UCD CENTRE FOR ECONOMIC RESEARCH

WORKING PAPER SERIES

2006

Small Sample Bias in Synthetic Cohort Models of Labor Supply

Paul J Devereux, University College Dublin

WP06/06

May 2006

UCD SCHOOL OF ECONOMICS
UNIVERSITY COLLEGE DUBLIN
BELFIELD DUBLIN 4

Small Sample Bias in Synthetic Cohort Models of Labor Supply

Paul J. Devereux[1]

April 26, 2006

[1]School of Economics, UCD Dublin (devereux@ucd.ie). I thank Dan Ackerberg, Joshua Angrist, Sandy Black, Janet Currie, Jin Hahn, Katerina Kyriazidou, Robert Moffitt, Donal O'Neill, and Olive Sweetman for helpful comments.

Abstract

In synthetic cohort models (cross-sectional data grouped at the cohort and year level), researchers often ignore potential biases induced by sampling error because they have 100 or 200 observations per group. I investigate small sample biases in the context of two synthetic cohort labor supply applications – a model of intertemporal labor supply of men (similar to that of Browning, Deaton, and Irish, 1985) and a female labor supply model (similar to that of Blundell, Duncan, and Meghir, 1998). My approach is to use the Current Population Survey to compare the estimates when group sizes are extremely large to those that arise from randomly drawing subsamples of observations from the large groups. This provides a natural framework for examining the extent of small sample biases and the group sizes required so that small sample biases are negligible. I augment this approach with Monte Carlo analysis so as to precisely quantify biases and coverage rates. I find that, in these two applications, thousands of observations per group are required before small sample issues can be ignored in estimation. In these applications, sampling error leads one to underestimate intertemporal labor supply elasticities for men, and conclude that the income response of female labor supply is zero or tiny when in fact it is quite large.

1 Introduction

Research on labor supply often uses panel data to control for unobserved individual effects. While this approach has many advantages, panel data are not always available. An alternative approach uses synthetic cohorts (cross-sectional data grouped at the cohort and year level) to control for time-invariant but unobserved effects.[1] Because this synthetic cohort approach provides consistent estimates even if the individual effects are correlated with the explanatory variables, it is widely used in labor supply estimation and in other areas of labor economics.

Unfortunately, these models are potentially biased because of sampling error when group sizes are small (Deaton, 1985). Results in the theoretical literature suggest that sampling error is not a problem in practice when there are at least 100 to 200 observations per group (Verbeek and Nijman 1992, 1993) and empirical researchers have typically not corrected for sampling error with group sizes of this magnitude.[2] In this paper, I show that this conclusion is not necessarily correct.

I investigate small sample biases in the context of two synthetic cohort applications – a model of intertemporal labor supply of men, and a female labor supply model. These applications contribute to the empirical literature by providing new estimates of labor supply elasticities using U.S. data from the Current Population Survey (CPS). My approach is to compare the estimates when group sizes are extremely large to those that arise from randomly drawing subsamples of observations from the large groups. This provides a natural framework for examining the extent of small sample biases and the group sizes required so that small sample biases are negligible. I augment this approach with Monte Carlo analysis so as to precisely quantify biases and coverage rates. I find that, in these two applications, thousands of observations per group are required before small sample issues can be ignored in estimation. In these data, sampling error leads one to underestimate intertemporal labor supply elasticities for men, and conclude that the income response of female labor supply is zero or tiny

[1] Even when panel data are available, sample sizes tend to be small, and they often suffer from serious attrition that leads to selection problems. Also, measurement error can be a major problem in fixed effects models in true panels.

[2] Blundell, Duncan, and Meghir (1998) report an average group size of 142 individuals in their study of female labor supply in the United Kingdom. Browning, Deaton, and Irish (1985) study male labor supply with an average group size of 190. Propper, Rees, and Green (2001) study the demand for private medical insurance using a synthetic cohort approach and group sizes that average about 80 people. None of these studies use estimation methods that are robust to sampling error.

when in fact it is quite large.

A further goal of the paper is to examine the performance of possible indicators of small sample bias in synthetic cohort models. I find that first stage F statistics are of limited value as it is not clear how large they need to be. I propose a related indicator that is closely related to the reliability ratio from the classical measurement error literature but is based on small-sample bias calculations rather than asymptotics as the number of groups goes to infinity. I apply the bias indicator to the two labor supply applications and also to the important female labor supply study of Blundell, Duncan, and Meghir (1998).

2 Model and Estimators

Consider the following linear model where i indexes individual, and t indexes time:

$$y_{it} = x'_{it}\beta + \alpha_i + u_{it} \qquad i = 1, ..., N \qquad t = 1, ..., T \tag{1}$$

where x_{it} is a k-dimensional column vector of exogenous variables, β is a k-dimensional parameter vector, and α_i is an individual effect. The error term, u_{it}, is assumed to be uncorrelated with x_{it} and α_i. The available data are a set of repeated cross-sections. Since the same individuals are not observed over time, one cannot 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 between 1960 and 1965. In a finite sample, taking means by cohort-year gives the following:

$$\overline{y}_{ct} = \overline{x}'_{ct}\beta + \overline{\alpha}_{ct} + \overline{u}_{ct} \tag{2}$$

The sample mean of x for group ct ($\overline{x}_{ct}$) is the mean of x over sample observations in cohort c at time t. The standard synthetic cohort approach is to replace $\overline{\alpha}_{ct}$ with cohort dummies and estimate equation (2) 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. 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 ($\overline{\alpha}_{ct}$) 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 $cov(\overline{\alpha}_{ct} - \alpha_c, \overline{x}_{ct}) \neq 0$, where α_c is the true cohort effect.

Taking expectations of equation (1) conditional on cohort and year gives the cohort population version:

$$y_{ct} = x'_{ct}\beta + \alpha_c + u_{ct} \tag{3}$$

Here y_{ct} and x_{ct} denote the population means of y and x, respectively, in cohort c at time t. 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 $\overline{x}_{ct}$ and $\overline{y}_{ct}$ are error-ridden measures of x_{ct} and y_{ct}.[3]

To define the estimators, I follow Deaton in stacking cohorts and surveys into a single index g (for cohort-year groups), that runs from 1 to G (which equals CT), and absorbing the cohort dummies into the x matrix. So now, equation (2) becomes

$$\overline{y}_g = \overline{x}'_g\beta + \overline{u}_g \tag{4}$$

Assume that the sampling error has the following structure, where y_g is the population mean of y in group g, and x_g is the equivalent population mean of x:

$$\begin{pmatrix} \overline{y}_g - y_g \\ \overline{x}_g - x_g \end{pmatrix} \sim iid \left(0, \frac{1}{n_g} \begin{bmatrix} \sigma_{00} & \sigma' \\ \sigma & \Sigma \end{bmatrix} \right) \tag{5}$$

[3]Ridder and Moffitt (2003) provide a similar exposition of the source of EWALD bias.

