

DISCUSSION PAPER SERIES

No. 6167

IMPROVED ERRORS-IN-VARIABLES ESTIMATORS FOR GROUPED DATA

Paul J. Devereux

LABOUR ECONOMICS



Centre for **E**conomic **P**olicy **R**esearch

www.cepr.org

Available online at:

www.cepr.org/pubs/dps/DP6167.asp

IMPROVED ERRORS-IN-VARIABLES ESTIMATORS FOR GROUPED DATA

Paul J. Devereux, University College Dublin and CEPR

Discussion Paper No. 6167
March 2007

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **LABOUR ECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Economic and Social Research Council, under which an ESRC Resource Centre operates within CEPR; the Esmée Fairbairn Charitable Trust; and the Bank of England. These organizations do not give prior review to the Centre's publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Paul J. Devereux

ABSTRACT

Improved Errors-in-Variables Estimators for Grouped Data*

In many economic applications, observations are naturally categorized into mutually exclusive and exhaustive groups. For example, individuals can be classified into cohorts and workers are employees of a particular firm. Grouping models are widely used in economics -- for example, cohort models have been used to study labour supply, wage inequality, consumption, and intergenerational transfer of human capital. The simplest grouping estimator involves taking the means of all variables for each group and then carrying out a group-level regression by OLS or weighted least squares. This estimator is biased in finite samples. I show that the standard errors in variables estimator (EVE) designed to correct for small sample bias is exactly equivalent to the Jack-knife Instrumental Variables Estimator (JIVE). Also EVE is closely related to the k-class of instrumental variables estimators. I then use results from the instrumental variables literature to develop an estimator (UEVE) with better finite-sample properties than existing errors in variables estimators. The theoretical results are demonstrated using Monte Carlo experiments. Finally, I use the estimators to implement a model of inter-temporal male labour supply using micro data from the United States Census. There are sizeable differences in the wage elasticity across estimators, showing the practical importance of the theoretical issues discussed in this paper even in circumstances where the sample size is quite large.

JEL Classification: c21 and j22

Keywords: errors-in-variables and grouped data

Paul J. Devereux
UCD School of Economics
Newman Building
University College Dublin
Belfield, Dublin 4
IRELAND
Email: Devereux@ucd.ie

* I thank Dan Akerberg, Joshua Angrist, Sandy Black, Janet Currie, Jin Hahn, Katerina Kyriazidou, Robert Moffitt, Donal O'Neill, Olive Sweetman, and Gautam Tripathi for helpful comments.

Submitted 14 February 2007

1 Introduction

In many economic applications, observations are naturally categorized into mutually exclusive and exhaustive groups. For example, individuals can be classified into cohorts and workers are employees of a particular firm. The simplest grouping estimator involves taking the means of all variables for each group and then carrying out a group-level regression by OLS or weighted least squares (if there are different numbers of observations in different groups). This estimator has been called the efficient Wald estimator (Angrist, 1991). For brevity, I refer to it as the EWALD estimator in this paper. Grouping estimators have been used in recent years to study labor supply (Angrist 1991, Blundell, Duncan and Meghir 1998), consumption (Mckenzie 2001), wage inequality (Card and Lemieux 1996), intergenerational transfers of human capital (Acemoglu and Pischke 2001), and many other topics.

Deaton (1985) points out that EWALD is biased in finite samples and proposes an errors-in-variables estimator (EVE) to correct for the effects of sampling error. The first contribution of this paper is to analyze the relationship between errors-in-variables estimators and bias-corrected instrumental variables estimators. These two types of estimators have been developed in separate literatures and, to my knowledge, the relationships between them have not been studied in either literature. I show that, in the grouping context, EVE is exactly equivalent to the Jackknife Instrumental Variables Estimator (JIVE) of Phillips and Hale (1977), Angrist, Imbens and Krueger (1995, 1999) and Blomquist and Dahlberg (1994, 1999). The relationship between EVE and the k-class of instrumental variables estimators is also developed.

The second contribution of this paper is to use the equivalence of EVE and JIVE to examine the small sample bias of EVE and to develop an errors-in-variables estimator (UEVE) that is approximately unbiased. Unlike many instrumental variables estimators, the UEVE estimator can be implemented in situations where the microdata are unavailable provided estimates of the variance of sampling errors can be obtained. The theoretical results are supported by Monte Carlo evidence that EVE often has substantial biases but UEVE is close to unbiased and tends to have lower finite sample variance than EVE. In the final section of the paper,

I estimate a model of intertemporal labor supply using a cohort approach in repeated cross-sectional data. There are sizeable differences in the wage elasticity across estimators, showing the practical importance of the theoretical issues discussed in this paper even in circumstances where the sample size is quite large.

2 The Grouping Model

Assume that there are G groups and n_g is the number of observations in group g . The sample mean of x for group g , \bar{x}_g , is the mean of x over all members of group g included in the sample. The population mean of x for that group (π_g) relates to the mean of x for all members of the underlying population who are in that group.

Consider the following model:

$$y_{gi} = \pi'_g \beta + u_{gi} \quad i = 1, \dots, n_g, \quad g = 1, \dots, G \quad (1)$$

$$x_{gi} = \pi_g + v_{gi} \quad (2)$$

Here π_g and β are $K \times 1$ vectors where K is the number of right hand side variables, and $\sum_{g=1}^G n_g = N$.

Taking means within groups,

$$\bar{y}_g = \pi'_g \beta + \bar{u}_g \quad (3)$$

$$\bar{x}_g = \pi_g + \bar{v}_g \quad (4)$$

Assume that the sampling error has the following structure:

$$\begin{pmatrix} \bar{u}_g \\ \bar{v}_g \end{pmatrix} \sim iid \left(0, \frac{1}{n_g} \begin{bmatrix} \varrho & \sigma' \\ \sigma & \Sigma \end{bmatrix} \right) \quad (5)$$

2.1 The Application: Intertemporal Male Labor Supply

Browning, Deaton, and Irish (1985) use repeated cross-sectional data from the British Family Expenditure Survey (FES) to estimate the intertemporal wage elasticity for men. As described below, in section 5, I take a similar approach to estimation using the Integrated Public Use Files from the United States Census (IPUMS) from years 1980, 1990, and 2000 (Ruggles et al., 2004).

MaCurdy (1981) shows that the intertemporal Frisch labor supply curve under certainty takes the form

$$y_{it} = x'_{it}\beta + \alpha_i + u_{it} \quad i = 1, \dots, N \quad t = 1, \dots, T \quad (6)$$

where i indexes individual, t indexes time, y_{it} is the log of hours worked, x_{it} is a k -dimensional column vector of exogenous variables (including the log wage), β is a k -dimensional parameter vector, and α_i is an individual effect that controls for the marginal utility of wealth. The error term, u_{it} , is assumed to be uncorrelated with x_{it} and α_i , but x_{it} may be correlated with α_i . MaCurdy (1981) and Altonji (1986) estimate this type of labor supply equation for men using individual fixed effects approaches with panel data from the Panel Study of Income Dynamics (PSID).

Assume that the available data are a set of repeated cross-sections. Since the same individuals are not observed over time, it is impossible to use standard fixed effects methods to allow α_i to be correlated with x_{it} . Deaton (1985) proposed identifying β by dividing the data into groups of cohorts indexed by c , e.g. men born in 1965. The intertemporal labor supply model can be estimated after grouping observations in each period at the cohort level, because the distribution of the marginal utility of wealth is time invariant at the cohort level.

In a finite sample, taking means by cohort-year gives the following:

$$\bar{y}_{ct} = \bar{x}'_{ct}\beta + \bar{\alpha}_{ct} + \bar{u}_{ct} \quad (7)$$

The sample mean of x for group ct (\bar{x}_{ct}) is the mean of x over sample observations in cohort c at time t .

