Minimum MSE estimation of a regression model with fixed effects from a series of cross-sections
Verbeek, M.J.C.M.; Nijman, T.E.

Publication date: 1994

Citation for published version (APA):
Minimum MSE Estimation of a Regression Model with Fixed Effects from a Series of Cross-sections

by

Marno Verbeek and Theo Nijman


Reprint Series no. 148
CENTER FOR ECONOMIC RESEARCH

Board
Harry Barkema
Helmut Bester
Eric van Damme, Chairman
Frank van der Duyn Schouten
Jeffrey James

Management
Jeffrey James (Director of Graduate Studies)
Arie Kapteyn (Scientific Director)
Marie-Louise Kemperman (Managing Director)

Scientific Council
Anton Barten
Eduard Bomhoff
Willem Buiter
Jacques Drèze
Jack Kleijn
Theo van de Klundert
Jean-Jacques Laffont
Merton Miller
Piet Moerland
Philippe Naert
Pieter Ruys
Université Catholique de Louvain
Erasmus University Rotterdam
Yale University
Université Catholique de Louvain
Tilburg University
Tilburg University
Université des Sciences Sociales de Toulouse
University of Chicago
Tilburg University
Nijenrode University
Tilburg University

Residential Fellows
Hans Bloemen
Lans Bovenberg
Hans Carlsson
Jay Pil Choi
Jan Magnus
Andrew Mountford
Bezalel Peleg
Mark Steel
Frank Verboven
Oscar Volij
Karl-Erik Wärneryd
CentER
CentER, Erasmus University Rotterdam
Gothenburg University and Lund University
Columbia University
CentER, LSE
CentER
Hebrew University of Jerusalem
CentER/Department of Econometrics, Tilburg University
CentER
Hebrew University of Jerusalem
Stockholm School of Economics

Address : P.O. Box 90153, 5000 LE Tilburg, The Netherlands
Phone : +31 13 663102
Telex : 52426 kub nl
Telefax : +31 13 663066
E-mail : center@kub.nl

ISSN 0924-7874

1994
Minimum MSE Estimation of a Regression Model with Fixed Effects from a Series of Cross-sections

by Marno Verbeek and Theo Nijman

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

* The authors thank Bertrand Melenberg, Arthur van Soest, Guglielmo Weber, two anonymous referees, and participants of the European Meeting of the Econometric Society, Cambridge, 1991, for helpful comments. Financial support by the Royal Netherlands Academy of Arts and Sciences (K.N.A.W.) is gratefully acknowledged.

0304-4076/93/$06.00 © 1993—Elsevier Science Publishers B.V. All rights reserved
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:

$$y_{it} = x_{it} \beta + \theta_i + \epsilon_{it}, \quad t = 1, \ldots, T,$$

where $i$ indexes individuals, $x_{it}$ is a $k$-dimensional row vector of exogenous variables, and interest lies in the $k$-dimensional parameter vector $\beta$. Throughout this paper the error term $\epsilon_{it}$ is assumed to be uncorrelated with the explanatory variables in $x_{it}$ and the individual effects $\theta_i$. 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 ($i = 1, \ldots, NT$).

In many applications, for example in life cycle models, economic theory suggests that the individual effects $\theta_i$ are correlated with the explanatory variables. Consequently, regression estimators which simply include $\theta_i$ in the error term ('random effects estimators') will be inconsistent. Moreover, a fixed effect treatment of the $\theta_i$ is not possible because repeated observations on the same individuals are not available. Deaton (1985) has shown how $\beta$ 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,

\[ \hat{y}_{ct} = \tilde{x}_{ct}\beta + \tilde{\theta}_{ct} + \tilde{\epsilon}_{ct}, \quad c = 1, \ldots, C, \quad t = 1, \ldots, T. \] (2)