One can define the EWALD estimator as:

$$\beta^{EWALD} = \left(\sum_{g=1}^{G} n_g \overline{x}_g \overline{x}'_g \right)^{-1} \left(\sum_{g=1}^{G} n_g \overline{x}_g \overline{y}_g \right) \tag{6}$$

Deaton (1985) shows that one can consistently estimate β as G goes to infinity with n_g fixed using the following errors in variables estimator (EVE):[4]

$$\beta^{EVE} = \left(\sum_{g=1}^{G} n_g \overline{x}_g \overline{x}'_g - G\widehat{\Sigma} \right)^{-1} \left(\sum_{g=1}^{G} n_g \overline{x}_g \overline{y}_g - G\widehat{\sigma} \right) \tag{7}$$

Here, $\widehat{\Sigma}$ and $\widehat{\sigma}$ are sample estimates of the relevant population parameters.

Devereux (2006) shows that the EVE estimator is biased in finite samples and suggests an alternative errors-in-variables estimator (UEVE) that is approximately unbiased to order $\frac{1}{N}$:

$$\beta^{UEVE} = \left(\sum_{g=1}^{G} n_g \overline{x}_g \overline{x}'_g - (G - K - 1)\,\widehat{\Sigma} \right)^{-1} \left(\sum_{g=1}^{G} n_g \overline{x}_g \overline{y}'_g - (G - K - 1)\,\widehat{\sigma} \right) \tag{8}$$

In addition to the estimators discussed above, I also report estimates from a version of EVE suggested by Verbeek and Nijman (1993). They show that EVE2 is consistent as the number of cohorts goes to infinity:

$$\beta^{EVE2} = \left(\sum_{g=1}^{G} n_g \overline{x}_g \overline{x}'_g - \left(\frac{T-1}{T} \right) G\widehat{\Sigma} \right)^{-1} \left(\sum_{g=1}^{G} n_g \overline{x}_g \overline{y}_g - \left(\frac{T-1}{T} \right) G\widehat{\sigma} \right) \tag{9}$$

Additionally, I implement the Limited Information Maximum Likelihood Estimator (LIML). The LIML estimator (Anderson and Rubin, 1949, 1950) is well known and provides a useful comparison to EWALD.

[4] Angrist and Krueger (1995), Bekker (1994), and Bekker and van der Ploeg (1999) also take a group asymptotic approach to consistency.

3 Indicators of small sample bias in EWALD

It is useful to have indicators of the likely bias of EWALD that researchers can report in practical applications and readers of applied work can calculate from published results. In this section, I suggest one such measure.

One approach to assessing the bias of EWALD is to compare the EWALD estimate to that of the bias-corrected estimators. Of these, Deaton's EVE is biased in finite samples (Verbeek and Nijman, 1993; Devereux, 2006), and LIML, as shall be seen in the applications, is sensitive to misspecification. Thus, a natural comparison is to UEVE, which is approximately unbiased in finite samples. However, when group sizes are small and identification is weak, UEVE estimates can be highly variable and any one estimate may give little indication about the magnitude and direction of bias of EWALD.

My approach is to concentrate on the denominator of the UEVE and EWALD estimators. Define Λ_{kk} as the element in the kth row and kth column of $\left(\sum_{g=1}^{G} n_g \overline{x}_g \overline{x}_g'\right)^{-1}$, and Ψ_{kk} as the element in the kth row and kth column of $\left(\sum_{g=1}^{G} n_g \overline{x}_g \overline{x}_g' - (G-K-1)\,\widehat{\Sigma}\right)^{-1}$. Then, for the kth element of $\overline{x}_g$ consider

$$\lambda_k = \frac{\Lambda_{kk}}{\Psi_{kk}} \tag{10}$$

If λ_k equals one, it suggests that there is no bias in the EWALD estimator of β_k coming through the denominator. As sampling error in $\overline{x}_g$ becomes more important, λ_k goes towards 0. In perverse cases where sampling error is really severe and the relevant element of $\left(\sum_{g=1}^{G} n_g \overline{x}_g \overline{x}_g' - (G-K-1)\,\widehat{\Sigma}\right)^{-1}$ is negative, λ_k is negative.

To get a sense about how λ_k differs from an indicator based on the equivalent group-asymptotic approximation, consider an example where there is one right hand side variable plus a constant. In this situation,

$$\lambda_k = \frac{\Lambda_{kk} - (G-K-1)\widehat{\Sigma}_{kk}}{\Lambda_{kk}} \tag{11}$$

while the equivalent indicator based on group asymptotic formula is $\frac{\Lambda_{kk} - G\widehat{\Sigma}_{kk}}{\Lambda_{kk}}$, the reliability ratio from the classical measurement error literature. That is, λ_k builds on the standard formulae by using a small sample correction rather than the group-asymptotic approximation.

In addition, λ_k is closely related to the first stage F statistic from the instrumental variables literature. The EWALD estimator has been shown (for example, by Durbin (1954) in the two-group case and Angrist (1991) in the general case) to be identical to the two stage least squares estimator (2SLS) in the case where estimation uses the microdata and group dummies are used as instruments. Thus, it is natural to report the F statistic for the excluded instruments in the first stage (the regression of x_{it} on the group indicators) as a measure of instrumental relevance (Staiger and Stock, 1997).

The first stage F statistic is closely related to the λ_k measure. For tractibility, consider the case where there is only the constant term plus one additional right hand side variable (the kth variable) and there are the same number of observations in each group. The F statistic for the excluded instruments in the first stage regression is easily calculated as

$$F_k = \left(\frac{1}{1-\lambda_k}\right)\left(\frac{G-K-1}{G-1}\right) \tag{12}$$

The advantage of λ_k over the F statistic is that it provides an estimate of the degree of bias. Both the F statistic and λ_k share the disadvantage that a weak value does not necessarily imply bias. In instrumental variables estimation, if the true coefficient and the OLS estimate are the same, 2SLS is not biased even if the F statistic is very low. Analogously, in the synthetic cohort case, if the explanatory variables are uncorrelated with the unobserved heterogeneity, EWALD is not biased even if λ_k is close to zero. That is, since we are only capturing bias from the denominator and in general $\sigma \neq 0$, λ_k does not provide an exact approximation to the degree of bias in EWALD. However, a value of λ_k that is much smaller than one suggests that researchers should take small sample issues seriously.

