Minimum MSE estimation of a regression model with fixed effects from a series of cross-sections

Marno Verbeek and Theo Nijman*

Tilburg University, 5000 LE Tilburg, The Netherlands

If panel data are not available but repeated cross-sections are, the parameters in a regression model with fixed individual effects can be estimated consistently using the cohort approach proposed by Deaton (1985). In this paper we show that Deaton's estimator is inconsistent if the number of time periods is small, even if the number of cohorts tends to infinity. Moreover, we propose an alternative estimator which does not suffer from a bias due to a small number of sampling periods and we introduce a new class of estimators, containing both estimators mentioned above. We discuss minimum mean squared error estimation within this class. Our results show that it may be optimal to eliminate only part of the measurement error in the cohort averages, since the implied bias is offset by a smaller variance.

1. Introduction

Many models suggested by economic theory can be written as regression models with fixed individual effects [see, e.g., Heckman and MaCurdy (1980), MaCurdy (1981), or Browning, Deaton, and Irish (1985)]. It is well known by now how such models can be estimated from panel data [see, e.g., Hsiao (1986)]. Unfortunately, however, panel data (i.e., repeated observations on the same individuals) are often not available, though series of cross-sections (i.e., observations on different individuals in a number of periods) are.

Recently, Deaton (1985) suggested a way to estimate parameters in a fixed effects regression model from a series of cross-sections. He proposed to construct a pseudo panel consisting of a number of cohorts (i.e., groups of individuals sharing some common characteristics such as age) followed over time, and

showed how consistent estimators of the regression parameters can be obtained from this pseudo panel. Since only a limited number of observations is available for each cohort, the observed cohort aggregates are considered as error-ridden measurements of the true cohort population values and an errors-in-variables technique is proposed to obtain consistent estimators.

In section 2 of this paper we introduce a class of estimators, containing Deaton's errors-in-variables estimator as well as the standard within estimator on the pseudo panel as special cases. We show that consistency of Deaton's estimator requires that the number of available cross-sections tends to infinity. In addition, we suggest a modified estimator which does not suffer from an inconsistency due to a small number of time periods.

In sections 3 and 4 we make some additional assumptions on the model of interest and the way in which the cohorts are constructed and, subsequently, we consider minimum mean squared error estimation of the slope parameters in a simple regression model with fixed individual effects. Section 5 concludes.

2. A new class of estimators

Consider the following linear model on an individual level:

$$y_{it} = x_{it}\beta + \theta_i + \varepsilon_{it}, \qquad t = 1, \dots, T, \tag{1}$$

where i indexes individuals, x_{it} is a k-dimensional row vector of exogenous variables, and interest lies in the k-dimensional parameter vector β. Throughout this paper the error term ε_{it} is assumed to be uncorrelated with the explanatory variables in x_{it} and the individual effects θ_i. Moreover, it is assumed that the data set is a series of T independent cross-sections with N individual observations each, where different values for the index i reflect different individuals ($i = 1, \dots, NT$).

In many applications, for example in life cycle models, economic theory suggests that the individual effects θ_i are correlated with the explanatory variables. Consequently, regression estimators which simply include θ_i in the error term ('random effects estimators') will be inconsistent. Moreover, a fixed effect treatment of the θ_i is not possible because repeated observations on the same individuals are not available. Deaton (1985) has shown how β can nevertheless be identified. Suppose, as in Deaton (1985) and Verbeek and Nijman (1992a) that the individual observations are split into C groups or cohorts (for example age intervals). Aggregating all observations to cohort level and denoting the average value of x_{it} in cohort c at time t by $\bar{x}_{ct}$ (and analogously for all other variables in the model), we obtain a model in terms of observed cohort

aggregates,

$$\bar{y}_{ct} = \bar{x}_{ct}\beta + \bar{\theta}_{ct} + \bar{\varepsilon}_{ct}, \qquad c = 1, \dots, C, \quad t = 1, \dots, T. \tag{2}$$