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

\[ y^*_c = x^*_c\beta + \theta_c + \epsilon^*_c, \quad c = 1, \ldots, C, \quad t = 1, \ldots, T, \] (3)

where the asterisks denote unobservable population cohort means and where \( \theta_c \) 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 \( \tilde{y}_{ct} \) and \( \tilde{x}_{ct} \) can be considered as error-ridden measurements of \( y^*_c \) and \( x^*_c \). It is assumed that the measurement errors are normally distributed with zero mean, independent of the true values,

\[ \begin{pmatrix} \tilde{y}_{ct} - y^*_c \\ \tilde{x}_{ct} - x^*_c \end{pmatrix} \sim N \begin{pmatrix} 0 \\ \begin{pmatrix} \sigma_{00} & \sigma' \\ \sigma & \Omega_2 \end{pmatrix} \end{pmatrix}. \] (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, \( \sigma_{00}, \sigma, \) and \( \Omega_2 \) can be estimated consistently from the individual observations. Without affecting our main conclusions we shall in the sequel assume that \( \Omega_2 \) and \( \sigma \) are known.

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

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

where

\[ M_{xx} = \frac{1}{CT} \sum_{c=1}^{C} \sum_{t=1}^{T} (\tilde{x}_{ct} - \bar{x}_c)(\tilde{x}_{ct} - \bar{x}_c), \] (6)

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

with \( \bar{x}_c = (1/T)\sum_{t=1}^{T} \tilde{x}_{ct} \). 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 $\hat{\beta}(\alpha)$ in more detail, and, from these results, derive some conclusions on the optimal choice of $\alpha$.

Because of the stochastic nature of $M_{xx}$, large sample theory appears to be required$^1$ to approximate the behaviour of $\hat{\beta}(\alpha)$. 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 $\epsilon_{ct}^* = 0$). The latter choice eliminates the measurement error problem and makes the value of $\alpha$ (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 $\Omega_1$ as the (asymptotic) within cohort variation in the true cohort means,

$$\Omega_1 = \lim_{C \to \infty} \frac{1}{CT} \sum_{\tau=1}^{T} \sum_{t=1}^{T} (x_{c\tau}^* - x_{ct}^*)(x_{c\tau}^* - x_{ct}^*),$$

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

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

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

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

\(^1\) Note that the expectation of $\hat{\beta}(\alpha)$ 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

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

Whether or not $\tilde{\beta}(\alpha)$ is consistent for $\beta$ depends on the value of $\alpha$ as well as on the relationship between $\Omega_2$ and $\sigma$.

If $\theta_i$ and $x_{it}$ are uncorrelated, it is easily verified that $\sigma = \Omega_2\beta$. Consequently, $\tilde{\beta}(\alpha)$ is consistent for any value of $\alpha$. On the other hand, if $\theta_i$ 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 $\theta$ cannot be estimated consistently), which carries over to the parameters of interest $\beta$, 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 $\alpha = 1$. Only if 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 $\alpha$ 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 $\alpha$. 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 $\alpha$ and the inconsistency has a minimum at $\alpha = (T - 1)/T$, there is room for a trade-off between variance and asymptotic bias, and it is not unlikely that an estimator with a value of $\alpha$ smaller than $\tau$ will perform better in terms of mean squared error. The results of Fuller (1987, p. 163ff.) also suggest that a smaller value of $\alpha$ 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 $\alpha$ 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 $\alpha$ and the cohort size $n_c$. The model considered in this section is

$$y_{it} = x_{it}\beta + \theta_i + \epsilon_{it}, \quad t = 1, \ldots, T,$$  \tag{12}$$
where now $x_{it}$ is only one-dimensional. As before, the individual effects $\theta_i$ may be correlated with the explanatory variable $x_{it}$. In particular, we shall impose the assumption [based on Mundlak (1978)] that the individual effects $\theta_i$ are correlated with the $x$'s in the following way:

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

(13)

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

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 $z$ 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 $z_t$ according to

$$x_{it} = \mu_t + \gamma_t z_t + u_{it},$$

(14)

where $E \{ u_{it} | z_i \} = 0$, $\text{cov} \{ u_{it}, u_{is} \} = \rho \sigma^2$. The $\mu_t$ are fixed unknown constants. The error term $u_{it}$ is assumed to be uncorrelated with $z_t$.