4 Application 1: Intertemporal Male Labor Supply

MaCurdy (1981) shows that the intertemporal Frisch labor supply curve under certainty is approximated as

$$h_{it} = \delta w_{it} + \alpha_i + \eta_t + u_{it} \tag{13}$$

where h_{it} is log hours worked, w_{it} is the log real wage, α_i is an individual effect that controls for the marginal utility of wealth, and η_t is a fixed time effect. 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) but obtain somewhat imprecise estimates of the intertemporal elasticity (δ).[5]

Browning, Deaton, and Irish (1985) take the alternative approach of using repeated cross-sectional data from the British Family Expenditure Survey (FES) to estimate the intertemporal wage elasticity for men. This works in life-cycle models because a key source of omitted variables bias, the marginal utility of wealth, is time-invariant and fixed within cohorts. I take a similar approach using the NBER extracts of the Merged Outgoing Rotation Groups (MORG) from the Current Population Survey (CPS) over the period 1979-1993. The sample consists of men who are aged 21 to 50 in 1979. Thus, the men are aged 36 to 65 in 1993. I include all individuals who report being employed in the private or public sector and exclude the self-employed and those who are not working for pay. The hours measure is “How many hours per week does ... USUALLY work at this job?” The wage measure for salaried workers is usual weekly earnings divided by usual weekly hours.[6] For hourly workers, the wage measure is the directly reported usual hourly wage. I exclude men who report working more than 80 hours per week or having real hourly wages (in 1983 dollars) less than $1 or greater than $100.

The men are divided into 6 evenly divided 5-year birth cohorts. Thus, there are 90 (6 by 15) cohort-year

[5] Angrist (1991) also uses the PSID but deals with measurement error by carrying out estimation using yearly means of the variables.

[6] The earnings measure is “How much does ... usually earn per week at this job before deductions?” This includes overtime, tips, and commissions and is topcoded at $999 between 1979 and 1988, and at $1923 from 1989 to 1993. I impute earnings for these individuals as 1.33 times the topcoded value.

groups. The labor supply equation comes from taking expectations of (13) by cohort and year:

$$h_{ct} = \delta w_{ct} + \alpha_c + \eta_t + u_{ct} \tag{14}$$

Here, α_c denotes indicator variables for the 6 cohorts, and the intertemporal wage elasticity equals δ.

As described earlier, the analysis involves starting with very large group sizes. The estimates obtained using the large group sizes are considered to be (approximately) free of small-sample bias. Then, I examine the consequences for EWALD of drawing random subsamples of observations and carrying out the estimation.[7] I report results for the following average number of observations per group: 100, 200, 500, 1000, and 2000. These constitute approximately 1%, 2%, 5%, 10%, and 20% of the sample respectively. In each case, I carry out 1000 replications and report features of the resulting distribution of estimates.

4.1 Results for Application 1 Using The Actual Data

There are 883,610 observations in total on 90 groups, an average of 9818 observations per group. With such large groups, one would expect the sample means to be close to the population means and small sample bias to be small. Note that the OLS estimate of the wage elasticity from the microdata is 0.0290.

The estimates and standard errors for the full sample are in table 1.[8] The EWALD estimate of the wage elasticity, δ, is 0.351. The other estimators give similar estimates of the wage elasticity in the region of 0.35 - 0.37. The fact that these estimates are so similar suggests that there is very little small sample bias in EWALD with these huge group sizes. Furthermore, the estimated value of λ is 0.99 which is close enough to one as to suggest that there is very little bias. The intertemporal wage elasticity is larger than that found by Browning et al. (1985) for British data but somewhat smaller than the U.S. estimates of Angrist (1991) from the PSID.

[7] This strategy is similar to that used by Buchinsky (1995) in his analysis of quantile regression models. As in his paper, the sampling is carried out with replacement.

[8] Because of the group structure of the data, the standard errors are based on the White "sandwich" variance estimator that allows errors to be correlated within group. For example, the variance covariance matrix for the EWALD estimator is estimated as $\left(\sum_{g=1}^{G} n_g \bar{x}_g \bar{x}_g'\right)^{-1} \Omega \left(\sum_{g=1}^{G} n_g \bar{x}_g \bar{x}_g'\right)^{-1}$ where $\Omega = \sum_{g=1}^{G} n_g \bar{x}_g e_g e_g \bar{x}_g'$ and e_g is the residual from the group-level regression.

Table 2 contains the wage elasticity estimates for different sample sizes. The median and trimmed mean of the EWALD estimates clearly increase with sample size from 0.18 with average group sizes of 100 to 0.33 with average group sizes of 2000. Thus, the EWALD estimates are very sensitive to sample size and the biases appear sizeable in the smaller samples. In contrast, the other estimators are more robust to small sample bias – the mean and median do not increase as the sample size increases. However, the EVE estimates tend to decrease as the group sizes increase, reflecting the fact that EVE is biased in finite samples.

4.2 Monte Carlo Based on Application 1

While the results above are indicative, one is limited by the fact that the true wage elasticity is unknown. Therefore, I now carry out a Monte Carlo simulation based on the above application. By construction, the artificial data are designed to share the characteristics of the data in the CPS. I start by taking the 90 cell means of the log wage from the full sample and treat these as the population means (w_{ct}). The next step is to choose population means for the dependent variable (h_{ct}): Given w_{ct}, I fix the values of δ, α_c (the coefficients on the cohort indicators), and η_t (the coefficients on the year dummies) to equal the EWALD estimates from the regression of $\overline{h}_{ct}$ on w_{ct} and the fixed cohort and year effects. Then, I construct h_{ct} as equal to $\delta w_{ct} + \alpha_c + \eta_t$.

Once I have the population means for each group, the Monte Carlo simulations involve repeatedly drawing values of u_{ict} and v_{ict} and estimating δ.

$$h_{ict} = h_{ct} + u_{ict} \tag{15}$$

$$w_{ict} = w_{ct} + v_{ict} \tag{16}$$

The u, v errors are assumed to be bivariate normal and the covariance matrix of u, v used in the simulations is estimated from the within-group variances in the data. Note that, with this setup, all the estimators considered are consistent as group sizes go to infinity with the number of groups fixed.