The problem with estimating β in (2) is that the aggregated individual effects $\bar{\theta}_{ct}$ are unobserved, not constant over time, and correlated with $\bar{x}_{ct}$ as long as x_{it} and θ_i are correlated. Consequently, the within estimator on the pseudo panel will, in general, not be consistent. Now, consider the cohort population version of (2),

$$y_{ct}^* = x_{ct}^*\beta + \theta_c + \varepsilon_{ct}^*, \qquad c = 1, \dots, C, \quad t = 1, \dots, T, \tag{3}$$

where the asterisks denote unobservable population cohort means and where θ_c is the cohort fixed effect, which is constant over time because population cohorts contain the same individuals in all periods (ignoring birth and death of individuals). Now $\bar{y}_{ct}$ and $\bar{x}_{ct}$ can be considered as error-ridden measurements of y_{ct}^* and x_{ct}^*. It is assumed that the measurement errors are normally distributed with zero mean, independent of the true values,

$$\begin{pmatrix} \bar{y}_{ct} - y_{ct}^* \\ \bar{x}_{ct}' - x_{ct}^{*\prime} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma_{00} & \sigma' \\ \sigma & \Omega_2 \end{pmatrix} \right). \tag{4}$$

In general, the measurement error variances depend on the number of observations within each cohort. Assuming, for convenience, that there are n_c observations available in each cohort for each period, the measurement errors tend to zero if $n_c \to \infty$. For any construction of cohorts, σ_{00}, σ, and Ω_2 can be estimated consistently from the individual observations. Without affecting our main conclusions we shall in the sequel assume that Ω_2 and σ are known.

We can now define a class of estimators for β, indexed by a parameter $\alpha \in [0, 1]$, as follows:

$$\tilde{\beta}(\alpha) = (M_{xx} - \alpha\Omega_2)^{-1}(m_{xy} - \alpha\sigma), \tag{5}$$

where

$$M_{xx} = \frac{1}{CT} \sum_{c=1}^{C} \sum_{t=1}^{T} (\bar{x}_{ct} - \bar{x}_c)(\bar{x}_{ct} - \bar{x}_c)', \tag{6}$$

$$m_{xy} = \frac{1}{CT} \sum_{c=1}^{C} \sum_{t=1}^{T} (\bar{x}_{ct} - \bar{x}_c)(\bar{y}_{ct} - \bar{y}_c), \tag{7}$$

with $\bar{x}_c = (1/T)\sum_{s=1}^{T} \bar{x}_{cs}$. The estimator proposed by Deaton (1985) eliminates the total measurement error variances and is characterized by $\alpha = 1$. On the

other hand, the within estimator based on the pseudo panel ignores the errors-in-variables problem and is characterized by $\alpha = 0$. In the sequel we shall analyze the properties of $\tilde{\beta}(\alpha)$ in more detail, and, from these results, derive some conclusions on the optimal choice of α.

Because of the stochastic nature of M_{xx}, large sample theory appears to be required[1] to approximate the behaviour of $\tilde{\beta}(\alpha)$. Several limiting arguments can be used to derive large sample approximations of the sampling behaviour of these estimators. Deaton (1985) uses asymptotics on CT, Collado (1991) uses $C \to \infty$, and Moffitt (1993) and Angrist (1991) use asymptotics on n_c (implicitly assuming that $\varepsilon_{ct}^* = 0$). The latter choice eliminates the measurement error problem and makes the value of α (asymptotically) irrelevant. However, as noted by Verbeek and Nijman (1992a), assuming $n_c \to \infty$ may not be an accurate approximation, unless the value of n_c is very high. Moreover, the accuracy of this approximation depends crucially on the way in which the cohorts are constructed. This point will become clearer in the discussion below. Moreover, in practical applications large cohort sizes n_c may lead to groups that are so heterogeneous that orthogonality conditions are violated. It should also be noted that many observations per cohort imply a small number of cohort observations in the pseudo panel, resulting in inefficient estimators.

Following Deaton (1985), Collado (1991), and Verbeek and Nijman (1992a), we will in this paper concentrate on the case where the number of observations per cohort (n_c) is finite. In analogy with real panel data it seems most natural to define asymptotics on the number of individual observations, thus on the number of cohorts C in the pseudo panel. One could also define asymptotics on T, but we will not do so, because it may be important to get consistent estimators for small values of T.