Under the assumptions above, it can be shown that the inconsistency in $\tilde{\beta}(x)$ is given by

$$\delta(x) = \lim_{c \to \infty} (\tilde{\beta}(x) - \beta) = \lambda \frac{1 + (T - 1)\rho}{T} \frac{(\tau - \alpha)\omega_2}{(\omega_1 + (\tau - \alpha)\omega_2)}.$$

(15)

where $\omega_1$ and $\omega_2$ are given by

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

(16)

$$\omega_2 = \lim_{c \to \infty} \frac{1}{CT} \sum_{c-1}^{C} \sum_{t=1}^{T} (\bar{x}_{it} - \bar{x}_{ct})^2 = n_c^{-1} \sigma^2_{\nu},$$

(17)

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.

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

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

In this section lower-case letters are used to stress that $\omega_1$ and $\omega_2$ are scalar.
Table 1
Relative inconsistencies of $\beta(\alpha)$, apart from $\lambda/\beta$.

<table>
<thead>
<tr>
<th>$\omega_1/\sigma^2_w$</th>
<th>$n_c$</th>
<th>$T = 2$</th>
<th>$T = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\alpha = 0$</td>
<td>$\alpha = 1$</td>
</tr>
<tr>
<td>0.025</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.50</td>
<td>- 0.50</td>
<td>0.43</td>
</tr>
<tr>
<td>50</td>
<td>0.21</td>
<td>- 0.19</td>
<td>0.23</td>
</tr>
<tr>
<td>100</td>
<td>0.13</td>
<td>- 0.08</td>
<td>0.15</td>
</tr>
<tr>
<td>200</td>
<td>0.07</td>
<td>- 0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.25</td>
<td>- 0.75</td>
<td>0.26</td>
</tr>
<tr>
<td>50</td>
<td>0.07</td>
<td>- 0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>100</td>
<td>0.04</td>
<td>- 0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>200</td>
<td>0.02</td>
<td>- 0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.13</td>
<td>- 0.19</td>
<td>0.15</td>
</tr>
<tr>
<td>50</td>
<td>0.03</td>
<td>- 0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>100</td>
<td>0.01</td>
<td>- 0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>200</td>
<td>0.01</td>
<td>- 0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

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 $\beta$ for any $\alpha \in [0, 1]$. If $\lambda \neq 0$, $\tilde{\beta}(\alpha)$ is inconsistent for $\beta$ (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 $\omega_1$ is small relative to $\omega_2$. Since $\omega_1$ 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 $\mu_i$'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 $\omega_2$, depends crucially on the ratio of $\omega_1$ and $\sigma^2_v$. Note that both $\omega_1$ and $\sigma^2_v$ 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$^7$ $\lambda/\beta$, i.e., we present $\delta(\alpha)/\lambda$] for $\omega_1/\sigma^2_v = 0.025, 0.10, \text{ and } 0.25$, $T = 2$ and 10, and $\rho = 0.5$, for different values of $n_c$ and $\alpha = 0$ and 1. Note that the inconsistencies are increasing in $\rho$. 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 $\lambda/\beta$ is the relative bias in the between estimator for $\beta$ 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 \( \omega_1 \) is small relative to the measurement error variance \( \sigma_e^2 \) 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 \( \alpha \) 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 \( \alpha \) 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 \( \alpha \) is smaller than \( \tau = (T - 1)/T \).

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

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

where

\[
V^* = \lim_{c \to \infty} V\left\{ \frac{1}{\sqrt{CT}} \sum_{c=1}^{C} \sum_{r=1}^{T} (\bar{x}_{ct} - \bar{x}_c)(\bar{\beta}_{ct} - \bar{\beta}_c + \bar{\varepsilon}_{ct} - \bar{\varepsilon}_c) - \alpha \sigma \right\}.
\]

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

\[
V^* = (\omega_1 + \tau \omega_2) [n_e^{-1}(\sigma_e^2 + \sigma_c^2 + \lambda^2 A \sigma_v^2)] + \tau n_e^{-2} \lambda^2 A^2 \sigma_v^4,
\]

where \( A = 1 + (T - 1)\rho)/T \). Since \( V^* \) does not depend on \( \alpha \), 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

\[
\text{MSE}\{\tilde{\beta}(\alpha)\} \approx \text{MSE}\{\tilde{\beta}(\alpha)\} = V\{\tilde{\beta}(\alpha)\} + \delta(\alpha)^2.
\]