In all tables, I report quantiles (10%, 25%, 50%, 75%, 90%) of the distribution of the estimators 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 the errors-in-variables 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 the mean bias and mean absolute error for these trimmed distributions. Note that in all cases considered in this paper, the trimmed mean of the EWALD estimator is extremely close to the untrimmed mean. The reported 90% coverage rates are calculated using conventional asymptotic estimates of variance.[9]

4.3 Results from the Monte Carlos Based on Application 1

The results from the Monte Carlos are in table 3. The true value of the wage elasticity is 0.351. The means and percentiles in table 3 are expressed as deviations from the true value. The first thing to notice is how similar the Monte Carlo estimates are to the estimates from the empirical application in table 2. For example, the EWALD mean bias of -0.166 in panel A implies an average coefficient of 0.185 (0.351 - 0.166) that is very similar to the equivalent value of 0.176 in panel A of table 2. This suggests that the Monte Carlo design has maintained essential features of the data. The EWALD estimator is almost unbiased in the samples of 10,000, a result that confirms that group sizes of this magnitude are sufficient to get approximately unbiased estimates in this application. The estimates for the smaller samples imply average biases of 88% for the groups of 100, 47% for the groups of 200, 19% for the groups of 500, 10% for the groups of 1000, and 5% for the groups of 2000.

In contrast to EWALD, LIML and UEVE perform quite well in terms of median and trimmed mean bias in

[9] To do so, I utilize the equivalence of EWALD to 2SLS where the instrument set, Z, is a set of group dummies. Thus, for EWALD, I use conventional 2SLS standard errors. Likewise, for LIML, conventional asymptotic standard errors are used. Denote the N by K matrix of right hand side variables in the microdata as X, and the vector of individual-level residuals as e (where e differs across estimators). For the errors-in-variables estimators of β, $Var(\beta)$ is estimated by $\frac{e'e}{N}(X'C'X)^{-1}$ where C for EVE, UEVE, and EVE2 equals $(BX(X'B'X)^{-1}X'B')$. Here $B^{EVE} = P_z - (\frac{G}{N-G})(I(N) - P_Z)$, $B^{UEVE} = P_z - (\frac{G-K-1}{N-G})(I(N) - P_Z)$, and $B^{EVE2} = P_z - (\frac{T-1}{T})(\frac{G}{N-G})(I(N) - P_Z)$. Note that the asymptotic variance of the EVE estimator is calculated using the fact that each group has the same number of observations and that EVE in this context is equivalent to the Jackknife Instrumental Variables Estimator (JIVE) estimator of Angrist, Imbens, and Krueger (AIK, 1999). The standard errors for EVE here are exactly those used for JIVE by AIK.

both small and large samples. The superiority of UEVE (based on finite sample formulae) compared to EVE (based on the group asymptotic approximation) is very clear in the table.

The final column of table 3 provides coverage rates for the various estimators. The coverage rates for the EWALD estimator are very poor in the smaller samples. For example, the 90% confidence intervals exclude the truth in 98% of cases in the samples of 100, 93% of cases in the samples of 200, and 73% of cases in the samples of 500. Not only are the EWALD estimates biased in small samples, researchers may be misled because the standard asymptotic confidence intervals generally exclude the truth. The coverage rates of the non-EWALD estimators are quite close to the nominal size of 0.90, and (as expected) are especially good with larger group sizes.

4.3.1 Diagnosing Small Sample Bias in the EWALD Estimator

The median value of the first stage F statistic, reported in table 3, increases as the sample size increases. However, it is difficult to know what size F statistic should be required. In the groups of 1000, about 50% of draws provide F statistics greater than 10. For these draws, the average value of the wage elasticity is 0.313 (approximately a 12% bias), and the coverage rate of the EWALD estimator is less than 0.5. Thus, even requiring the F statistic to be greater than 10 would lead to incorrect inference.

The values for the attenuation bias indicator, λ, do provide a warning that small sample bias may be present. As expected, the median value of λ for the log wage increases on average as group sizes fall. The median values are also quite a bit less than one – 0.49 for the groups of 100, 0.65 for the groups of 200, and up to 0.99 for the full sample. While the estimated λ differs across draws, it does not vary enough to seriously mislead: For example, the value of λ lies between [-0.03, 0.70] for the groups of 100, [0.36, 0.78] for the groups of 200, [0.71, 0.88] for the groups of 500, and [0.85, 0.93] for the groups of 1000. Thus, irrespective of which sample a researcher draws, there is ample warning from these statistics that small sample bias is likely if the EWALD estimator is used. Note also that λ provides a good approximation of the exact amount of bias – the EWALD estimates are approximately equal to $\lambda\delta$. This is the case in this application because the σ term plays little role, as indicated by the OLS estimate of the intertemporal elasticity being very close to zero.

5 Application 2: Female Labor Supply

Blundell and MaCurdy (1999) provide a summary of empirical work on the labor supply of married women. There is a large cross-sectional literature that treats wages and other income as exogenous (Ransom 1987) or uses age or education as instruments for wages (Kooreman and Kapteyn, 1986). However, if taste for work is related to wages, other income, or education, cross-sectional estimation likely provides inconsistent estimates of wage and income elasticities.

A natural approach to identification in this context is to use exogenous shifts in average wages when individuals are grouped by cohort and year. Blundell, Duncan, and Meghir (BDM 1998) estimate labor supply functions for married and cohabiting women in the United Kingdom over the period 1978-1992 using this approach in the FES.[10] Unlike cross-sectional approaches, they allow for differences in preferences across education and age groups that are assumed to remain constant over time. They identify labor supply elasticities using changes in relative wages across groups of women defined by education and birth cohort. The idea is that differential wage growth across birth cohorts by education group reflects changes in the demand for labor, possibly skill-biased technological change, and can be excluded from the labor supply equation.

I take a similar approach using the same 1979-1993 CPS files as in the male application. I group married working women aged between 20 and 50 by birth cohort and by education into 8 groups (4 10-year birth cohorts and 2 education groups). The four birth cohorts consist of women born in 1931-1940, 1941-1950, 1951-1960, and 1961-1970. For education groups, I split the sample between women with high school or less, and women with more than high school (including those with some college). The wage and hours measures are the same as in the male labor supply application and husband's weekly earnings is the measure of other income. I report

[10] They actually implement the estimator on the individual-level data using cohort-year indicators as instruments, a process that is equivalent to grouping by cohort-year.

results for the wage elasticity and the income elasticity (evaluated at the mean levels of wages, hours, and income) from the following specification:

$$H_{ct} = \alpha + \beta w_{ct} + \delta y_{ct} + f_c + f_t + \epsilon_{ct} \tag{17}$$

Here H_{ct} is hours worked, w_{ct} is the log wage, y_{ct} is other income, f_c represents cohort fixed effects, and f_t represents year fixed effects.