Defining Ω_1 as the (asymptotic) within cohort variation in the true cohort means,

$$\Omega_1 = \plim_{C \to \infty} \frac{1}{CT} \sum_{c=1}^{C} \sum_{t=1}^{T} (x_{ct}^* - x_c^*)(x_{ct}^* - x_c^*)', \tag{8}$$

where $x_c^* = (1/T)\sum_{s=1}^{T} x_{cs}^*$, it is easily verified that

$$\plim_{C \to \infty} M_{xx} - \alpha\Omega_2 = \Omega_1 + (\tau - \alpha)\Omega_2, \tag{9}$$

where $\tau = (T - 1)/T$. In the sequel we shall assume that this is a positive definite matrix. Furthermore,

$$\plim_{C \to \infty} m_{xy} - \alpha\sigma = \Omega_1\beta + (\tau - \alpha)\sigma. \tag{10}$$

---

[1] Note that the expectation of $\tilde{\beta}(\alpha)$ will not be defined in many cases, because $M_{xx} - \alpha\Omega_2$ may be singular [see also Fuller (1987, p. 164)].

From (9), (10), and (5) it follows directly that

$$\operatorname*{plim}_{c \to \infty} \tilde{\beta}(\alpha) = (\Omega_1 + (\tau - \alpha)\Omega_2)^{-1}(\Omega_1 \beta + (\tau - \alpha)\sigma). \tag{11}$$

Whether or not $\tilde{\beta}(\alpha)$ is consistent for β depends on the value of α as well as on the relationship between Ω_2 and σ.

If θ_i and x_{it} are uncorrelated, it is easily verified that $\sigma = \Omega_2 \beta$. Consequently, $\tilde{\beta}(\alpha)$ is consistent for any value of α. On the other hand, if θ_i and x_{it} are correlated, it is easily seen that $\tilde{\beta}(\alpha)$ is inconsistent for finite T unless $\alpha = \tau$. This implies that Deaton's errors-in-variables estimator, $\tilde{\beta}(1)$, is inconsistent if T is fixed and, consequently, in conflict with Deaton's (1985) assertion that $\tilde{\beta}(1)$ is 'clearly consistent as $CT \to \infty$'. The inconsistency of $\tilde{\beta}(1)$ is caused by the incidental parameters problem (implying that θ cannot be estimated consistently), which carries over to the parameters of interest β, as long as the measurement errors are not correctly eliminated. Within our class of estimators it is easy to adjust Deaton's estimator to attain consistency by choosing $\alpha = \tau$ instead of $\alpha = 1$. If only a few cross-sections are available, the difference between the two estimators may be substantial. Of course, if $T \to \infty$, $\tau \to 1$ and consistency of both estimators is guaranteed.

The choice of α will not only affect the consistency of the estimators, it will also have an effect on their variances. From (5) one can readily derive that the variance of $\tilde{\beta}(\alpha)$ is increasing in α. Consequently, Deaton's errors-in-variables estimator ($\alpha = 1$) not only is inconsistent, it also has a larger variance than the consistent estimator in our class. More generally, because the variance is monotonically increasing in α and the inconsistency has a minimum at $\alpha = (T - 1)/T$, there is room for a trade-off between variance and asymptotic bias, and is it not unlikely that an estimator with a value of α smaller than τ will perform better in terms of mean squared error. The results of Fuller (1987, p. 163ff.) also suggest that a smaller value of α may improve the moment properties of the estimators.

To illustrate the trade-off between large sample variance and large sample bias, we shall in the next two sections consider a simple illustrative case and present expressions for the probability limit and the corresponding variance of $\tilde{\beta}(\alpha)$ given α and n_c.

3. The bias in $\tilde{\beta}(\alpha)$ in a simplified case

In this section we shall consider a specific example with a specific way of constructing the cohorts, and derive analytical expressions for the probability limit of $\tilde{\beta}(\alpha)$, as a function of α and the cohort size n_c. The model considered in this section is

$$y_{it} = x_{it}\beta + \theta_i + \varepsilon_{it}, \qquad t = 1, \dots, T, \tag{12}$$

where now x_{it} is only one-dimensional. As before, the individual effects θ_i may be correlated with the explanatory variable x_{it}. In particular, we shall impose the assumption [based on Mundlak (1978)] that the individual effects θ_i are correlated with the x's in the following way:

$$\theta_i = \bar{x}_i \lambda + \xi_i, \tag{13}$$

where $E\{\xi_i | x_{it}\} = 0$ for all $t = 1, \dots, T$ and $\bar{x}_i = (1/T)(x_{i1} + \cdots + x_{iT})$.