The standard cohort approach is to use the EWALD estimator – replace $\bar{\alpha}_{ct}$ with cohort dummies and estimate equation (7) by OLS or weighted least squares (if there are different numbers of observations in different groups). This estimator provides consistent estimates as N goes to infinity even if α_i is correlated with x_{it} . Deaton notes that the EWALD estimator yields biased estimates for finite N because the cohort effect ($\bar{\alpha}_{ct}$) is not constant over time due to different individuals being sampled in the cohort in different time periods. That is, EWALD is biased in small samples because $cov(\bar{\alpha}_{ct} - \alpha_c, \bar{x}_{ct}) \neq 0$, where α_c is the true cohort effect.

Taking expectations of equation (6) conditional on cohort and year gives the cohort population version:

$$y_{ct} = x'_{ct}\beta + \alpha_c + u_{ct} \quad (8)$$

$$x_{ict} = x_{ct} + v_{ict} \quad (9)$$

Here y_{ct} and x_{ct} denote the population means of y and x , respectively, in cohort c at time t . Note that equations (8) and (9) take the same form as equations (1) and (2) above. Since the population in each cohort is assumed fixed over time, the cohort effect (α_c) is constant over time and can be replaced by cohort dummies. Now, the small sample bias of EWALD can be interpreted as a measurement error problem as \bar{x}_{ct} and \bar{y}_{ct} are error-ridden measures of x_{ct} and y_{ct} .

While the application in this paper is a cohort model, one should note that other models fit in this framework. For example, firm-level regressions in which some or all of the right hand side variables are averages across a sample of workers within the firm (see Mairesse and Greenan (1999) for an explicit description of how firm-level regressions using matched firm-worker data fit in this framework). Finally, there are many contexts in which instruments naturally take a binary or categorical form such as quarter of birth (Angrist and Krueger, 1991), or lottery numbers (Angrist, 1990). As we shall see, models with dichotomous instruments will tend to fit into the framework used here.

2.2 Existing Grouping Estimators

Define the EWALD estimator:

$$\beta^{EWALD} = \left(\sum_{g=1}^G n_g \bar{x}_g \bar{x}'_g \right)^{-1} \left(\sum_{g=1}^G n_g \bar{x}_g \bar{y}_g \right) \quad (10)$$

The EWALD estimator has been shown (for example, by Angrist (1991)) to be identical to the two stage least squares estimator where group indicators are used as instruments for x_{gi} :

$$\left(\sum_{g=1}^G n_g \bar{x}_g \bar{x}'_g \right)^{-1} \left(\sum_{g=1}^G n_g \bar{x}_g \bar{y}_g \right) = \left(\sum_{g=1}^G x'_g P_g x_g \right)^{-1} \left(\sum_{g=1}^G x'_g P_g y_g \right)$$

Here

$$x'_g = [x_{g,1}, \dots, x_{g,n_g}] \quad (11)$$

$$y'_g = [y_{g,1}, \dots, y_{g,n_g}] \quad (12)$$

$$P_g = \frac{1}{n_g} l_g l'_g \quad (13)$$

l_g denotes the n_g dimensional vector of ones, and x_g is an $n_g \times K$ matrix.

Deaton (1985) shows that the EWALD estimator is inconsistent when the number of groups is taken to infinity with the number of observations per group held fixed:

$$p \lim \beta^{EWALD} = p \lim \left(\frac{1}{G} \sum_{g=1}^G n_g \pi_g \pi'_g + \Sigma \right)^{-1} p \lim \left(\frac{1}{G} \sum_{g=1}^G n_g \pi_g \pi'_g \beta + \sigma \right) \quad (14)$$

The bias here that arises from estimating π_g is somewhat analogous to the incidental parameters problem in panel data (Neyman and Scott, 1948). Given equation (14), Deaton shows that one can consistently estimate β

as G goes to infinity with n_g fixed using the following errors in variables estimator (EVE):

$$\beta^{EVE} = \left(\sum_{g=1}^G n_g \bar{x}_g \bar{x}'_g - G \widehat{\Sigma} \right)^{-1} \left(\sum_{g=1}^G n_g \bar{x}_g \bar{y}_g - G \widehat{\sigma} \right) \quad (15)$$

$\widehat{\Sigma}$ and $\widehat{\sigma}$ are sample estimates of the relevant population parameters. Here, $M_g = I_g - P_g$.

$$\widehat{\Sigma} = \frac{1}{G} \sum_{g=1}^G \widehat{\Sigma}_g \quad (16)$$

$$\widehat{\Sigma}_g = \frac{1}{n_g - 1} \sum_{i=1}^{n_g} (x_{gi} - \bar{x}_g) (x_{gi} - \bar{x}_g)' = \frac{1}{n_g - 1} x'_g M_g x_g \quad (17)$$

$$\widehat{\sigma} = \frac{1}{G} \sum_{g=1}^G \widehat{\sigma}_g \quad (18)$$

$$\widehat{\sigma}_g = \frac{1}{n_g - 1} \sum_{i=1}^{n_g} (x_{gi} - \bar{x}_g) (y_{gi} - \bar{y}_g) = \frac{1}{n_g - 1} x'_g M_g y_g \quad (19)$$

McClellan and Staiger (1999) implement a similar estimator using GMM.

3 Errors-in-Variables Estimators and Bias-Corrected Instrumental Variables

In the next sections, I show that, like EWALD, the EVE estimator can be understood as an instrumental variables estimator. In fact, the EVE estimator can be shown to be exactly identical to the Jackknife Instrumental Variables Estimator (JIVE) and to be closely related to the k-class estimators. Then, results from the instrumental variables literature are used to calculate the small-sample bias of EVE and develop an errors-in-variables estimator that is approximately unbiased in finite samples (UEVE).

3.1 The JIVE Estimator

Consider a standard instrumental variables model:

$$Y = X\beta + \epsilon \quad (20)$$

$$X = Z\Pi + \eta \quad (21)$$

X is an N by K matrix that may include endogenous variables, and Z is an N by G matrix of instruments. Assume that ϵ and η are homoskedastic with $K + 1$ by $K + 1$ variance matrix $\Sigma_{\epsilon\eta}$. Denote the probability limits of $Z'Z/N$ and $X'X/N$ as Σ_z and Σ_x respectively. Define $P_z = Z(Z'Z)^{-1}Z'$. The 2SLS estimator is

$$\beta_{2SLS} = (X'P_zX)^{-1}(X'P_zY) \quad (22)$$

While β_{2SLS} is consistent as N goes to infinity, it is now well known (see Nagar (1959), Phillips and Hale (1977), Bound, Jaeger, Baker (1995), Staiger and Stock (1997), and others) that it is biased in finite samples when there are many instruments Z relative to the dimension of X . The JIVE and k-class estimators have been proposed as alternatives to 2SLS with better bias properties in finite samples.

Phillips and Hale (1977, henceforth PH), Angrist, Imbens, and Krueger (1995, 1999, henceforth AIK), and Blomquist and Dahlberg's (1994, 1999, henceforth BD) JIVE estimator works as follows: Let $Z(i)$ and $X(i)$ denote matrices equal to Z and X with the i th row removed. Consider the following estimate of Π for observation i :

$$\tilde{\Pi}(i) = (Z(i)'Z(i))^{-1}(Z(i)'X(i))$$

Define \hat{X}_{JIVE} to be the $N \times K$ dimensional matrix with i th row $Z_i\tilde{\Pi}(i)$. The JIVE estimator is

$$\beta_{JIVE} = (\hat{X}'_{JIVE}X)^{-1}(\hat{X}'_{JIVE}Y)$$

Note the intuition behind the JIVE estimator: In forming the “predicted value” of X for observation i , one uses a Π coefficient estimated on all observations other than i . This eliminates overfitting problems in the first stage.

The following lemma is adapted from AIK (it is proved in appendix 1).