Note that the setup is quite similar to BDM who also group women aged between 20 and 50 by birth cohort and by education into 8 groups (4 10-year birth cohorts and 2 education groups).[11] BDM include some controls for number of children that are not present in the CPS files and use estimated after-tax wages rather than the pre-tax wages utilized here.[12]

5.1 Results from Application 2 Using the Actual Data

There are 498,667 observations in total on 110 groups, an average of 4533 observations per group. I first assess whether the group sizes are sufficiently high so that there is no small sample bias in the EWALD estimator. The OLS estimates of the wage elasticity and income elasticity from the microdata are 0.12 and -0.03 respectively.

The estimates for the full sample are in table 4. For brevity, I will concentrate on the results for the income elasticity. The EWALD estimate of the income elasticity is -0.28 using the full sample. The EVE, EVE2, and UEVE estimators all give estimates of the income elasticity between -0.36 and -0.40. Thus, it appears likely that even with 4533 observations per group on average, there is substantial small sample bias in the EWALD estimator. The value of the λy bias indicator for other income (0.72) is a lot less than one, also suggesting that the EWALD estimate may be biased in the full sample. The LIML estimate is completely different from all the

[11] Their four cohorts consist of individuals born in 1930-1939, 1940-1949, 1950-1959, and 1960-1969. Their two education groups are those who left school at the minimum legal age and those who continued beyond the minimum. Their measure of hours of work is “usual weekly hours, including usual overtime” and the wage is constructed as “usual weekly earnings, including usual overtime pay” divided by “usual weekly hours, including usual overtime”. Their measure of other income is usual weekly household nondurable consumption less the woman’s usual weekly earnings.

[12] BDM also include a selection correction for non-participation in some specifications. For simplicity, I report results without the correction. Including a selection correction does not change any of my conclusions.

others, suggesting that there are problems using LIML in this situation.

The magnitudes of the UEVE estimates are in keeping with much of the prior literature on female labor supply. Evaluated at the mean female wage of \$7.13, the mean hours worked of 35.40, and the mean weekly other income of \$414.26, the compensated wage elasticity for women is 0.55. Hyslop (2001) and Pencavel (1998) find compensated wage elasticities of about 0.4 to 0.5; Ransom (1987) and Hausman and Ruud (1984) find elasticities of about 0.75. In contrast, BDM find much lower compensated wage elasticities for the UK.

The results from randomly generating average group sizes of 200, 500, 1000, and 2000 are in tables 5 and 6. As in the male labor supply example, there is clear evidence of small sample bias in the EWALD estimator. For example, the mean income elasticity is -0.28 for the full sample, -0.19 for the groups of 2000, -0.16 for the groups of 1000, -0.12 for the groups of 500, and -0.09 for the groups of 200.

5.2 Monte Carlo Based on Application 2

I now carry out a Monte Carlo simulation based on the above application. By construction, the artificial data are designed to share the characteristics of the CPS data. I start by taking the 110 cell means from the full sample. Because there is reason to believe that these do not equal the population means, I adjust them as follows:

$$w_{ct} = 0.75\overline{w}_{ct} + 0.25\overline{w} \quad (18)$$

$$y_{ct} = 0.75\overline{y}_{ct} + 0.25\overline{y} \quad (19)$$

In this equation $\overline{w}$ and $\overline{y}$ refer to the means of w and y over the whole sample. The next step is to choose the population means for the dependent variable (H_{ct}): Given w_{ct}, I fix the values of β, δ, f_c, and f_t to equal the EWALD estimates from the regression of $\overline{H_{ct}}$ on w_{ct} and y_{ct}. Then I construct H_{ct} as equal to $\alpha+\beta w_{ct}+\delta y_{ct}+f_c+f_t$.

The Monte Carlo simulations involve repeatedly drawing values of u_{ict}, v_{ict}, e_{ict} and estimating β.

$$H_{ict} = H_{ct} + u_{ict} \tag{20}$$

$$w_{ict} = w_{ct} + v_{ict} \tag{21}$$

$$y_{ict} = y_{ct} + e_{ict} \tag{22}$$

The u, v, e errors are assumed to be trivariate normal and the covariance matrix of u, v, e used in the simulations is estimated from the within-group variances and covariances in the data.

When there is more than one explanatory variable, F statistics for each individual explanatory variable can be misleading. It is important that the excluded instruments have explanatory power for each variable, conditional on the predicted values of the other variables. Therefore, I report a test that the rank of the first stage coefficients of the excluded instruments is at least as great as the number of endogenous explanatory variables. It is utilized by BDM and is one of the tests of rank proposed by Robin and Smith (1994).[13]

5.3 Results of the Monte Carlo Simulations

The results of the Monte Carlos are in tables 7 and 8 (for brevity, I just describe results for the income elasticity in table 8). The values of β and δ used to generate the data imply an income elasticity of -0.363 and an uncompensated wage elasticity of 0.331. As in the previous monte carlo, the reported percentiles and trimmed means are deviations from the true elasticity. One can see that the EWALD estimator is very biased with biases of 0.28 when average group size is 200, 0.23 in the groups of 500, 0.18 in the groups of 1000, 0.13 in the groups of 2000, 0.07 in the groups of 4500 (approximately the size of the full sample), and 0.04 when each group size

[13] This exposition follows closely that of BDM: Let $\widehat{\Pi}$ be a consistent and asymptotically normal estimator of a p by k reduced form matrix of the first stage parameters on the excluded instruments (k endogenous variables and p excluded instruments). Let Ω be the covariance matrix of $\sqrt{n}Vec(\widehat{\Pi})$ where n is the sample size. Calculate the eigenvalues of $\widehat{\Pi}'\widehat{\Pi}$. Under the null hypothesis that the rank of the matrix is r, the test statistic is the sum of the k-r smallest eigenvalues of $\widehat{\Pi}'\widehat{\Pi}$. This has for limiting distribution a mixture of (p - r)(k - r) one-degree-of-freedom chi-square distributions. The weights are calculated as the characteristic roots of the matrix $(D'_{k-r} \otimes C'_{p-r})\Omega(D_{k-r} \otimes C_{p-r})$ where D_{k-r} is a k by (k - r) matrix formed by the eigenvectors corresponding to the k - r lowest eigenvalues of $\widehat{\Pi}'\widehat{\Pi}$. Analogously, C_{p-r} is a p by (p - r) matrix formed by the eigenvectors corresponding to the p - r lowest eigenvalues of $\widehat{\Pi}\widehat{\Pi}'$.

is 10,000. Given the true value of the income elasticity of -0.36, these biases are quite large. The coverage rates of the EWALD estimator are also very poor even when the group sizes get very large.