Our next assumption concerns the construction of the cohorts. As in Verbeek and Nijman (1992a), it is assumed that grouping is based on some variable[2] z such that the support of the density of z is split into C intervals with equal probability mass, each interval corresponding with a particular cohort. In practice, the variable z is often based on the date of birth (resulting in age cohorts), but it is also possible to base z on more than one underlying variable. Finally, we have to specify the relationships between the cohort-identifying variable z and the other variables in the model. We shall assume that the regressor variable x_{it} is correlated with z_i according to

$$x_{it} = \mu_t + \gamma_t z_i + v_{it}, \tag{14}$$

where[3] $E\{v_{it} | z_i\} = 0$, $\text{cov}\{v_{it}, v_{is}\} = \rho\sigma_v^2$. The μ_t are fixed unknown constants. The error term ε_{it} is assumed to be uncorrelated with z_i.

Under the assumptions above, it can be shown[4] that the inconsistency in $\tilde{\beta}(\alpha)$ is given by[5]

$$\delta(\alpha) = \operatorname*{plim}_{C \to \infty}(\tilde{\beta}(\alpha) - \beta) = \lambda \frac{1 + (T-1)\rho}{T} \frac{(\tau - \alpha)\omega_2}{\omega_1 + (\tau - \alpha)\omega_2}, \tag{15}$$

where ω_1 and ω_2 are given by

$$\omega_1 = \frac{1}{T}\sum_{t=1}^{T}\left(\mu_t - \frac{1}{T}\sum_{s=1}^{T}\mu_s\right)^2 + \frac{1}{T}\sum_{t=1}^{T}\left(\gamma_t - \frac{1}{T}\sum_{s=1}^{T}\gamma_s\right)^2, \tag{16}$$

$$\omega_2 = \operatorname*{plim}_{C \to \infty} \frac{1}{CT}\sum_{c=1}^{C}\sum_{t=1}^{T}(\bar{x}_{ct} - x_{ct}^*)^2 = n_c^{-1}\sigma_v^2, \tag{17}$$

---

[2] For mathematical convenience, this variable z is assumed to be absolutely continuous, distributed independently across individuals. Without loss of generality the variance of z is normalized to one.

[3] The equicorrelation of v_{it} over time is chosen for mathematical convenience only and can easily be relaxed to more general autocorrelation patterns.

[4] Technical details are given in the appendix of the working paper version [Verbeek and Nijman (1992b)], available upon request.

[5] In this section lower-case letters are used to stress that ω_1 and ω_2 are scalar.

Table 1

Relative inconsistencies of $\tilde{\beta}(\alpha)$, apart from λ/β.

ω_1/σ_v^2	n_c	$T = 2$		$T = 10$	
		$\alpha = 0$	$\alpha = 1$	$\alpha = 0$	$\alpha = 1$
0.025	10	0.50	—	0.43	− 0.37
	50	0.21	− 0.50	0.23	− 0.05
	100	0.13	− 0.19	0.15	− 0.02
	200	0.07	− 0.08	0.08	− 0.01
0.10	10	0.25	− 0.75	0.26	− 0.06
	50	0.07	− 0.08	0.08	− 0.01
	100	0.04	− 0.04	0.05	− 0.01
	200	0.02	− 0.01	0.02	− 0.00
0.25	10	0.13	− 0.19	0.15	− 0.02
	50	0.03	− 0.03	0.04	− 0.00
	100	0.01	− 0.02	0.02	− 0.00
	200	0.01	− 0.01	0.01	− 0.00

respectively. As long as $\lambda = 0$ (i.e., as long as the individual effects and the explanatory variable are uncorrelated), the estimator is consistent for β for any $\alpha \in [0, 1]$. If $\lambda \neq 0$, $\tilde{\beta}(\alpha)$ is inconsistent for β (when T is fixed) unless $\alpha = \tau$. Note that the inconsistencies do not disappear if $T \to \infty$ unless $\rho = 0$, i.e., unless the x_{it}'s are uncorrelated over time (conditional upon z_i) or unless $\alpha - \tau \to 0$ as $T \to \infty$ (which is the case for Deaton's errors-in-variables estimator).