Lemma 1 : For the model in equations (20) and (21), assume that we can write an estimator $\widehat{\beta}$ in the form

$$\widehat{\beta} = (X'CX)^{-1}(X'C'Y) \quad (23)$$

where C is an $N \times N$ matrix such that the elements of C are of stochastic order $Op(1/\sqrt{N})$ and

$$CX = Z\Pi + C\eta \quad (24)$$

Then, with $\Sigma_z = p \lim(Z'Z/N)$, the approximate bias of $\widehat{\beta}$ to order $\frac{1}{N}$ equals

$$\frac{\sigma_{\epsilon\eta}(\Pi'\Sigma_z\Pi)^{-1}}{N} [\text{trace}(C) - K - 1]$$

3.2 Relationship of EVE to JIVE

Define $\bar{x}_g(i)$ as the mean of x over all observations in group g *except observation i* . In the grouping context, the JIVE estimator can be written as

$$\beta^{JIVE} = \left(\sum_{g=1}^G \sum_{i=1}^{n_g} \bar{x}_g(i)x'_{gi} \right)^{-1} \left(\sum_{g=1}^G \sum_{i=1}^{n_g} \bar{x}_g(i)y_{gi} \right) \quad (25)$$

That is, the instrument for x for any observation i equals the mean value of x in the group where the mean is calculated over all observations *except observation i* . Mechanically, $\bar{x}_g(i) = \left(\frac{n_g \bar{x}_g - x_{gi}}{n_g - 1}\right)$. Note that

$$\sum_{i=1}^{n_g} \frac{n_g \bar{x}_g - x_{gi}}{n_g - 1} x'_{gi} = \frac{n_g}{n_g - 1} \sum_{i=1}^{n_g} \bar{x}_g x'_{gi} - \frac{1}{n_g - 1} \sum_{i=1}^{n_g} x_{gi} x'_{gi} \quad (26)$$

$$= \frac{n_g}{n_g - 1} \bar{x}_g (n_g \bar{x}_g)' - \frac{1}{n_g - 1} \left(\sum_{i=1}^{n_g} (x_{gi} - \bar{x}_g) (x_{gi} - \bar{x}_g)' + n_g \bar{x}_g \bar{x}'_g \right) \quad (27)$$

$$= \frac{n_g^2 - n_g}{n_g - 1} \bar{x}_g \bar{x}'_g - \frac{1}{n_g - 1} \sum_{i=1}^{n_g} (x_{gi} - \bar{x}_g) (x_{gi} - \bar{x}_g)' \quad (28)$$

$$= n_g \bar{x}_g \bar{x}'_g - \frac{1}{n_g - 1} \sum_{i=1}^{n_g} (x_{gi} - \bar{x}_g) (x_{gi} - \bar{x}_g)' \quad (29)$$

$$= n_g \bar{x}_g \bar{x}'_g - \frac{1}{n_g - 1} x'_g M_g x_g \quad (30)$$

$$= n_g \bar{x}_g \bar{x}'_g - \widehat{\Sigma}_g \quad (31)$$

It follows that

$$\sum_{g=1}^G \sum_{i=1}^{n_g} \bar{x}_g(i) x'_{gi} = \sum_{g=1}^G n_g \bar{x}_g \bar{x}'_g - \sum_{g=1}^G \widehat{\Sigma}_g \quad (32)$$

Likewise, we have

$$\sum_{g=1}^G \sum_{i=1}^{n_g} \bar{x}_g(i) y_{gi} = \sum_{g=1}^G n_g \bar{x}_g \bar{y}'_g - \sum_{g=1}^G \widehat{\sigma}_g \quad (33)$$

Using equations (16) - (19), one can write the JIVE estimator of β as

$$\left(\sum_{g=1}^G \sum_{i=1}^{n_g} \bar{x}_g(i) x'_{gi} \right)^{-1} \left(\sum_{g=1}^G \sum_{i=1}^{n_g} \bar{x}_g(i) y_{gi} \right) = \left(\sum_{g=1}^G n_g \bar{x}_g \bar{x}'_g - G \widehat{\Sigma} \right)^{-1} \left(\sum_{g=1}^G n_g \bar{x}_g \bar{y}'_g - G \widehat{\sigma} \right) \quad (34)$$

showing the exact equivalence of the EVE and JIVE estimators.

EVE is also closely related to the k-class estimators which take the form

$$(X' P_z X - \gamma X' M_z X)^{-1} (X' P_z Y - \gamma X' M_z Y)$$

For example, Nagar's estimator (Nagar, 1959) has $\gamma = \frac{G-K+1}{N-G+K-1}$. Donald and Newey (2001) suggest the Bias-Adjusted 2SLS (B2SLS) estimator in which $\gamma = \frac{G-K-1}{N-G+K+1}$. With grouped data, the k-class estimators become

$$\left(\sum_{g=1}^G x'_g P_g x_g - \gamma \sum_{g=1}^G x'_g M_g x_g \right)^{-1} \left(\sum_{g=1}^G x'_g P_g y_g - \gamma \sum_{g=1}^G x'_g M_g y_g \right)$$

Using equations (16) - (19), this can be written as

$$\left(\sum_{g=1}^G n_g \bar{x}_g \bar{x}'_g - \gamma \sum_{g=1}^G (n_g - 1) \hat{\Sigma}_g \right)^{-1} \left(\sum_{g=1}^G n_g \bar{x}_g \bar{y}_g - \gamma \sum_{g=1}^G (n_g - 1) \hat{\sigma}_g \right) \quad (35)$$

Note that if n_g is constant across groups, EVE takes the form of a k-class estimator with $\gamma = G / (\sum_{g=1}^G n_g - G)$.

3.3 Developing An Unbiased EVE Estimator (UEVE)

PH and AIK show that the approximate bias of JIVE to order $\frac{1}{N}$ is proportional to

$$\text{trace}(C^{JIVE}) - K - 1 \quad (36)$$

where K is the number of right hand side variables. In the grouping context, C^{JIVE} is block diagonal with each block equal to

$$P_g - \frac{1}{n_g - 1} M_g \quad (37)$$

It follows that

$$\text{trace}(C^{JIVE}) = \sum_{g=1}^G \text{trace} \left(P_g - \frac{1}{n_g - 1} M_g \right) \quad (38)$$

$$= 0 \quad (39)$$

Thus, the approximate bias of JIVE, and hence EVE, is proportional to $-K - 1$.

Consider a generalized EVE estimator (GEVE) that has the following form:

$$\beta^{GEVE} = \left(\sum_{g=1}^G n_g \bar{x}_g \bar{x}'_g - \zeta G \widehat{\Sigma} \right)^{-1} \left(\sum_{g=1}^G n_g \bar{x}_g \bar{y}_g - \zeta G \widehat{\sigma} \right) \quad (40)$$

Using relationships developed above, this estimator equals

$$\left(\sum_{g=1}^G \left(x'_g P_g x_g - \frac{\zeta}{n_g - 1} x'_g M_g x_g \right) \right)^{-1} \left(\sum_{g=1}^G \left(x'_g P_g y_g - \frac{\zeta}{n_g - 1} x'_g M_g y_g \right) \right) \quad (41)$$

$$= \left(\sum_{g=1}^G (x'_g C_g^{GEVE} x_g) \right)^{-1} \left(\sum_{g=1}^G (x'_g C_g^{GEVE} y_g) \right) \quad (42)$$

Here C_g^{GEVE} equals $P_g - \frac{\zeta}{n_g - 1} M_g$. Thus C^{GEVE} is block diagonal with typical block equal to C_g^{GEVE} . I now show that the GEVE estimator satisfies the conditions of the lemma in section 3.1. To satisfy the lemma, $C^{GEVE} Z\Pi$ must equal $Z\Pi$. In the grouping context, $Z\Pi$ is a block diagonal matrix with typical block equal to $l_g \pi'_g$ where l_g is a $n_g \times 1$ vector of ones. Given the block diagonal structure of C^{GEVE} and of $Z\Pi$, $C^{GEVE} Z\Pi$ equals $Z\Pi$ if, in each block, $C_g^{GEVE} l_g \pi'_g$ equals $l_g \pi'_g$.

