0.019)

Table 3: Average estimates and standard errors, $T = 10$, $N = 2000$

				$T = 10$; $N = 5000$; $\alpha = 0.5$; $\beta = 0.5$							
cohort effects (%)				$C = 20$				$C = 100$			
x	$y0$	y		IV		AIV		IV		AIV	
25	0	0	α	0.470	(0.015)	0.469	(0.015)	0.386	(0.016)	0.381	(0.017)
			β	0.524	(0.012)	0.539	(0.025)	0.592	(0.013)	0.640	(0.024)
25	25	0	α	0.478	(0.012)	0.475	(0.013)	0.409	(0.014)	0.400	(0.015)
			β	0.518	(0.010)	0.539	(0.025)	0.574	(0.012)	0.630	(0.024)
25	50	0	α	0.482	(0.010)	0.479	(0.012)	0.424	(0.012)	0.414	(0.014)
			β	0.515	(0.009)	0.533	(0.024)	0.562	(0.011)	0.623	(0.023)
50	25	0	α	0.489	(0.009)	0.487	(0.009)	0.450	(0.010)	0.446	(0.010)
			β	0.509	(0.007)	0.515	(0.017)	0.540	(0.008)	0.565	(0.017)
50	50	0	α	0.490	(0.008)	0.489	(0.008)	0.456	(0.009)	0.451	(0.010)
			β	0.508	(0.007)	0.514	(0.017)	0.535	(0.007)	0.561	(0.017)
25	25	25	α	0.906	(0.052)	0.486	(0.009)	0.891	(0.022)	0.440	(0.011)
			β	0.184	(0.064)	0.529	(0.021)	0.198	(0.027)	0.611	(0.021)
25	50	25	α	0.886	(0.055)	0.487	(0.008)	0.875	(0.023)	0.447	(0.010)
			β	0.199	(0.069)	0.528	(0.021)	0.211	(0.029)	0.608	(0.020)
50	25	25	α	0.849	(0.059)	0.492	(0.007)	0.844	(0.026)	0.464	(0.008)
			β	0.227	(0.064)	0.512	(0.015)	0.232	(0.027)	0.554	(0.015)
25	25	50	α	0.963	(0.039)	0.491	(0.007)	0.960	(0.016)	0.462	(0.008)
			β	0.140	(0.060)	0.526	(0.018)	0.145	(0.024)	0.600	(0.018)
25	50	50	α	0.950	(0.042)	0.492	(0.006)	0.949	(0.017)	0.465	(0.007)
			β	0.151	(0.065)	0.525	(0.018)	0.154	(0.026)	0.598	(0.017)
50	25	50	α	0.923	(0.049)	0.495	(0.005)	0.926	(0.020)	0.476	(0.006)
			β	0.170	(0.061)	0.510	(0.012)	0.167	(0.025)	0.547	(0.012)
50	50	50	α	0.912	(0.051)	0.495	(0.005)	0.916	(0.021)	0.478	(0.006)
			β	0.179	(0.064)	0.510	(0.012)	0.175	(0.026)	0.545	(0.012)

Table 4: Average estimates and standard errors, $T = 10$, $N = 5000$

found.

A third conclusion is related to the size of the standard errors. In general, the standard errors do not provide any indication of the degree of bias that can be expected. When the number of cohorts is increased, standard errors typically *decrease* somewhat, while the bias *increases* substantially. Note that the sampling variation of all estimators is rather low. On the one hand, this is a positive result as it indicates that the estimators are quite accurate, even if the cohort effects are relatively unimportant. On the other hand, it is of concern because in cases where the estimators are severely biased, high standard errors are not warning against their use. Apparently, sampling variation in the cohort averages produces estimates that have a low standard error, even if this variation is hardly related to genuine cohort variation.

Fourth, when cohort effects are introduced in the main equation, the IV estimator is severely biased. With the true autoregressive coefficient equal to 0.5, IV produces estimates in the range of 0.80-0.95, depending upon the importance of the cohort effects. Similarly, the coefficient estimates for the exogenous regressor are in the range of 0.15-0.30, again with a true value of 0.5. It is interesting to note that the precision of the IV estimators is reduced when cohort effects are introduced in the main equation, but insufficiently to generate a potential warning for its bias. Also note that the augmented IV estimator is hardly affected by the presence of cohort effects in the process for y_{it}. Rather, its biases appears to be slightly smaller than in their absence.

Fifth, in the cases where the estimators are consistent, there generally appears to be a negative small sample bias in the estimation of the autoregressive coefficient. Similar results are reported for alternative estimators and data generating processes by their respective authors (Collado, 1998, Girma, 2000). We also find a positive bias in the coefficient for the exogenous regressor (which is not present in Collado's simulations). These biases become smaller when sample sizes increase (particularly when the number of observations per cohort increase), and when the cohort effects become more important. This is particularly true of the cohort effect in the process for the exogenous variable. For reasonable cohort sizes, the biases seem to be acceptable.

Sixth, for the genuine panel data case, the results of Nickell (1981) imply that for $T = 10$ the within estimator has a probability limit (for $N \to \infty$) of 0.33. Clearly, the cohort-based estimators perform better in terms of small T bias. This confirms our earlier remark that estimating a dynamic model from cohort level data can be computationally simpler than from genuine panel data.