The bias in $\tilde{\beta}(\alpha)$ may be very large if ω_1 is small relative to ω_2. Since ω_1 denotes the within variance of the true cohort means, it is thus important to choose the cohorts such that this variation (over time and over cohorts) is large. If the expectation of x_{it} does not depend on t (the μ_t's do not vary with t), this requires that there is time variation in the correlation between x_{it} and the cohort-identifying variable z_i. The extent to which an increasing cohort size n_c reduces the inconsistency, through its effect on ω_2, depends crucially on the ratio of ω_1 and σ_v^2. Note that both ω_1 and σ_v^2 are implied by the choice of the cohort-identifying variable and are independent of the number of observations in each cohort. Increasing n_c reduces the bias in all estimators, unless $\omega_1 = 0$.

In table 1 we present values[6] for the relative inconsistency $\delta(\alpha)/\beta$ of $\tilde{\beta}(\alpha)$ [apart from the factor[7] λ/β, i.e., we present $\delta(\alpha)/\lambda$] for $\omega_1/\sigma_v^2 = 0.025, 0.10$, and 0.25, $T = 2$ and 10, and $\rho = 0.5$, for different values of n_c and $\alpha = 0$ and 1. Note that the inconsistencies are increasing in ρ. From the table we can see that the

---

[6] A '—' in the tables indicates that $\omega_1 + (\tau - \alpha)\omega_2$ is not positive (definite), as was assumed above.

[7] It can be shown that λ/β is the relative bias in the between estimator for β on individual data and is an upper bound for the relative bias in OLS or GLS estimators [cf. Mundlak (1978)].

inconsistencies may be fairly large if the within variance of the true cohort means ω_1 is small relative to the measurement error variance σ_v^2 and if small cohort sizes are chosen. A larger value of T does not have much effect on the bias if $\alpha = 0$, but reduces the bias if $\alpha = 1$.

It is important to realize that a choice of α which reduces the bias may do so at the cost of a larger variance. To see this, we shall in the next section consider the variance and the MSE of our estimators $\tilde{\beta}(\alpha)$.

4. Minimum MSE estimation in a simplified case

In this section we shall derive the choice of α which minimizes the mean squared error of the estimator in the example of the previous section. Our results suggest that neither Deaton's errors-in-variables estimator ($\alpha = 1$) nor the adjusted errors-in-variables estimator ($\alpha = \tau$) are optimal in finite samples and that the optimal value of α is smaller than $\tau = (T - 1)/T$.

The asymptotic variance of $\tilde{\beta}(\alpha)$ can be written as

$$V\{\tilde{\beta}(\alpha)\} = \frac{1}{CT}(\omega_1 + (\tau - \alpha)\omega_2)^{-2}V^*, \tag{18}$$

where

$$V^* = \lim_{C \to \infty} V\left\{\frac{1}{\sqrt{CT}}\sum_{c=1}^{C}\sum_{t=1}^{T}(\bar{x}_{ct} - \bar{x}_c)(\bar{\theta}_{ct} - \bar{\theta}_c + \bar{\varepsilon}_{ct} - \bar{\varepsilon}_c) - \alpha\sigma\right\}. \tag{19}$$

Under the additional assumption of normality of the cohort aggregates $\bar{x}_{ct}$, $\bar{\theta}_{ct}$, and $\bar{\varepsilon}_{ct}$, one obtains that

$$V^* = (\omega_1 + \tau\omega_2)\left[n_c^{-1}(\sigma_\xi^2 + \sigma_\varepsilon^2 + \lambda^2 A\sigma_v^2)\right] + \tau n_c^{-2}\lambda^2 A^2 \sigma_v^4, \tag{20}$$

where $A = (1 + (T - 1)\rho)/T$. Since V^* does not depend on α, it immediately follows that the within estimator $\tilde{\beta}(0)$ has smallest variance, while Deaton's errors-in-variables estimator $\tilde{\beta}(1)$ has largest variance in our class.