$$\begin{aligned} C_g^{GEVE} l_g \pi'_g &= \left(P_g - \frac{\zeta}{n_g - 1} M_g \right) l_g \pi'_g \\ &= \left(\frac{(n_g - 1) P_g l_g \pi'_g - \zeta l_g \pi'_g + \zeta P_g l_g \pi'_g}{n_g - 1} \right) \\ &= \left(\frac{(n_g - 1) l_g \pi'_g - \zeta l_g \pi'_g + \zeta l_g \pi'_g}{n_g - 1} \right) \\ &= l_g \pi'_g \end{aligned}$$

The penultimate step uses the fact that $P_g l_g = l_g$. Given that GEVE satisfies the conditions of the lemma,

the approximate bias of GEVE is proportional to

$$\sum_{g=1}^G \text{trace} \left(P_g - \frac{\zeta}{n_g - 1} M_g \right) - K - 1 \quad (43)$$

$$= G - \sum_{g=1}^G \frac{\zeta}{n_g - 1} (n_g - 1) - K - 1 \quad (44)$$

$$= G - G\zeta - K - 1 \quad (45)$$

Setting this equal to zero, one obtains

$$\zeta = \frac{G - K - 1}{G} \quad (46)$$

This implies that the estimator

$$\beta^{UEVE} = \left(\sum_{g=1}^G n_g \bar{x}_g \bar{x}'_g - (G - K - 1) \hat{\Sigma} \right)^{-1} \left(\sum_{g=1}^G n_g \bar{x}_g \bar{y}'_g - (G - K - 1) \hat{\sigma} \right) \quad (47)$$

is approximately unbiased to order $\frac{1}{N}$. Comparing UEVE (unbiased EVE) to EVE, we can see that they differ in that EVE subtracts off too much of the sampling variance of \bar{x}_g in the denominator and so overcorrects for the sampling error. Thus EVE will typically be biased away from EWALD and the bias of EVE will tend to increase with K (the number of right hand side variables).

Verbeek and Nijman (1993) suggest a version of EVE specifically for cohort models where cohorts are followed over T periods.

$$\beta^{EVE2} = \left(\sum_{g=1}^G n_g \bar{x}_g \bar{x}'_g - \left(\frac{T-1}{T} \right) G \hat{\Sigma} \right)^{-1} \left(\sum_{g=1}^G n_g \bar{x}_g \bar{y}'_g - \left(\frac{T-1}{T} \right) G \hat{\sigma} \right) \quad (48)$$

Comparison with UEVE shows that EVE2 is not in general unbiased in finite samples.

When there are the same number of observations in each group, UEVE takes the k-class form with $\gamma = (G - K - 1) / (\sum_{g=1}^G n_g - G)$. Comparing this to B2SLS (where $\gamma = (G - K - 1) / (\sum_{g=1}^G n_g - G + K + 1)$), the only difference is an additional $K + 1$ in the B2SLS denominator and this term becomes unimportant when N

gets reasonably large.

Papers in the cohort literature have typically done asymptotics as the number of cohorts goes to infinity (Collado (1997), Verbeek and Nijman (1993)) or the number of groups goes to infinity (Deaton, 1985). Having an estimator (UEVE) that is approximately unbiased when there are a small number of groups may be important as in many practical applications there are limits on the number of birth-year or birth-decade cohorts that can be used.

In appendix 2, I verify that UEVE is consistent as the number of groups goes to infinity and derive its variance under the group-asymptotic sequence. While I only consider the homoskedastic case, it is easy to verify that UEVE is group-asymptotically consistent if Σ_g differs across groups. In contrast, k-class estimators are not group-asymptotically consistent in the presence of heteroskedasticity (see Akerberg and Devereux, 2003). Bekker and van der Ploeg (1999) also show that LIML is not consistent in the heteroskedastic case.

4 Monte Carlo Simulations

In this section, I present results from Monte Carlo simulations that provide some insight about the performance of the estimators. The data are divided into a set of mutually exclusive and exhaustive groups indexed by g . These groups are allocated into mutually exclusive and exhaustive cohorts indexed by c that are supersets of these groups: The model includes a constant, a continuous variable (x), and fixed cohort effects. The model is as follows with the x_{igc} referring to the value of x for person i in group g in cohort c :

$$x_{igc} = f_c + f_g + v_{igc} \quad (49)$$

$$y_{igc} = \beta_0 + \beta_1(f_c + f_g) + h_c + u_{igc} \quad (50)$$

All the error terms (f_c, f_g, h_c, u_{igc}) are distributed $N(0, 1)$. The error term, v_{igc} , that determines the degree of sampling error in \bar{x}_{gc} is distributed $N(0, 2)$. All errors are drawn independently (so $\sigma = 0$). The value of β_0 is

set to 0, and β_1 is set equal to 1. The model is estimated using 50 groups with 5 observations per group.

I report quantiles (10%, 25%, 50%, 75%, 90%) of the distribution of the estimator around the true parameter vector. The 50% quantile is thus the median bias of the estimator. I also report the median absolute error of the estimator. Mean biases and mean squared errors of the estimators are a bit more problematic. This is because JIVE and Nagar type estimators are known not to have second moments. This makes their means extremely sensitive to outliers and makes mean squared errors meaningless. To address this issue, I trim the distributions of all the estimators (at the 5th and 95th percentiles) and report mean bias and mean absolute error for these trimmed distributions. I also report 90% coverage rates for the estimators using the group-asymptotic standard errors derived in appendix 2. In addition to the EWALD, EVE, and UEVE estimators, I report results for B2SLS which has the same finite sample properties as UEVE under homoskedasticity.

4.1 Monte Carlo Results

The results are in Table 1. The results, in panels A-C, show how increasing the number of cohorts affects the performance of the estimators. Since cohort fixed effects are included in the specification, increasing the number of cohorts increases the number of control variables. The main result from panels A-C is that, as suggested by the bias formulae, the bias of EVE increases as the number of cohorts increases: The trimmed mean bias goes from 0.04 with 2 cohorts, to 0.15 with 10 cohorts, to 0.92 with 25 cohorts. Indeed, with 25 cohorts, the bias of EVE is much larger than the bias from EWALD. Also, the spread of EVE increases as the number of cohorts increases. On the other hand, the UEVE estimator remains approximately unbiased as the number of cohorts is increased.

In panels D-F, the sampling error problem is increased by increasing the variance of v_{igc} to 5. As expected, the bias of EWALD is greater than before, but the clear advantage of UEVE over EVE is still evident.

In all panels, the 90% coverage rates of UEVE are quite close to 0.90, suggesting that the group-asymptotic standard errors work quite well even though there are only 50 groups. Overall, it is clear from the Monte Carlos

that UEVE is a significant improvement over EVE in terms of both bias and variance.

5 An Application to Intertemporal Male Labor Supply

I apply the estimators to the labor supply model from section 2.1, using U.S. Census microdata from 1980, 1990, and 2000 (Ruggles et al., 2004). The sample consists of men who are aged 25 to 40 in 1980. Thus, the men are aged 35 to 50 in 1990, and 45 to 60 in 2000. The hours measure used is annual hours worked in the preceding calendar year, and the wage measure is average hourly earnings in that year. Earnings are topcoded in all three Census files (at \$75,000 in 1980, \$140,000 in 1990, and \$175,000 dollars in 2000). I impute earnings for topcoded values as 1.33 times the topcoded value. I exclude individuals who did not work any hours in the preceding calendar year or who report working more than 80 hours per week.

Because the Census samples are large, I define a cohort by birth year and by region of birth. Thus, there are 144 cohorts (16 birth years times 9 Census regions), and 432 (144 cohorts by 3 years) groups. As described in section 2.1, the labor supply equation is a log-linear hours-wage equation: The log of weekly hours in each group is a function of the log wage, indicator variables for the 144 cohorts, and indicator variables for the 3 years. In addition, I include controls for marital status (a dummy that equals one if the individual is currently married and living with their spouse), number of children in the household aged less than five, and number of children in the household aged five or more. The estimating sample is composed of 2,915,397 men. I carry out separate analyses by education level, and by race. Descriptive statistics for the sample are in Table 2.