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

\[
\text{MSE}\{\tilde{\beta}(\alpha)\} = \left( \lambda^2 A^2 (\tau - \alpha)^2 + \frac{V^* n_e^2}{C T \sigma_v^4} \right)(n_e \omega_1/\sigma_e^2 + (\tau - \alpha)^{-2}).
\]
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 $\alpha$. As $\omega_1 + (\tau - \alpha)\omega_2 > 0$, we obtain that $\text{MSE} \{ \hat{\beta}(\alpha) \}$ is minimal if $\alpha = \alpha^*$, where

$$\alpha^* = \tau - \frac{1}{NT} V^* \frac{n_c^2}{\lambda^2 A^2 \sigma_\epsilon^2 \omega_1}.$$ \hspace{1cm} (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 $\alpha$ smaller than $\tau$, even though the implied estimator will suffer from an inconsistency.

Using the expression for $V^*$ one can derive that $\alpha^*$ is increasing in $\lambda$ and $NT$ as well as increasing in $\omega_1/\sigma_\epsilon^2$. Since $\omega_1/\sigma_\epsilon^2$ 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 $\omega_1/\sigma_\epsilon^2$). This is caused by the fact that the (asymptotic) bias is bounded [cf. (15)], while the variance may increase rapidly if $\omega_1/\sigma_\epsilon^2$ decreases (in particular if $\alpha$ is large).

Empirically, it can make a huge difference in terms of mean squared error which value of $\alpha$ 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 $\alpha$ ($\alpha = \alpha^{opt}$). For analyzing the relative performance of these estimators, two scaling parameters are irrelevant (corresponding to the scaling of $y_t$ and $x_{it}$): $\sigma_\epsilon^2 + \sigma_\tau^2$ and $\sigma_\tau^2$. In addition, we arbitrarily normalize all mean squared errors such that the MSE of the optimal-$\alpha$ estimator equals one for $n_c = 50$. Consequently, relative MSEs depend on $\omega_1/\sigma_\epsilon^2$, $\lambda^2 \sigma_\tau^2/(\sigma_\epsilon^2 + \sigma_\tau^2)$, $\rho$, 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, $\lambda^2 \sigma_\tau^2/(\sigma_\epsilon^2 + \sigma_\tau^2) = 0.5$. Smaller values of this parameter reduce the value of $\alpha^{opt}$ 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-$\alpha$ 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-$\alpha$ estimator. The differences between the latter
### Table 2