Using the asymptotic bias and variance, we approximate the mean squared error (in finite samples) of $\tilde{\beta}(\alpha)$ by

$$\text{MSE}\{\tilde{\beta}(\alpha)\} \approx \widehat{\text{MSE}}\{\tilde{\beta}(\alpha)\} = V\{\tilde{\beta}(\alpha)\} + \delta(\alpha)^2. \tag{21}$$

Using (15) and (18) this can be written as

$$\widehat{\text{MSE}}\{\tilde{\beta}(\alpha)\} = \left(\lambda^2 A^2(\tau - \alpha)^2 + \frac{V^*}{CT}\frac{n_c^2}{\sigma_v^4}\right)(n_c\omega_1/\sigma_v^2 + (\tau - \alpha))^{-2}. \tag{22}$$

Obviously, if $\lambda = 0$ (i.e., if the individual effects are uncorrelated with the explanatory variables), minimum MSE is obtained for $\alpha = 0$, the within estimator on the pseudo panel of cohort data. Assuming $\lambda \neq 0$, we can minimize (22) with respect to α. As $\omega_1 + (\tau - \alpha)\omega_2 > 0$, we obtain that $\widetilde{MSE}\{\hat{\beta}(\alpha)\}$ is minimal if $\alpha = \alpha^*$, where

$$\alpha^* = \tau - \frac{1}{NT} V^* \frac{n_c^2}{\lambda^2 A^2 \sigma_v^2 \omega_1}. \tag{23}$$

Thus, a minimal mean squared error is obtained for $\alpha^{opt} = 0$ if $\alpha^* \leq 0$ or $\alpha^{opt} = \alpha^*$ if $\alpha^* > 0$. Note that $\alpha^* \leq \tau$, such that $\alpha > \tau$ will never be optimal. Consequently, in finite samples it may be advantageous in terms of MSE to choose α smaller than τ, even though the implied estimator will suffer from an inconsistency.

Using the expression for V^* one can derive that α^* is increasing in λ and NT as well as increasing in ω_1/σ_v^2. Since ω_1/σ_v^2 is an indicator for the quality of the cohort-identifying variable z [see also Verbeek and Nijman (1992a)], this latter point implies that it is unadvisable to eliminate much measurement error if cohorts are chosen poorly (low ω_1/σ_v^2). This is caused by the fact that the (asymptotic) bias is bounded [cf. (15)], while the variance may increase rapidly if ω_1/σ_v^2 decreases (in particular if α is large).

Empirically, it can make a huge difference in terms of mean squared error which value of α one is using. To illustrate this, we present numerical values of MSEs of the four estimators considered in this paper: the standard within estimator ($\alpha = 0$), Deaton's errors-in-variables estimator ($\alpha = 1$), the adjusted errors-in-variables estimator ($\alpha = \tau$), and the estimator using the optimal value of α ($\alpha = \alpha^{opt}$). For analyzing the relative performance of these estimators, two scaling parameters are irrelevant (corresponding to the scaling of y_{it} and x_{it}): $\sigma_\varepsilon^2 + \sigma_\xi^2$ and σ_v^2. In addition, we arbitrarily normalize all mean squared errors such that the MSE of the optimal-α-estimator equals one for $n_c = 50$. Consequently, relative MSEs depend on ω_1/σ_v^2, $\lambda^2 \sigma_v^2/(\sigma_\varepsilon^2 + \sigma_\xi^2)$, ρ, on sample sizes N and T, and on cohort size n_c.

Numerical values for relative MSEs are given in table 2 (for $N = 1000$) and table 3 ($N = 5000$), where we have taken the parameter values from table 1, with, in addition, $\lambda^2 \sigma_v^2/(\sigma_\varepsilon^2 + \sigma_\xi^2) = 0.5$. Smaller values of this parameter reduce the value of α^{opt} and deteriorate the relative performance of the adjusted errors-in-variables estimator, while larger values lead to smaller differences in the MSEs of the adjusted errors-in-variables estimator and the optimal-α estimator. The overall view is that both the within estimator on the pseudo panel ($\alpha = 0$) and the unadjusted errors-in-variables estimator ($\alpha = 1$) perform rather poorly, in particular for small cohort sizes n_c. The reason for this is that biases are relatively important and cause a substantial deterioration of the mean squared errors compared to the optimal-α estimator. The differences between the latter

Table 2

The relative MSE of $\tilde{\beta}(\alpha)$ in comparison with $\tilde{\beta}(\alpha^{\mathrm{opt}})$ with $n_c = 50$; $N = 1000$.