5.1 Results

The estimated coefficients and standard errors by education group are in Table 3. First consider the EWALD results in the first column: These suggest a wage elasticity of about 0.4 for all four education groups. The presence of young children in the household leads to lower hours worked, with the effects being larger for the less educated groups. The presence of older children reduces hours for the lowest two education groups, but

there is no evidence of this effect for the higher educated. In all four samples, married men work significantly longer hours than other men. These results are all consistent with expectations.

The EVE estimates, in the second column, are much less precisely estimated than their EWALD equivalents. For all but the lowest education group, the coefficient estimates are generally bigger in absolute terms than EWALD, suggesting the EWALD bias is an attenuation bias for these samples. EVE seems particularly problematic in the high school dropout sample in that the number of children coefficients and the married coefficient have perverse signs. However, these coefficients are very imprecisely estimated.

The UEVE estimates in the third column are quantitatively quite different from both EWALD and EVE. The wage elasticity is uniformly higher than EWALD across education groups, and is estimated to be about 0.6 for the high school dropouts and for college graduates, with values about 0.45 for the other groups. Likewise, the negative effects of children on labor supply (and the positive effects of marriage) are estimated to be larger using UEVE than using EWALD, with the effects of both old and young children being negative and statistically significant for all four education groups. As expected, the UEVE estimates are less precisely estimated than EWALD but more precisely estimated than EVE. The B2SLS estimates and standard errors are generally very close to UEVE, suggesting that heteroskedasticity is not a serious problem in this application.

In Table 4, I estimate the specification by race. Based on the EWALD estimates, one would conclude that the wage elasticity is significantly higher for whites than non-whites. In contrast, the UEVE estimates are very similar for both racial groups. Thus, the particularly low EWALD elasticity for non-whites appears to be a symptom of finite sample bias in this relatively small sample. Note that the EVE estimates are very imprecise and generally have the wrong sign in the non-whites sample. In contrast, EWALD, UEVE, and EVE are all quite similar in the sample of whites, reflecting the fact that the sample is very large.

The preferred estimates in Tables 3 and 4 are the UEVE estimates as the theory and Monte Carlo evidence suggests that these are approximately unbiased. These suggest an intertemporal wage elasticity of approximately 0.4 - 0.6 for all groups of men. This elasticity is larger than that found by Browning et al. (1985) for

British data but somewhat smaller than the estimates of Angrist (1991) using U.S. data from the Panel Study of Income Dynamics. The variation in the estimated elasticities and standard errors across estimators in Tables 3 and 4 implies that the choice of estimator may be of great importance in empirical practice.

One interesting feature of the application is that EWALD and UEVE estimates are quite different despite the fact that the sample size is large relative to the number of groups. There are two features of the specification that help explain why finite sample issues are relevant to these seemingly large samples. The first is that there are four endogenous variables (wages, children aged less than 5, children aged 5 or more, and marital status) and the group-means of these variables are correlated. The second is that both cohort and year fixed effects are included and conditioning on these reduces the cross-group variance of wages substantially because the variance of wages over time within cohorts is much lower than the variance of wages across cohorts in the cross-section.

While reasonably small numbers of observations may be sufficient for precisely estimating group means, the presence of cohort and year fixed effects in cohort models increases enormously the likelihood of serious small sample biases in EWALD and the number of observations required to eliminate biases. Thus, even if the variance of \bar{x}_g is low because there are many observations per group, it may still be sizeable relative to the cross-group variance in π_g . Given the equivalence of EWALD to the 2SLS estimator using the microdata and group indicators as instruments, the finding here is similar to that of Bound et al. (1995) that 2SLS can be very biased in overidentified linear models even if the number of observations is very large. Devereux (2006) provides another example where EWALD suffers from small sample bias even with very large sample sizes.

6 Conclusions

This paper has two main results: The first finding is that, with grouped data, the EVE estimator is identical to JIVE and therefore is biased in finite samples. Second, I show that one can use results from the instrumental variables literature to construct an unbiased EVE estimator (UEVE) that is approximately unbiased in finite

samples. Monte Carlo experiments support the theoretical results and show that the UEVE estimator has both lower bias and variance than EVE. In the intertemporal labor supply application, the EWALD, EVE, and UEVE estimates of the intertemporal wage elasticity are often quite different. This suggests that the choice of grouping estimator is very relevant in practice.

While the UEVE estimator is closely related to instrumental variables estimators, there are situations where the instrumental variables estimators are infeasible but the UEVE estimator can be implemented using estimates of the group means and sampling variances. For example, Angrist (1990), used restricted Social Security Administration (SSA) data to examine the effects of Vietnam draft eligibility on earnings. For confidentiality reasons, the SSA would not provide individual-level data but did provide information on first and second moments of the variables by group. In this type of situation, the UEVE estimator could be implemented but conventional instrumental variables estimators are not feasible.

References

- [1] Acemoglu, D. and J. S. Pischke (2001), 'Changes in the wage structure, family income, and children's education', *European Economic Review*, May, 45, 890-904.
- [2] Ackerberg, D. and P. Devereux (2003), 'Improved Jive Estimators for Overidentified Linear Models with and without Heteroskedasticity', mimeo.
- [3] Altonji, J. (1986), 'Intertemporal Substitution in Labor Supply: Evidence from Micro Data' *Journal of Political Economy*, 94, S176-S215.
- [4] Angrist, J.D. (1990) 'Lifetime earnings and the Vietnam era draft lottery: Evidence from Social Security Administrative Records' *American Economic Review*, 80(3), 313-336.
- [5] Angrist, J.D. (1991) "Grouped Data Estimation and Testing in Simple Labor Supply Models", *Journal of Econometrics*, 47, 243-265.
- [6] Angrist, J.D. and A.B. Krueger (1991), 'Does compulsory school attendance affect schooling and earnings?', *Quarterly Journal of Economics*, 106, 979-1014.
- [7] Angrist, J.D., G.W. Imbens, and A.B. Krueger (1995), 'Jackknife Instrumental Variables Estimation', NBER technical working paper.
- [8] Angrist, J.D., G.W. Imbens, and A.B. Krueger (1999), 'Jackknife Instrumental Variables Estimation', *Journal of Applied Econometrics*, 14, 57-67.
- [9] Bekker P.A., and J. van der Ploeg (1999), 'Instrumental variable estimation based on grouped data', mimeo.
- [10] Blomquist S. and M. Dahlberg (1994), 'Small sample properties of jackknife instrumental variables estimators: experiments with weak instruments', mimeo, Uppsala University.

- [11] Blomquist S. and M. Dahlberg (1999), 'Small sample properties of LIML and jackknife IV estimators: experiments with weak instruments', *Journal of Applied Econometrics*, 14, 69-88.
- [12] Blundell, Richard, Alan Duncan, and Costas Meghir (1998) "Estimating Labor Supply Responses using Tax Reforms", *Econometrica*, 66, 4, 827-861.
- [13] Bound, J., D. Jaeger, and R. Baker (1995), 'Problems with instrumental variables estimation when the correlation between instruments and the endogenous explanatory variable is weak', *Journal of the American Statistical Association*, June, 443-450.
- [14] Browning, M., A. Deaton, and M. Irish (1985), "A profitable approach to labor supply and commodity demands over the life-cycle", *Econometrica*, 53(3), 503-544.
- [15] Card, D. and T. Lemieux (1996), "Wage dispersion, returns to skill, and black-white wage differentials" *Journal of Econometrics* 74, 319-361.
- [16] Collado, M.D. (1997), 'Estimating dynamic models from time series of independent cross-sections', *Journal of Econometrics*, 82, 37-62.
- [17] Deaton, A. (1985), 'Panel data from a time series of cross-sections', *Journal of Econometrics*, 30, 109-126.
- [18] Devereux, P. (2006), 'Small Sample Bias in Synthetic Cohort Models of Labor Supply', UCD Centre for Economic Research Working Paper WP06/06.
- [19] Donald S. and W. Newey (2001), 'Choosing the number of instruments', *Econometrica*, 69(5), 1161-1191.
- [20] Mairesse, Jacques, and Nathalie Greenan (1999) "Using Employee-Level Data in a Firm-Level Econometric Study" in *The Creation and Analysis of Employer-Employee Matched Data*, J.I. Lane, J.R. Spletzer, J.J.M. Theeuwes and K.R. Troske, Elsevier Science B.V.