6 Concluding remarks

Several authors have argued that the estimation of dynamic models at the individual level is possible on the basis of repeated cross-sections and present alternative estimators that are consistent under appropriate conditions. The proposed estimators vary widely in the degree of computational complexity and in the way they are motivated and presented. For example, Moffitt (1993) presents a simple estimator based on OLS where the lagged dependent variable is replaced by a predicted value. In contrast, Collado (1998) requires aggregation of all observations into cohorts, after which a GMM estimator, with a measurement error correction, is employed using the cohort aggregates. Girma (2000) presents an IV estimator that involves the use of other observations from the cohort to approximate the lagged dependent variable and to act as instruments. This paper reviewed the identification conditions underlying these alternative estimators and presented computationally attractive alternatives. Our main focus was upon the identifying conditions. Consistency refers to the case where the time dimension is fixed and the number of individuals grows, while the number of instruments is constant. Note that McKenzie (2001) provides a detailed discussion of the asymptotic properties of the estimators using alternative multidimensional limits.

The three most important conclusions from our analysis are the following. First, one can only identify individual dynamics without having individual time-series or panel data, by assuming the existence of a set of instruments that is (1) observed for each person in the entire sample, and (2) appropriate for each variable in the model of interest. Such instruments can be used to aggregate the data in a number of mutually exclusive groups (cohorts). While such an assumption may be unrealistic in certain applications, it is, generally, not testable due to its identifying nature. The Monte Carlo study illustrates the fact that imposing incorrect identifying assumptions may produce severely biased estimators that exhibit little sampling variation. Overall, the size of the standard errors is no indication for the amount of bias that is present.

Second, if one assumes the absence of cohort effects in the unobservables of the main equation, a simple IV estimator based upon OLS using the cohort aggregates, is consistent (provided all variables exhibit genuine cohort-specific variation). The estimator proposed by Moffitt (1993) is inconsistent in the presence of time-varying exogenous regressors, while the estimator presented by Girma (2000) for this model is unnecessarily complicated. Given moderate sample sizes, the sampling variation of the IV estimator is mainly driven by the importance of the cohort effects in the exogenous variables. The Monte Carlo study reveals that standard errors are fairly low, even when only 25% of the innovations in the process for the exogenous variable can be attributed to cohort-specific effects. When the number of instruments (cohorts) is large relative to the number of individuals, a small sample bias is present in the IV estimator. When cohort sizes are 100 or more, these biases seem to be acceptable, provided sufficient cohort-specific variation is present in the exogenous variables.

Third, if one allows for time-invariant cohort effects in the unobservables of the main equation, an augmented IV estimator, based upon the within estimator using cohort-aggregates, is consistent. This requires that all exogenous variables exhibit genuine time-varying cohort-specific variation. Again, sufficiently large cohorts are required to reduce the small sample bias in this estimator. Notably, the bias that is present in the within estimator for the dynamic model using genuine panel data (see Nickell, 1981), is much larger than what is found for similar estimators employed upon cohort aggregates.

References

[1] Alessie, R., M.P. Devereux and G. Weber (1997), "Intertemporal Consumption, Durables and Liquidity Constraints: A Cohort Analysis", *European Economic Review*, 41, 37-59.

[2] Anderson, T.W. and C. Hsiao (1981), "Estimation of Dynamic Models with Error Components", *Journal of the American Statistical Association*, 76, 598-606.

[3] Angrist, J.D. (1991), "Grouped-data estimation and testing in simple labor-supply models", *Journal of Econometrics*, 47, 243-266.

[4] Arellano, M. and S. Bond (1991), "Some Test of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", *Review of Economic Studies*, 58, 277-297.

[5] Blundell, R., M. Browning, and C. Meghir (1994), "Consumer Demand and the Life-Cycle Allocation of Household Expenditure", *Review of Economic Studies*, 61, 57-80.

[6] Collado, M.D. (1998), "Estimating Dynamic Models from Time Series of Independent Cross-Sections", *Journal of Econometrics,* 82, 37-62.

[7] Dargay, J.M. and P.C. Vythoulkas (1999), "Estimation of a Dynamic Car Ownership Model: A Pseudo Panel Approach", *Journal of Transport Economics and Policy*, 33, 283-302.

[8] Deaton, A. (1985), "Panel Data from Times Series of Cross-Sections", *Journal of Econometrics*, 30, 109-126.

[9] Girma, S. (2000), "A Quasi-Differencing Approach to Dynamic Modelling from a Time Series of Independent Cross-Sections", *Journal of Econometrics,* 98, 365-383.

[10] Hamilton, J.D. (1994), *Time Series Analysis*, Princeton University Press, Princeton.

[11] McKenzie, D.J. (2001), "Asymptotic Theory for Heterogeneous Dynamic Pseudo-Panels", working paper, Stanford University.

[12] Moffitt, R. (1993), "Identification and Estimation of Dynamic Models with a Time Series of Repeated Cross-Sections", *Journal of Econometrics*, 59, 99-124.

[13] Nickell, S. (1981), "Biases in Dynamic Models with Fixed Effects", *Econometrica*, 49, 1417-1426.

[14] Verbeek, M. (1996), "Pseudo Panel Data", in: L. Mátyás and P. Sevestre, eds., *The Econometrics of Panel Data: Handbook of the Theory with Applications*, 2nd revised edition, Kluwer Academic Publishers, 280-292.

[15] Verbeek, M. and Th. Nijman (1992), "Can Cohort Data Be Treated as Genuine Panel Data?", *Empirical Economics*, 17, 9-23.

[16] Verbeek, M. and Th. Nijman (1993),"Minimum MSE Estimation of a Regression Model with Fixed Effects from a Series of Cross Sections", *Journal of Econometrics*, 59, 125-136.