ω_1/σ_v^2	n_c	α^{opt}	$T = 2$			
			$\alpha = 0$	$\alpha = \alpha^{\mathrm{opt}}$	$\alpha = \tau$	$\alpha = 1$
0.025	10	0.417	3.805	1.975	2.634	—
	50	0.319	1.229	1.000	1.145	6.692
	100	0.197	0.885	0.855	0.959	1.992
	200	0	0.781	0.781	0.866	1.166
0.10	10	0.461	4.821	1.414	1.469	43.387
	50	0.363	1.159	1.000	1.027	1.731
	100	0.241	0.967	0.948	0.972	1.181
	200	0	0.921	0.921	0.945	1.018
0.25	10	0.470	3.568	1.182	1.196	8.029
	50	0.372	1.080	1.000	1.010	1.267
	100	0.250	0.987	0.977	0.987	1.069
	200	0.005	0.966	0.966	0.975	1.005
ω_1/σ_v^2	n_c	α^{opt}	$T = 10$			
			$\alpha = 0$	$\alpha = \alpha^{\mathrm{opt}}$	$\alpha = \tau$	$\alpha = 1$
0.025	10	0.858	10.813	2.526	2.954	15.951
	50	0.824	3.412	1.000	1.061	1.385
	100	0.782	1.667	0.787	0.824	0.925
	200	0.670	0.913	0.678	0.706	0.742
0.10	10	0.883	22.846	1.659	1.687	3.314
	50	0.849	3.046	1.000	1.010	1.093
	100	0.807	1.459	0.917	0.925	0.954
	200	0.723	0.994	0.875	0.883	0.895
0.25	10	0.888	20.006	1.297	1.303	1.892
	50	0.854	2.116	1.000	1.004	1.038
	100	0.812	1.233	0.963	0.966	0.978
	200	0.728	1.000	0.944	0.948	0.952

estimator and the adjusted errors-in-variables estimator ($\alpha = \tau$) are fairly small, in particular if n_c is not too small. If N increases, biases affect the MSE more heavily, causing a deterioration of the performance of the estimators with $\alpha = 0$ and $\alpha = 1$.

Summarizing the results, it has been shown that Deaton's (1985) errors-in-variables estimator ($\alpha = 1$) will never be optimal in terms of mean squared error and that the adjusted errors-in-variables estimator ($\alpha = \tau$) will be optimal in large samples only. In most cases an intermediate value of α leads to a minimal MSE. If λ is near zero, the within estimator on the pseudo panel ($\alpha = 0$) will be optimal, because in this case the bias will be negligible.

Table 3

The relative MSE of $\tilde{\beta}(\alpha)$ in comparison with $\tilde{\beta}(\alpha^{\text{opt}})$ with $n_c = 50$; $N = 5000$.

ω_1/σ_v^2	n_c	α^{opt}	$T = 2$			
			$\alpha = 0$	$\alpha = \alpha^{\text{opt}}$	$\alpha = \tau$	$\alpha = 1$
0.025	10	0.483	16.045	2.219	2.367	—
	50	0.464	3.424	1.000	1.029	18.640
	100	0.439	1.585	0.841	0.862	3.566
	200	0.390	0.936	0.761	0.778	1.399
0.10	10	0.492	21.034	1.427	1.438	189.31
	50	0.473	2.348	1.000	1.005	3.507
	100	0.448	1.279	0.947	0.951	1.563
	200	0.399	0.989	0.920	0.924	1.093
0.25	10	0.494	14.404	1.183	1.186	32.408
	50	0.474	1.650	1.000	1.002	1.936
	100	0.450	1.129	0.977	0.979	1.223
	200	0.401	0.996	0.966	0.968	1.037