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

<table>
<thead>
<tr>
<th>$\omega_1/\sigma^2$</th>
<th>$n_i$</th>
<th>$\alpha^{opt}$</th>
<th>$\alpha = 0$</th>
<th>$\alpha = \alpha^{opt}$</th>
<th>$\alpha = \tau$</th>
<th>$\alpha = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$T = 2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.025</td>
<td>10</td>
<td>0.417</td>
<td>3.805</td>
<td>1.975</td>
<td>2.634</td>
<td>6.692</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.319</td>
<td>1.229</td>
<td>1.000</td>
<td>1.145</td>
<td>2.634</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.197</td>
<td>0.885</td>
<td>0.855</td>
<td>0.959</td>
<td>1.992</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0</td>
<td>0.781</td>
<td>0.781</td>
<td>0.866</td>
<td>1.186</td>
</tr>
<tr>
<td>0.10</td>
<td>10</td>
<td>0.461</td>
<td>4.821</td>
<td>1.414</td>
<td>1.469</td>
<td>43.387</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.363</td>
<td>1.159</td>
<td>1.000</td>
<td>1.027</td>
<td>1.731</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.241</td>
<td>0.967</td>
<td>0.948</td>
<td>0.972</td>
<td>1.181</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0</td>
<td>0.921</td>
<td>0.921</td>
<td>0.945</td>
<td>1.018</td>
</tr>
<tr>
<td>0.25</td>
<td>10</td>
<td>0.470</td>
<td>3.568</td>
<td>1.182</td>
<td>1.196</td>
<td>8.029</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.372</td>
<td>1.080</td>
<td>1.000</td>
<td>1.010</td>
<td>1.267</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.250</td>
<td>0.987</td>
<td>0.977</td>
<td>0.987</td>
<td>1.069</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.005</td>
<td>0.966</td>
<td>0.966</td>
<td>0.975</td>
<td>1.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$T = 10$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.025</td>
<td>10</td>
<td>0.858</td>
<td>10.813</td>
<td>2.526</td>
<td>2.954</td>
<td>15.951</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.824</td>
<td>3.412</td>
<td>1.000</td>
<td>1.061</td>
<td>1.385</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.782</td>
<td>1.667</td>
<td>0.787</td>
<td>0.824</td>
<td>0.925</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.670</td>
<td>0.913</td>
<td>0.678</td>
<td>0.706</td>
<td>0.742</td>
</tr>
<tr>
<td>0.10</td>
<td>10</td>
<td>0.883</td>
<td>22.846</td>
<td>1.659</td>
<td>1.687</td>
<td>3.314</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.849</td>
<td>3.046</td>
<td>1.000</td>
<td>1.010</td>
<td>1.093</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.807</td>
<td>1.459</td>
<td>0.917</td>
<td>0.925</td>
<td>0.954</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.723</td>
<td>0.994</td>
<td>0.875</td>
<td>0.883</td>
<td>0.895</td>
</tr>
<tr>
<td>0.25</td>
<td>10</td>
<td>0.888</td>
<td>20.006</td>
<td>1.297</td>
<td>1.303</td>
<td>1.892</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.854</td>
<td>2.116</td>
<td>1.000</td>
<td>1.004</td>
<td>1.038</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.812</td>
<td>1.233</td>
<td>0.963</td>
<td>0.966</td>
<td>0.978</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.728</td>
<td>1.000</td>
<td>0.944</td>
<td>0.948</td>
<td>0.952</td>
</tr>
</tbody>
</table>

Estimator and the adjusted errors-in-variables estimator ($\alpha = \tau$) are fairly small, in particular if $n_i$ 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 $\alpha$ leads to a minimal MSE. If $\lambda$ is near zero, the within estimator on the pseudo panel ($\alpha = 0$) will be optimal, because in this case the bias will be negligible.