As in the previous Monte Carlo, LIML and UEVE perform well in terms of median bias, with LIML being clearly best in terms of median bias when the group sizes are very small.[14] Once again, UEVE is clearly superior to EVE and EVE2 in terms of trimmed mean and median bias as well as in terms of median absolute error. UEVE is approximately median unbiased except in the groups of 200. In this case it appears that the model is sufficiently poorly identified that the UEVE bias correction does not work. However, the coverage rates of the UEVE estimator remain excellent so the lack of information in the sample about the value of the income elasticity is conveyed to the researcher. In contrast, note that the EWALD coverage rate is zero and so will tend to mislead the researcher into believing that the income elasticity is very small. The LIML coverage rates are a little low in the smaller samples but become close to 0.9 as the sample sizes grow.[15]

The F statistics for the wage and other income increase with sample size. They are both substantial for the groups of 4500 and the groups of 10000. Thus, as in the first application, sizeable first stage F statistics do not ensure that there is no small sample bias. The proportion of times the p value from the rank test is less than 0.01 is also reported in table 8 (as $\Pr op(P < 0.01)$). It also appears of limited usefulness as the p values are very low in the larger samples despite sizeable biases. The values for the attenuation biases λ, do provide a warning that small sample bias may be present. The λs increase with sample size and are much less than one even for the large samples, providing a useful warning to researchers about biases in the EWALD estimator despite the large group sizes.[16]

[14]The strong peformance of LIML contrasts with its huge spread when the actual data are used in tables 5 and 6. My experimentation suggests that the presence of a group-specific random effect (ϵ_{ct}) in the actual data causes LIML to behave strangely (when monte carlos are carried out with a group random effect, LIML performs very poorly). It appears that LIML can be very sensitive to these random effects and this is problematic as such error components are likely in synthetic cohort models. In contrast, it is easy to show that EVE and UEVE are group-asymptotically consistent in the presence of a random group-specific error component.

[15]The poor LIML coverage in the smaller samples is consistent with the theoretical arguments of Bekker (1994). In the literature, there is some Monte Carlo evidence of poor LIML coverage using asymptotic standard errors when there are many weak instruments (for example, Flores-Lagunes, 2002, Table 18).

[16]Note that the EWALD bias is not as great as predicted by λ because the σ term offsets somewhat the bias coming through the denominator in this application.

5.4 Calculating λ in BDM's Data

One nice feature of λ is that it can be approximated even in cases when one does not have the micro-data; in this section I calculate it using information from the paper by BDM.[17] To do so, I take advantage of the fact that BDM are extremely thorough in reporting the reduced forms in their paper. In particular, BDM provide the first stage estimates from the regression of the wage and other income on the group indicators. Thus, one knows the means of these variables for each group. It is then straightforward to calculate the cross-group variance in the wage and other income. BDM do not give the number of observations in each individual group so I calculate the cross-group variance under the assumption that there are the same number of observations in each group. This assumption is counterfactual but given that they exclude groups with fewer than 50 observations and there are on average only 142 observations per group, the variance of the group means is probably close to the weighted variance. While BDM do not give information on the average within-group variance of log wages and other income, they report the within-year variance for these variables. To calculate the λs, I assume that the average within-group variance of the log wage and other income is 80% of the average within-year variance. This assumption approximates what one finds in the CPS. Further, I assume zero correlation between the log wage and other income. Given these assumptions, I use the information in tables 9, 10, and 15 of BDM to estimate that λ_w equals 0.63 and λ_y equals 0.30.[18]

BDM also report p values for rank tests for both the log wage and other income individually (i.e. testing rank one against rank zero in each reduced form). The p value for the wage is 0%, and the p value for other income is 0.038%. For comparison, the median p value for these rank tests for the CPS female labor supply application are 0% and 0.05% for the wage and other income respectively in the groups of 500. This similarity of the rank test statistics is consistent with the finding that the BDM λ statistics are fairly similar to those in the groups of 500 in table 8. Given the relatively small group sizes used by BDM, it is not surprising that the λ

[17] BDM could not provide me with their data extract for confidentiality reasons, so I estimate λ using the wealth of information provided in their paper.

[18] The estimated λ values are somewhat sensitive to the assumption about within-group variances. For example, if the within group variance is assumed to be only 50% of the within-year variance, this implies that λ_w equals 0.77 and λ_y equals 0.56. On the other hand, variation in the assumed covariance of wages and other income has very little effect on the estimated λs.

values are quite a lot smaller than one in their application.

6 Conclusions

In this paper, I have shown using real world data and applications that having 100 or 200 observations per group is not necessarily sufficient for biases in the EWALD estimator to be small. In the male labor supply application, biases are not terribly large once there are 500 or 1000 observations per group. In contrast, in the female labor supply application considered, many thousands of observations per group are necessary for the EWALD estimator to have small biases. Indeed, in the female labor supply application, the Monte Carlos suggest that there are still biases of the order of 10% in the income elasticity when there are 10,000 observations per group. These results reflect the fact that the variance of wages over time within cohorts is much lower than the variance of wages across cohorts in the cross-section. Thus, while reasonably small numbers of observations may be sufficient for precisely estimating group means, the presence of cohort and year fixed effects in synthetic cohort models increases enormously the likelihood of serious small sample biases in EWALD and the number of observations required to eliminate biases.[19]

The paper also suggests λ as a diagnostic for small sample bias in the EWALD estimator and shows how it is related to the first stage F statistic from the instrumental variables literature and the reliability ratio from the classical measurement error model. A value of λ that is much less than one suggests likely attenuation biases in the EWALD estimator. On the other hand, there are many cases where the first-stage F statistic is quite large and bias is sizeable. The bias diagnostic proves particularly useful when group sizes are very small as in this case bias-corrected estimators such as UEVE can be extremely variable. In this situation, the λ indicator remains much less than one in every single draw of the data.

In some instances when group sizes are small, identification may be sufficiently poor that no estimator provides reliable estimates. In these cases, reseachers should consider obtaining more or better data or estimating

[19] For example, when the year indicators are omitted from the first Monte Carlo, the median bias of EWALD falls from -0.17 to -0.04 in the groups of 100 (compared to the true value of 0.351).

a different model. One approach is to reduce the number of groups, so the number of observations per group increases (see Verbeek and Nijman 1992 for an analysis of the choice of number of groups). As an alternative to grouping, Moffitt (1993) proposes projecting the individual-level explanatory variables onto continuous functions of time invariant and time varying variables. Moffitt does not discuss these estimators in a small sample context but the small sample benefits of smoothing may be large in some contexts.