- [21] MaCurdy T. (1981), 'An Empirical Model of Labor Supply in a Life-Cycle Setting', *Journal of Political Economy*, 89, 1059-1085.
- [22] McClellan M., and D. Staiger (1999), 'Estimating treatment effects using hospital-level variation in treatment intensity', mimeo.
- [23] McKenzie, D.J. (2001), 'Consumption Growth in a Booming Economy: Taiwan 1976-96' mimeo.
- [24] Nagar, A.L. (1959), 'The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations', *Econometrica*, 27, 575-595.
- [25] Neyman, J. and E. Scott (1948), 'Consistent Estimates Based on Partially Consistent Observations', *Econometrica*, 16, 1-32.
- [26] Phillips and Hale (1977) "The Bias of Instrumental Variable Estimators of Simultaneous Equation Systems" *International Economic Review*, Vol. 18, No. 1., pp. 219-228.
- [27] Ruggles, Steven, Matthew Sobek, Trent Alexander, Catherine A. Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, and Chad Ronnander (2004). Integrated Public Use Microdata Series: Version 3.0 [Machine-readable database]. Minneapolis, MN: Minnesota Population Center [producer and distributor].
- [28] Staiger D. and J.H. Stock (1997), 'Instrumental variables regression with weak instruments', *Econometrica*, 65, 557-586.
- [29] Verbeek, M. and T. Nijman (1993), "Minimum MSE estimation of a regression model with fixed effects from a series of cross-sections" *Journal of Econometrics* 59, 125-136.

7 Appendix 1: Proof of Lemma 1

Following AIK, first I derive the bias of $\widehat{\beta} = (X'C'X)^{-1}(X'C'Y)$ relative to the bias of $\widehat{\beta}(Z\Pi)$ where $\widehat{\beta}(Z\Pi) = (\Pi'Z'X)^{-1}(\Pi'Z'Y)$

$$\begin{aligned}\widehat{\beta} - \widehat{\beta}(Z\Pi) &= (X'C'X)^{-1}(X'C'Y) - \widehat{\beta}(Z\Pi) \\ &= (\Pi'Z'X + \eta'C'X)^{-1}(\Pi'Z'Y + \eta'C'Y) - \widehat{\beta}(Z\Pi)\end{aligned}$$

Defining $R = (\Pi'Z'X)^{-1}$, this can be written as

$$\begin{aligned}\widehat{\beta} - \widehat{\beta}(Z\Pi) &= (R^{-1}(I + R\eta'C'X))^{-1}(\Pi'Z'Y + \eta'C'Y) - \widehat{\beta}(Z\Pi) \\ &= (I + R\eta'C'X)^{-1}(R\Pi'Z'Y + R\eta'C'Y) - \widehat{\beta}(Z\Pi)\end{aligned}$$

Expanding $(I + R\eta'C'X)^{-1}$ around $R\eta'C'X = 0$ and ignoring terms of order less than $1/N$ gives

$$\begin{aligned}\widehat{\beta} - \widehat{\beta}(Z\Pi) &= (I - R\eta'C'X)(R\Pi'Z'Y + R\eta'C'Y) - \widehat{\beta}(Z\Pi) + o_p(1/N) \\ &= R\Pi'Z'Y + R\eta'C'Y - R\eta'C'XR\Pi'Z'Y - R\eta'C'XR\eta'C'Y - (\Pi'Z'X)^{-1}(\Pi'Z'Y) + O_p(1/N) \\ &= R\Pi'Z'Y + R\eta'C'Y - R\eta'C'XR\Pi'Z'Y - R\eta'C'XR\eta'C'Y - R\Pi'Z'Y + O_p(1/N) \\ &= R\eta'C'Y - R\eta'C'X\widehat{\beta}(Z\Pi) - R\eta'C'XR\eta'C'Y + O_p(1/N) \\ &= R\eta'C'\epsilon - R\eta'C'X(\widehat{\beta}(Z\Pi) - \beta) - R\eta'C'XR\eta'C'Y + O_p(1/N)\end{aligned}$$

The term $R\eta'C'XR\eta'C'Y$ is of order lower than $1/N$. Expanding the i th row of $X(\widehat{\beta}(Z\Pi) - \beta)$, one gets

$$X_i(\widehat{\beta}(Z\Pi) - \beta) = X_i(\Pi'Z'X)^{-1}(\Pi'Z'\epsilon) = Z_i\Pi(\Pi'Z'Z\Pi)^{-1}(\Pi'Z'\epsilon) + O_p(1/\sqrt{N})$$

Then, expanding NR around $R_0 = p \lim(\Pi'Z'Z\Pi/N)^{-1} = (\Pi'\Sigma_z\Pi)^{-1}$, one can write

$$R\eta'C'\epsilon - R\eta'C'X(\widehat{\beta}(Z\Pi) - \beta) = \frac{1}{N}(R_0\eta'C'\epsilon - R_0\eta'C'P_{Z\Pi}\epsilon) + O_p(1/N)$$

where $P_{Z\Pi} = Z\Pi(\Pi'Z'Z\Pi)^{-1}\Pi'Z'$. $E(\widehat{\beta} - \beta) = E(\widehat{\beta} - \widehat{\beta}(Z\Pi)) + E(\widehat{\beta}(Z\Pi) - \beta)$. The approximate bias of $\widehat{\beta}(Z\Pi)$ equals $-\sigma_{\epsilon\eta}(\Pi'\Sigma_z\Pi)^{-1}/N$. Hence

$$\begin{aligned} E(\widehat{\beta} - \beta) &= \frac{\sigma_{\epsilon\eta}(\Pi'\Sigma_z\Pi)^{-1}}{N} E(C' - C'P_{Z\Pi} - 1) \\ &= \frac{\sigma_{\epsilon\eta}(\Pi'\Sigma_z\Pi)^{-1}}{N} \text{trace}(C - C'P_{Z\Pi} - 1) \end{aligned}$$

Then, because $CZ\Pi = Z\Pi$, $C'P_{Z\Pi} = P_{Z\Pi}$. Also, since $\text{trace}(P_{Z\Pi}) = K$,

$$E(\widehat{\beta} - \beta) = \frac{\sigma_{\epsilon\eta}(\Pi'\Sigma_z\Pi)^{-1}}{N} (\text{trace}(C) - K - 1)$$

Note that if the homoskedasticity assumption is violated, $\text{trace}(\epsilon\eta'C'P_{Z\Pi})$ now depends on the exact form of heteroskedasticity so the bias formula no longer has this simple form.