The relative MSE of $\hat{\beta}(x)$ in comparison with $\tilde{\beta}(x^{\text{rep}})$ with $n_e = 50; N = 5000$.

### Table 3

<table>
<thead>
<tr>
<th>$\omega_1/\sigma^2_n$</th>
<th>$n_e$</th>
<th>$x^{\text{rep}}$</th>
<th>$\alpha = 0$</th>
<th>$\alpha = x^{\text{rep}}$</th>
<th>$\alpha = \tau$</th>
<th>$\alpha = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>T = 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.025</td>
<td>10</td>
<td>0.483</td>
<td>16.045</td>
<td>2.219</td>
<td>2.367</td>
<td>18.640</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.464</td>
<td>3.424</td>
<td>1.000</td>
<td>1.029</td>
<td>3.566</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.439</td>
<td>1.585</td>
<td>0.841</td>
<td>0.862</td>
<td>0.778</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.390</td>
<td>0.936</td>
<td>0.761</td>
<td>0.778</td>
<td>1.399</td>
</tr>
<tr>
<td>0.10</td>
<td>10</td>
<td>0.492</td>
<td>21.034</td>
<td>1.427</td>
<td>1.438</td>
<td>189.31</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.473</td>
<td>2.348</td>
<td>1.000</td>
<td>1.005</td>
<td>3.507</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.448</td>
<td>1.279</td>
<td>0.947</td>
<td>0.951</td>
<td>1.563</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.399</td>
<td>0.989</td>
<td>0.920</td>
<td>0.924</td>
<td>1.093</td>
</tr>
<tr>
<td>0.25</td>
<td>10</td>
<td>0.494</td>
<td>14.404</td>
<td>1.183</td>
<td>1.186</td>
<td>32.048</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.474</td>
<td>1.650</td>
<td>1.000</td>
<td>1.002</td>
<td>1.936</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.450</td>
<td>1.129</td>
<td>0.977</td>
<td>0.979</td>
<td>1.223</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.401</td>
<td>0.996</td>
<td>0.966</td>
<td>0.968</td>
<td>1.037</td>
</tr>
<tr>
<td><strong>T = 10</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.025</td>
<td>10</td>
<td>0.892</td>
<td>51.051</td>
<td>2.726</td>
<td>2.818</td>
<td>44.778</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.885</td>
<td>14.910</td>
<td>1.000</td>
<td>1.012</td>
<td>1.825</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.876</td>
<td>6.250</td>
<td>0.779</td>
<td>0.786</td>
<td>0.998</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.859</td>
<td>2.418</td>
<td>0.668</td>
<td>0.674</td>
<td>0.736</td>
</tr>
<tr>
<td>0.10</td>
<td>10</td>
<td>0.897</td>
<td>111.46</td>
<td>1.668</td>
<td>1.674</td>
<td>8.174</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.890</td>
<td>12.230</td>
<td>1.000</td>
<td>1.000</td>
<td>1.249</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.881</td>
<td>4.145</td>
<td>0.916</td>
<td>0.918</td>
<td>0.987</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.865</td>
<td>1.719</td>
<td>0.874</td>
<td>0.876</td>
<td>0.897</td>
</tr>
<tr>
<td>0.25</td>
<td>10</td>
<td>0.898</td>
<td>96.916</td>
<td>1.298</td>
<td>1.299</td>
<td>3.793</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.891</td>
<td>7.065</td>
<td>1.000</td>
<td>1.001</td>
<td>1.106</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.882</td>
<td>2.556</td>
<td>0.963</td>
<td>0.963</td>
<td>0.993</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.866</td>
<td>1.341</td>
<td>0.944</td>
<td>0.944</td>
<td>0.954</td>
</tr>
</tbody>
</table>