References

[1] Altonji, J. (1986), ’Intertemporal Substitution in Labor Supply: Evidence from Micro Data’ *Journal of Political Economy*, 94, S176-S215.

[2] Anderson T. and H. Rubin (1949), ’Estimation of the parameters of a single equation in a complete system of stochastic equations’, *Annals of Mathematical Statistics*, 20, 46-63.

[3] Anderson T. and H. Rubin (1950), ’The asymptotic properties of estimators of the parameters of a single equation in a complete system of stochastic equations’, *Annals of Mathematical Statistics*, 21, 570-582.

[4] Angrist, J.D. (1991) ”Grouped Data Estimation and Testing in Simple Labor Supply Models”, *Journal of Econometrics*, 47, 243-265.

[5] Angrist, J.D., G.W. Imbens, and A.B. Krueger (1999), ’Jackknife Instrumental Variables Estimation’, *Journal of Applied Econometrics*, 14, 57-67.

[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] Bekker Paul (1994), ’Alternative approximations to the distributions of instrumental variable estimators’, *Econometrica* 62, 657-681.

[8] Bekker Paul and Jan van der Ploeg (1999), ’Instrumental Variable Estimation Based on Grouped Data’, mimeo.

[9] Blundell, Richard, Alan Duncan, and Costas Meghir (1998) ”Estimating Labor Supply Responses using Tax Reforms”, *Econometrica*, 66, 4, 827-861.

[10] Blundell, Richard and Thomas MaCurdy (1999). “Labor Supply: a Review of Alternative Approaches.” in *Handbook of Labor Economics* Volume 3A, Orley Ashenfelter and David Card eds., Amsterdam, Elsevier Science.

[11] 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.

[12] Buchinsky, Moshe (1995), ’Estimating the Asymptotic Covariance Matrix for Quantile Regression Models: A Monte Carlo Study’, *Journal of Econometrics*, 68, 303-338.

[13] Deaton, A. (1985), ’Panel data from a time series of cross-sections’, *Journal of Econometrics*, 30, 109-126.

[14] Devereux P. (2006), ’Improved Errors-in-Variables Estimators for Grouped Data’, UCD Centre for Economic Research WP06/02.

[15] Durbin, J. (1954), ’Errors in Variables’, *Review of the International Statistical Institute,* 22, 23-32.

[16] Flores-Lagunes, Alfonso (2002), ’Estimation Methods Robust to Weak Instruments’, mimeo (November).

[17] Hausman, Jerry, and Paul Ruud (1984). ”Family Labor Supply with Taxes” *American Economic Review* 74(2): 242-248.

[18] Hyslop, Dean (2001). “Rising U.S. Earnings Inequality and Family Labor Supply: The Covariance Structure of Intrafamily Earnings.” *American Economic Review* 91(4): 755-777.

[19] Kooreman, Peter, and Arie Kapteyn (1986). ”Estimation of Rationed and Unrationed Household Labour Supply Functions Using Flexible Functional Forms.” *Economic Journal* 96(382): 398-412.

[20] MaCurdy T. (1981), ’An Empirical Model of Labor Supply in a Life-Cycle Setting’, *Journal of Political Economy*, 89, 1059-1085.

[21] Moffitt, Robert “Identification and estimation of dynamic models with a time series of repeated cross-sections” *Journal of Econometrics* 59 (1993) 99-123.

[22] Pencavel, John (1998). “The Market Work Behavior and Wages of Women: 1975-94.” *Journal of Human Resources* 33(4): 771-804.

[23] Propper, C., H. Rees, and K. Green (2001), ”The demand for private medical insurance in the UK: A cohort analysis”, *Economic Journal*, 111 (May), C180-C200.

[24] Ransom, Michael (1987). ”An Empirical Model of Discrete and Continuous Choice in Family Labor Supply.” *Review of Economics and Statistics* 69(3): 465-472.

[25] Ridder, G., and R. Moffitt (2003). ”The Econometrics of Data Combination”, Chapter for the forthcoming *Handbook of Econometrics*.

[26] Robin, J.M., and R.J. Smith (2000), ’Tests of rank’, *Econometric Theory*, 16(2), 151-75.

[27] Staiger D. and J.H. Stock (1997), ’Instrumental variables regression with weak instruments’, *Econometrica*, 65, 557-586.

[28] Verbeek, M. (1992), ”Pseudo Panel Data”, in *The Econometrics of Panel Data: Handbook of Theory and Applications*, Laszlo Matyas and Patrick Sevestre (editors), 303-315.

[29] Verbeek, M. and T. Nijman (1992), ”Can Cohort Data be Treated as Genuine Panel Data?” *Empirical Economics*, 17: 9-23.

[30] Verbeek, M. and T. Nijman (1993), ”Minimum MSE estimation of a regression model with fixed effects from a series of cross-sections” J*ournal of Econometrics* 59, 125-136.

Table 1: Male Labor Supply - Estimates of Wage Elasticity from Full Sample

	EWALD	EVE	UEVE	EVE2	LIML
Estimate	0.351	0.356	0.355	0.356	0.368
S.E.	0.018	0.018	0.018	0.018	0.018

For the Full Sample (883,610 observations), F=82.066, λ=0.988.
Also included are year and cohort fixed effects.

Table 2: Male Labor Supply (Wage Elasticity)