8 Appendix 2: Group-Asymptotic Properties of UEVE

8.1 Consistency of UEVE as $G \rightarrow \infty$

Deaton (1985) shows that EVE is consistent as G goes to infinity. In this section, I show the consistency of UEVE. Assume that $p \lim \frac{1}{G} \left(\sum_{g=1}^G n_g \pi_g \pi'_g \right) = \Omega$, a positive definite matrix. The probability limit of

$\frac{1}{G} \left(\sum_{g=1}^G n_g \bar{x}_g \bar{x}'_g - (G - K - 1) \widehat{\Sigma} \right)$ as G goes to infinity is as follows:

$$\begin{aligned}
& p \lim \frac{1}{G} \left(\sum_{g=1}^G n_g \bar{x}_g \bar{x}'_g - (G - K - 1) \widehat{\Sigma} \right) \\
&= p \lim \frac{1}{G} \left(\sum_{g=1}^G n_g (\pi_g + \bar{v}_g) (\pi_g + \bar{v}_g)' \right) - p \lim \frac{G - K - 1}{G} p \lim \frac{1}{G} \sum_{g=1}^G \widehat{\Sigma}_g \\
&= p \lim \frac{1}{G} \left(\sum_{g=1}^G n_g \pi_g \pi'_g \right) + p \lim \frac{1}{G} \left(\sum_{g=1}^G n_g \bar{v}_g \bar{v}'_g \right) - p \lim \frac{1}{G} \sum_{g=1}^G \widehat{\Sigma}_g + p \lim \frac{(K + 1)}{G} \frac{1}{G} \sum_{g=1}^G \widehat{\Sigma}_g \\
&= p \lim \frac{1}{G} \left(\sum_{g=1}^G n_g \pi_g \pi'_g \right) + p \lim \frac{1}{G} \left(\sum_{g=1}^G n_g \frac{\Sigma}{n_g} \right) - p \lim \frac{1}{G} \sum_{g=1}^G \Sigma \\
&= p \lim \frac{1}{G} \left(\sum_{g=1}^G n_g \pi_g \pi'_g \right)
\end{aligned} \tag{51}$$

Likewise,

$$p \lim \frac{1}{G} \left(\sum_{g=1}^G n_g \bar{x}_g \bar{y}'_g - (G - K - 1) \widehat{\sigma} \right) = p \lim \frac{1}{G} \left(\sum_{g=1}^G n_g \pi_g \pi'_g \beta \right) \tag{52}$$

Together, (51) and (52) establish the consistency of the UEVE estimator as G goes to infinity.

8.2 Group-Asymptotic Variance of UEVE

To simplify notation, define

$$p \lim(M_{xx}) - \alpha \Sigma = \Omega$$

$$p \lim(M_{xy}) - \alpha \sigma = \Omega \beta$$

$$p \lim(M_{yy}) - \varrho = \beta' \Omega \beta$$

where $M_{xx} = (1/G) \sum_{g=1}^G n_g \bar{x}_g \bar{x}'_g$, $M_{xy} = (1/G) \sum_{g=1}^G n_g \bar{x}_g \bar{y}_g$, $M_{yy} = (1/G) \sum_{g=1}^G n_g \bar{y}_g \bar{y}'_g$, $\Omega = p \lim \frac{1}{G} \left(\sum_{g=1}^G n_g \pi_g \pi'_g \right)$ and the probability limits are taken as G goes to infinity. Define the UEVE estimator as

$$\tilde{\beta} = (M_{xx} - \alpha \widehat{\Sigma})^{-1} (M_{xy} - \alpha \widehat{\sigma}) \quad (53)$$

where α equals $(G - K - 1) / G$. Assume that the sampling error is normally distributed:

$$\begin{pmatrix} \bar{u}_g \\ \bar{v}_g \end{pmatrix} \sim N \left(0, \frac{1}{n_g} \begin{bmatrix} \varrho & \sigma' \\ \sigma & \Sigma \end{bmatrix} \right) \quad (54)$$

The exposition here closely follows Deaton (1985). Expanding (53) around β gives

$$\tilde{\beta} - \beta = \Omega^{-1} [(M_{xy} - M_{xx}\beta) - \alpha(\sigma - \Sigma\beta)] - \alpha\Omega^{-1} [(\widehat{\sigma} - \widehat{\Sigma}\beta) - (\sigma - \Sigma\beta)] + O_p(G^{-1}) \quad (55)$$

The assumption of sampling under normality ensures that the second term is asymptotically independent of the first. Since the terms in equation (55) are sample averages centered around their means, by using a Central Limit Theorem for independent but not identically distributed random variables one can show that $\sqrt{G}(\tilde{\beta} - \beta)$ is asymptotically normally distributed.

The asymptotic variance of $\tilde{\beta}$ depends on the asymptotic variance of $\Omega^{-1} [(M_{xy} - M_{xx}\beta) - \alpha(\widehat{\sigma} - \widehat{\Sigma}\beta)]$.

Deaton shows that

$$GV\{M_{xy} - M_{xx}\beta\} = p \lim (M_{xx})(\varrho + \beta' \Sigma \beta - 2\sigma' \beta) + (\sigma - \Sigma\beta)(\sigma - \Sigma\beta)'$$

Given that $\widehat{\Sigma} = \frac{1}{G} \sum_{g=1}^G \widehat{\Sigma}_g$ and $\widehat{\sigma} = \frac{1}{G} \sum_{g=1}^G \widehat{\sigma}_g$,

$$V\{\widehat{\sigma} - \widehat{\Sigma}\beta\} = \frac{1}{G^2} V\left\{ \sum_{g=1}^G (\widehat{\sigma}_g - \widehat{\Sigma}_g \beta) \right\} = \frac{1}{G^2} \sum_{g=1}^G V\{\widehat{\sigma}_g - \widehat{\Sigma}_g \beta\} \quad (56)$$

Then, following Deaton, sampling under normality implies that the asymptotic variance of $\widehat{\sigma}_g - \widehat{\Sigma}_g\beta$ is

$$n_g V\{\widehat{\sigma}_g - \widehat{\Sigma}_g\beta\} = \Sigma[(\varrho + \beta'\Sigma\beta - 2\sigma'\beta) + (\sigma - \Sigma\beta)(\sigma - \Sigma\beta)'] \quad (57)$$

Because $(M_{xy} - M_{xx}\beta)$ and $(\widehat{\sigma} - \widehat{\Sigma}\beta)$ are asymptotically independently distributed, the asymptotic variance-covariance matrix of $\widetilde{\beta}$ is given by

$$GV\{\widetilde{\beta}\} = \Omega^{-1}[A + \alpha^2 B]\Omega^{-1} \quad (58)$$

where

$$\begin{aligned} A &= p \lim(M_{xx})(\varrho + \beta'\Sigma\beta - 2\sigma'\beta) + (\sigma - \Sigma\beta)(\sigma - \Sigma\beta)' \\ B &= \frac{1}{G} \sum_{g=1}^G \frac{1}{n_g} \Sigma[(\varrho + \beta'\Sigma\beta - 2\sigma'\beta) + (\sigma - \Sigma\beta)(\sigma - \Sigma\beta)'] \end{aligned}$$

To evaluate the variance-covariance matrix in practice requires estimates of Ω and ϱ . These can be estimated as follows:

$$\begin{aligned} \widetilde{\Omega} &= M_{xx} - \alpha \widehat{\Sigma} \\ \widetilde{\varrho} &= M_{yy} - \widetilde{\beta}' \widetilde{\Omega} \widetilde{\beta} \end{aligned}$$

Thus, the variance-covariance matrix can be estimated as

$$V\{\widetilde{\beta}\} = \frac{1}{G} \widetilde{\Omega}^{-1}[\widetilde{A} + \alpha^2 \widetilde{B}]\widetilde{\Omega}^{-1} \quad (59)$$

where

$$\tilde{A} = M_{xx}(M_{yy} - \tilde{\beta}'\tilde{\Omega}\tilde{\beta} + \tilde{\beta}'\hat{\Sigma}\tilde{\beta} - 2\hat{\sigma}'\tilde{\beta}) + (\hat{\sigma} - \hat{\Sigma}\tilde{\beta})(\hat{\sigma} - \hat{\Sigma}\tilde{\beta})' \quad (60)$$

$$\tilde{B} = \frac{1}{G} \sum_{g=1}^G \frac{1}{n_g} \hat{\Sigma}[(\tilde{q} + \tilde{\beta}'\hat{\Sigma}\tilde{\beta} - 2\hat{\sigma}'\tilde{\beta}) + (\hat{\sigma} - \hat{\Sigma}\tilde{\beta})(\hat{\sigma} - \hat{\Sigma}\tilde{\beta})'] \quad (61)$$

Note that the analogous variance-covariance matrices for the EWALD, EVE, and B2SLS estimators are calculated by evaluating (59) using values of α equal 0 for EWALD, 1 for EVE, and $((N - G)/(N - G + K + 1))(G - K - 1)/G$ for B2SLS.