### 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 ($\alpha$) 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-$\alpha$ 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


Reprint Series, CentER, Tilburg University, The Netherlands:


No. 5  Th. ten Raa and F. van der Ploeg, A statistical approach to the problem of negatives in input-output analysis, *Economic Modelling*, vol. 6, no. 1, 1989, pp. 2 - 19.


No. 8  Th. van de Klundert and F. van der Ploeg, Wage rigidity and capital mobility in an optimizing model of a small open economy, *De Economist*, vol. 137, nr. 1, 1989, pp. 47 - 75.


<table>
<thead>
<tr>
<th>No.</th>
<th>Authors</th>
<th>Title and Details</th>
</tr>
</thead>
</table>


<table>
<thead>
<tr>
<th>No.</th>
<th>Authors</th>
<th>Title</th>
<th>Journal</th>
<th>Volume</th>
<th>Issue</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>95</td>
<td>A. van den Nouweland, P. Borm and S. Tijs</td>
<td>Allocation rules for hypergraph communication situations</td>
<td>International Journal of Game Theory</td>
<td>20</td>
<td>3</td>
<td>255 - 268</td>
</tr>
<tr>
<td>96</td>
<td>E.J. Bonhoff</td>
<td>Monetary reform in Eastern Europe</td>
<td>European Economic Review</td>
<td>36</td>
<td>2/3</td>
<td>454 - 458</td>
</tr>
<tr>
<td>97</td>
<td>F. van der Ploeg and A. de Zeeuw</td>
<td>International aspects of pollution control</td>
<td>Environmental and Resource Economics</td>
<td>2</td>
<td>2</td>
<td>117 - 139</td>
</tr>
<tr>
<td>98</td>
<td>P.E.M. Borm and S.H. Tijs</td>
<td>Strategic claim games corresponding to an NTU-game</td>
<td>Games and Economic Behavior</td>
<td>4</td>
<td>1</td>
<td>58 - 71</td>
</tr>
<tr>
<td>99</td>
<td>A. van Soest and P. Kooreman</td>
<td>Coherency of the indirect translog demand system with binding nonnegativity constraints</td>
<td>Journal of Econometrics</td>
<td>44</td>
<td>3</td>
<td>391 - 400</td>
</tr>
<tr>
<td>100</td>
<td>Th. ten Raa and E.N. Wolff</td>
<td>Secondary products and the measurement of productivity growth</td>
<td>Regional Science and Urban Economics</td>
<td>21</td>
<td>4</td>
<td>581 - 615</td>
</tr>
<tr>
<td>101</td>
<td>P. Kooreman and A. Kapteyn</td>
<td>On the empirical implementation of some game theoretic models of household labor supply</td>
<td>The Journal of Human Resources</td>
<td>25</td>
<td>4</td>
<td>584 - 598</td>
</tr>
<tr>
<td>102</td>
<td>H. Bester</td>
<td>Bertrand equilibrium in a differentiated duopoly</td>
<td>International Economic Review</td>
<td>33</td>
<td>2</td>
<td>433 - 448</td>
</tr>
<tr>
<td>103</td>
<td>J.A.M. Potters and S.H. Tijs</td>
<td>The nucleolus of a matrix game and other nucleoli</td>
<td>Mathematics of Operations Research</td>
<td>17</td>
<td>1</td>
<td>164 - 174</td>
</tr>
<tr>
<td>104</td>
<td>A. Kapteyn, P. Kooreman and A. van Soest</td>
<td>Quantity rationing and concavity in a flexible household labor supply model</td>
<td>Review of Economics and Statistics</td>
<td>72</td>
<td>1</td>
<td>55 - 62</td>
</tr>
<tr>
<td>105</td>
<td>A. Kapteyn and P. Kooreman</td>
<td>Household labor supply: What kind of data can tell us how many decision makers there are?</td>
<td>European Economic Review</td>
<td>36</td>
<td>2/3</td>
<td>365 - 371</td>
</tr>
<tr>
<td>106</td>
<td>Th. van de Klundert and S. Smulders</td>
<td>Reconstructing growth theory: A survey</td>
<td>De Economist</td>
<td>140</td>
<td>2</td>
<td>177 - 203</td>
</tr>
<tr>
<td>107</td>
<td>N. Rankin</td>
<td>Imperfect competition, expectations and the multiple effects of monetary growth</td>
<td>The Economic Journal</td>
<td>102</td>
<td>413</td>
<td>743 - 753</td>
</tr>
<tr>
<td>109</td>
<td>S. van Wijnbergen</td>
<td>Trade reform, policy uncertainty, and the current account: A non-expected-utility approach</td>
<td>American Economic Review</td>
<td>82</td>
<td>3</td>
<td>626 - 633</td>
</tr>
<tr>
<td>No.</td>
<td>Author(s)</td>
<td>Title</td>
<td>Journal/Book</td>
<td>Pages</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>----------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
<td>-------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>113</td>
<td>P.J. Deschamps</td>
<td>Expectations and intertemporal separability in an empirical model of consumption and investment under uncertainty</td>
<td><em>Empirical Economics</em>, vol. 17, no. 3, 1992</td>
<td>419 - 450</td>
<td></td>
<td></td>
</tr>
<tr>
<td>115</td>
<td>G.W. Imbens</td>
<td>An efficient method of moments estimator for discrete choice models with choice-based sampling</td>
<td><em>Econometrica</em>, vol. 60, no. 5, 1992</td>
<td>1187 - 1214</td>
<td></td>
<td></td>
</tr>
<tr>
<td>116</td>
<td>P. Borm</td>
<td>On perfectness concepts for bimatrix games</td>
<td><em>OR Spektrum</em>, vol. 14, no. 1, 1992</td>
<td>33 - 42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>K. Wärneryd</td>
<td>Communication, correlation and symmetry in bargaining</td>
<td><em>Economics Letters</em>, vol. 39, no. 3, 1992</td>
<td>295 - 300</td>
<td></td>
<td></td>
</tr>
<tr>
<td>121</td>
<td>M.R. Baye, D. Kovenock and C.G. de Vries</td>
<td>It takes two to tango: equilibria in a model of sales</td>
<td><em>Games and Economic Behavior</em>, vol. 4, no. 4, 1992</td>
<td>493 - 510</td>
<td></td>
<td></td>
</tr>
<tr>
<td>123</td>
<td>S. van Wijnbergen</td>
<td>Intertemporal speculation, shortages and the political economy of price reform</td>
<td><em>The Economic Journal</em>, vol. 102, no. 415, 1992</td>
<td>1395 - 1406</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