ω_1/σ_v^2	n_c	α^{opt}	$T = 10$			
			$\alpha = 0$	$\alpha = \alpha^{\text{opt}}$	$\alpha = \tau$	$\alpha = 1$
0.025	10	0.892	51.051	2.726	2.818	44.778
	50	0.885	14.910	1.000	1.012	1.825
	100	0.876	6.250	0.779	0.786	0.998
	200	0.859	2.418	0.668	0.674	0.736
0.10	10	0.897	111.46	1.668	1.674	8.174
	50	0.890	12.230	1.000	1.000	1.249
	100	0.881	4.145	0.916	0.918	0.987
	200	0.865	1.719	0.874	0.876	0.897
0.25	10	0.898	96.916	1.298	1.299	3.793
	50	0.891	7.065	1.000	1.001	1.106
	100	0.882	2.556	0.963	0.963	0.993
	200	0.866	1.341	0.944	0.944	0.954

5. Concluding remarks

In this paper attention is paid to the estimation of a fixed effects regression model from repeated cross-sections using a cohort approach. In this approach the individual observations are grouped into cohorts based on one or more time-invariant characteristics, like date of birth or sex, which results in a pseudo panel of cohort data. As shown by Deaton (1985), if the individual effects are correlated with the explanatory variables, the standard within estimator based on the pseudo panel, will be inconsistent, a problem which can be viewed as an errors-in-variables problem.

In this paper we introduce a class of estimators including the errors-in-variables estimator proposed by Deaton (1985) and the standard within estimator. It is shown that for a fixed number of cross-sections (T) both these estimators are inconsistent. In addition, from this class a consistent estimator (for fixed T) is derived, which boils down to performing a within transformation on the pseudo panel (i.e., taking deviations from cohort means) and adjusting the moments matrices in the least squares estimator by eliminating a fraction $\tau = (T - 1)/T$ of the (estimated) measurement error variance.

Our class of estimators is indexed by the fraction (α) of the measurement error variance that is eliminated ($\alpha \in [0, 1]$). The larger this fraction is, the smaller is the variance of the resulting estimator. In terms of mean squared error this implies that it is optimal to eliminate a fraction smaller than $(T - 1)/T$ of the measurement error. Since this optimal fraction depends on unknown parameters, choosing $\alpha = (T - 1)/T$ might be tempting in applications. The difference between this adjusted errors-in-variables estimator and the optimal-α-estimator in terms of mean squared error was fairly small in the example considered in sections 3 and 4. Although Deaton's errors-in-variables estimator (corresponding with $\alpha = 1$) is consistent if the number of cross-sections T tends to infinity, it does not perform very well in terms of mean squared error, since both the bias and the variance are larger than for the adjusted errors-in-variables estimator [$\alpha = (T - 1)/T$]. Even for moderate T the difference can be substantial.

References

Angrist, J.D., 1991, Grouped-data estimation and testing in simple labor-supply models, Journal of Econometrics 47, 243–266.

Browning M., A. Deaton, and M. Irish, 1985, A profitable approach to labor supply and commodity demands over the life cycle, Econometrica 53, 503–543.

Collado, M.D., 1991, Estimating dynamic models from time series of cross-sections, Paper presented at the Econometric Society European Meeting, Cambridge, 1991.

Deaton, A., 1985, Panel data from time series of cross-sections, Journal of Econometrics 30, 109–126.

Fuller, W.A., 1987, Measurement error models (Wiley, New York, NY).

Heckman, J.J. and T.E. MaCurdy, 1980, A life cycle model of female labour supply, Review of Economic Studies 47, 47–74.

Heckman, J.J. and R. Robb, 1985, Alternative models for evaluating the impact of interventions: An overview, Journal of Econometrics 30, 239–267.

Hsiao, C., 1986, Analysis of panel data (Cambridge University Press, Cambridge).

MaCurdy, T.E., 1981, An empirical model of labor supply in a life-cycle setting, Journal of Political Economy 89, 1059–1085.

Moffitt, R., 1993, Identification and estimation of dynamic models with a time series of repeated cross-sections, Journal of Econometrics, this issue.

Mundlak, Y., 1978, On the pooling of time series and cross section data, Econometrica 46, 69–85.

Verbeek, M. and T.E. Nijman, 1992a, Can cohort data be treated as genuine panel data?, Empirical Economics 17, 9–23.

Verbeek, M. and T.E. Nijman, 1992b, Minimum MSE estimation of a regression model with fixed effects from a series of cross sections, CentER discussion paper no. 9201 (CentER, Tilburg University, Tilburg).