Table 1: Monte Carlo Experiments

	10%	25%	Median Bias	75%	90%	Median Abs. Err.	Trimmed Mean Bias	Trimmed MAE	90% C.I. Coverage
Panel A: 2 Cohorts									
EWALD	-0.40	-0.34	-0.29	-0.23	-0.18	0.29	-0.29	0.29	0.13
EVE	-0.16	-0.08	0.02	0.14	0.28	0.11	0.04	0.11	0.91
UEVE	-0.18	-0.11	-0.01	0.10	0.22	0.10	-0.00	0.10	0.90
B2SLS	-0.19	-0.11	-0.02	0.09	0.21	0.10	-0.01	0.10	0.90
Panel E: 10 Cohorts									
EWALD	-0.41	-0.35	-0.29	-0.22	-0.17	0.29	-0.29	0.29	0.16
EVE	-0.11	-0.01	0.12	0.29	0.51	0.15	0.15	0.18	0.92
UEVE	-0.20	-0.11	-0.01	0.10	0.24	0.11	-0.00	0.11	0.90
B2SLS	-0.21	-0.13	-0.04	0.08	0.20	0.11	-0.02	0.11	0.88
Panel C: 25 Cohorts									
EWALD	-0.44	-0.37	-0.28	-0.21	-0.13	0.28	-0.29	0.29	0.30
EVE	0.01	0.26	0.62	1.31	2.89	0.71	0.92	0.93	0.91
UEVE	-0.24	-0.14	-0.02	0.13	0.32	0.14	0.00	0.14	0.89
B2SLS	-0.27	-0.17	-0.06	0.07	0.23	0.14	-0.05	0.13	0.86
Panel D: 2 Cohorts, Greater Sampling Error									
EWALD	-0.61	-0.56	-0.50	-0.44	-0.39	0.50	-0.50	0.50	0.01
EVE	-0.26	-0.13	0.05	0.32	0.79	0.20	0.13	0.25	0.93
UEVE	-0.30	-0.19	-0.03	0.18	0.52	0.19	0.02	0.20	0.89
B2SLS	-0.30	-0.20	-0.05	0.16	0.48	0.19	-0.00	0.19	0.87
Panel E: 10 Cohorts, Greater Sampling Error									
EWALD	-0.62	-0.56	-0.50	-0.44	-0.38	0.50	-0.50	0.50	0.01
EVE	-0.21	-0.01	0.30	0.86	2.03	0.37	0.52	0.58	0.97
UEVE	-0.32	-0.20	-0.04	0.20	0.56	0.20	0.02	0.21	0.88
B2SLS	-0.34	-0.23	-0.09	0.13	0.42	0.20	-0.04	0.20	0.84
Panel F: 25 Cohorts, Greater Sampling Error									
EWALD	-0.65	-0.58	-0.50	-0.42	-0.35	0.50	-0.50	0.50	0.05
EVE	-7.04	-3.10	-1.39	1.66	5.85	2.55	-0.72	3.09	0.78
UEVE	-0.39	-0.25	-0.07	0.24	0.72	0.25	0.02	0.27	0.87
B2SLS	-0.43	-0.30	-0.15	0.08	0.42	0.24	-0.09	0.24	0.81

NOTE: Results for 10000 Monte Carlo replications.

Table 2: Means of Variables (Standard Deviations in Parentheses)

Census year 1980	.36	(.48)
Census year 1990	.34	(.47)
Census year 2000	.30	(.46)
Number of Children Under Age 5	.21	(.52)
Number of Children Aged 5+	.91	(1.13)
Log(wage)	2.57	(.65)
Log(hours)	7.59	(.49)
Education<12	.11	(.31)
Education=12	.33	(.47)
Education 13-15	.27	(.45)
Education>15	.29	(.46)
White	.89	(.31)
Married	.75	(.43)

NOTE: The sample includes 2,915,397 observations.

Table 3: Labor Supply Estimates by Education Level

<i>Education Less than 12 Years (N = 309,862)</i>				
	EWALD	EVE	UEVE	B2SLS
Log Wage	.37*	.67*	.60*	.61*
	(.04)	(.14)	(.08)	(.08)
Children >5	-.04*	.19	-.18*	-.16*
	(.01)	(.13)	(.06)	(.05)
Children <5	-.34*	.53	-.89*	-.85*
	(.06)	(.51)	(.25)	(.21)
Married	.54*	-2.72	1.90*	1.68*
	(.10)	(1.44)	(.69)	(.56)
<i>12 Years of Education (N = 948,523)</i>				
	EWALD	EVE	UEVE	B2SLS
Log Wage	.37*	.55*	.46*	.45*
	(.03)	(.05)	(.04)	(.03)
Children >5	-.05*	-.22*	-.11*	-.11*
	(.01)	(.05)	(.02)	(.02)
Children <5	-.36*	-.92*	-.56*	-.56*
	(.04)	(.15)	(.06)	(.06)
Married	.70*	2.31*	1.27*	1.26*
	(.09)	(.45)	(.18)	(.18)
<i>Education of 13 – 15 Years (N = 800,969)</i>				
	EWALD	EVE	UEVE	B2SLS
Log Wage	.40*	.31*	.44*	.43*
	(.04)	(.15)	(.05)	(.05)
Children >5	.00	-.22	-.05*	-.06*
	(.01)	(.12)	(.02)	(.02)
Children <5	-.17*	-.97*	-.34*	-.35*
	(.04)	(.44)	(.08)	(.08)
Married	.55*	2.96*	1.04*	1.08*
	(.10)	(1.38)	(.25)	(.25)
<i>16 or More Years of Education (N = 856,043)</i>				
	EWALD	EVE	UEVE	B2SLS
Log Wage	.46*	.72*	.59*	.58*
	(.05)	(.14)	(.08)	(.08)
Children >5	-.01	-.06*	-.04*	-.04*
	(.01)	(.02)	(.01)	(.01)
Children <5	-.12*	-.19*	-.17*	-.17*
	(.03)	(.08)	(.05)	(.05)
Married	.54*	.57	.60*	.62*
	(.11)	(.36)	(.19)	(.19)

NOTE: Also included in the regressions are cohort dummies, and year dummies. Standard errors in parentheses.
 * indicates significant at the 5% level

Table 4: Labor Supply Estimates by Race

<i>Non-Whites (N = 307,846)</i>				
	EWALD	EVE	UEVE	B2SLS
Log Wage	.20*	.36*	.35*	.37*
	(.04)	(.15)	(.07)	(.07)
Children >5	.03*	.33	-.09	-.09
	(.01)	(.19)	(.06)	(.06)
Children <5	-.18*	.89	-.58*	-.57*
	(.05)	(.64)	(.20)	(.19)
Married	.41*	-2.37	1.28*	1.24*
	(.09)	(1.52)	(.47)	(.44)
<i>Whites (N = 2,607,551)</i>				
	EWALD	EVE	UEVE	B2SLS
Log Wage	.36*	.38*	.37*	.37*
	(.03)	(.03)	(.03)	(.03)
Children >5	-.04*	-.08*	-.06*	-.06*
	(.01)	(.01)	(.01)	(.01)
Children <5	-.27*	-.41*	-.35*	-.35*
	(.04)	(.05)	(.04)	(.04)
Married	.87*	1.29*	1.10*	1.10*
	(.11)	(.16)	(.14)	(.14)

NOTE: Also included in the regressions are cohort dummies, and year dummies. Standard errors in parentheses.

* indicates significant at the 5% level