	10%	25%	Median Estimate	75%	90%	Trimmed Mean
Panel A: 100 Observations per group						
EWALD	0.122	0.147	0.176	0.205	0.236	0.176
EVE	0.280	0.374	0.508	0.749	1.219	0.597
UEVE	0.216	0.271	0.338	0.433	0.559	0.358
EVE2	0.263	0.342	0.447	0.637	0.916	0.509
LIML	0.238	0.290	0.354	0.438	0.521	0.365
F=1.866, λ=0.456						
Panel B: 200 Observations per group						
EWALD	0.182	0.206	0.230	0.257	0.280	0.231
EVE	0.312	0.359	0.423	0.498	0.587	0.432
UEVE	0.264	0.303	0.348	0.398	0.451	0.351
EVE2	0.296	0.342	0.400	0.465	0.541	0.406
LIML	0.279	0.317	0.359	0.408	0.459	0.364
F=2.722, λ=0.627						
Panel C: 500 Observations per group						
EWALD	0.255	0.271	0.290	0.310	0.324	0.290
EVE	0.327	0.350	0.378	0.409	0.435	0.380
UEVE	0.306	0.326	0.351	0.379	0.399	0.352
EVE2	0.321	0.343	0.370	0.400	0.424	0.372
LIML	0.321	0.342	0.368	0.395	0.420	0.368
F=5.204, λ=0.805						
Panel D: 1000 Observations per group						
EWALD	0.288	0.301	0.317	0.332	0.344	0.317
EVE	0.328	0.344	0.363	0.383	0.399	0.363
UEVE	0.318	0.332	0.351	0.370	0.384	0.351
EVE2	0.325	0.341	0.360	0.379	0.394	0.360
LIML	0.334	0.348	0.366	0.386	0.403	0.367
F=9.416, λ=0.892						
Panel E: 2000 Observations per group						
EWALD	0.312	0.321	0.333	0.345	0.353	0.333
EVE	0.334	0.344	0.357	0.370	0.380	0.357
UEVE	0.328	0.338	0.350	0.364	0.373	0.351
EVE2	0.332	0.342	0.355	0.368	0.378	0.355
LIML	0.343	0.353	0.367	0.380	0.392	0.367
F=17.732, λ=0.943						

Results from 1000 random draws (with replacement) from the Full sample.
Also included are year and cohort fixed effects.

Table 3: Monte Carlo for Male Labor Supply (Wage Elasticity)

	10%	25%	Median Bias	75%	90%	Median Abs. Err.	Trimmed Mean Bias	Trimmed MAE	90% C.I. Coverage
				Panel A: 100 Observations per group					
EWALD	-0.217	-0.193	-0.166	-0.140	-0.117	0.166	-0.166	0.166	0.017
EVE	-0.066	0.024	0.138	0.336	0.673	0.154	0.202	0.219	0.968
UEVE	-0.126	-0.071	-0.008	0.074	0.173	0.072	0.004	0.074	0.884
EVE2	-0.083	-0.007	0.087	0.236	0.463	0.115	0.128	0.153	0.955
LIML	-0.107	-0.058	-0.001	0.064	0.131	0.061	0.004	0.060	0.804
F=1.901, λ=0.489									
				Panel B: 200 Observations per group					
EWALD	-0.154	-0.135	-0.113	-0.090	-0.071	0.113	-0.112	0.112	0.069
EVE	-0.036	0.007	0.060	0.128	0.207	0.069	0.070	0.080	0.838
UEVE	-0.077	-0.044	-0.004	0.042	0.090	0.043	0.000	0.042	0.879
EVE2	-0.048	-0.009	0.039	0.101	0.169	0.057	0.048	0.064	0.866
LIML	-0.067	-0.037	0.000	0.038	0.077	0.037	0.001	0.037	0.867
F=2.798, λ=0.653									
				Panel C: 500 Observations per group					
EWALD	-0.088	-0.074	-0.057	-0.040	-0.025	0.057	-0.057	0.057	0.274
EVE	-0.024	-0.003	0.021	0.050	0.077	0.029	0.024	0.031	0.835
UEVE	-0.043	-0.023	-0.002	0.023	0.046	0.023	0.000	0.023	0.886
EVE2	-0.030	-0.009	0.014	0.042	0.068	0.027	0.017	0.028	0.858
LIML	-0.039	-0.021	0.000	0.022	0.042	0.022	0.000	0.021	0.875
F=5.500, λ=0.823									
				Panel D: 1000 Observations per group					
EWALD	-0.055	-0.044	-0.032	-0.018	-0.007	0.032	-0.031	0.031	0.521
EVE	-0.019	-0.005	0.010	0.028	0.043	0.018	0.011	0.018	0.856
UEVE	-0.028	-0.016	-0.001	0.015	0.030	0.016	0.000	0.015	0.891
EVE2	-0.021	-0.008	0.007	0.024	0.039	0.017	0.008	0.017	0.873
LIML	-0.027	-0.015	0.000	0.015	0.028	0.015	0.000	0.014	0.891
F=9.989, λ=0.903									
				Panel E: 2000 Observations per group					
EWALD	-0.034	-0.026	-0.017	-0.007	0.002	0.017	-0.016	0.017	0.678
EVE	-0.014	-0.005	0.005	0.017	0.026	0.011	0.006	0.011	0.873
UEVE	-0.019	-0.011	0.000	0.010	0.020	0.011	0.000	0.010	0.895
EVE2	-0.016	-0.007	0.004	0.015	0.025	0.011	0.004	0.011	0.884
LIML	-0.019	-0.010	0.000	0.010	0.020	0.010	0.000	0.010	0.893
F=18.991, λ=0.949									
				Panel F: 10000 Observations per group					
EWALD	-0.012	-0.008	-0.004	0.001	0.005	0.005	-0.003	0.005	0.852
EVE	-0.007	-0.004	0.001	0.006	0.010	0.005	0.001	0.004	0.891
UEVE	-0.008	-0.005	0.000	0.005	0.009	0.005	0.000	0.004	0.897
EVE2	-0.008	-0.004	0.001	0.005	0.010	0.005	0.001	0.004	0.895
LIML	-0.008	-0.004	0.000	0.004	0.009	0.004	0.000	0.004	0.897
F=90.998, λ=0.989									

True value of wage elasticity is 0.351.
Also included are cohort and year fixed effects. Results are from 2000 Monte Carlo replications.

Table 4: Female Labor Supply - Estimates from Full Sample

	EWALD	EVE	UEVE	EVE2	LIML
Wage Elasticity					
Estimate	0.249	0.372	0.333	0.360	1.848
S.E.	0.067	0.108	0.093	0.103	2.460
Income Elasticity					
Estimate	-0.276	-0.398	-0.359	-0.386	-1.826
S.E.	0.068	0.111	0.096	0.106	2.368

For the full sample (498,667 observations), Fw=33.728, Fy=18.062, λw=0.736, λy=0.717
Also included are cohort and year fixed effects.

Table 5: Female Labor Supply (Wage Elasticity)

	10%	25%	Median Estimate	75%	90%	Trimmed Mean
Panel A: 200 Observations per group						
EWALD	0.019	0.056	0.096	0.139	0.176	0.097
EVE	-0.859	-0.329	-0.066	0.188	0.703	-0.078
UEVE	-0.592	-0.079	0.141	0.444	1.088	0.185
EVE2	-1.095	-0.358	-0.025	0.331	0.987	-0.034
LIML	-2.903	-0.569	0.665	1.521	3.506	0.517
Fw=2.446, Fy=1.774, λw=0.183, λy=0.138						
Panel B: 500 Observations per group						
EWALD	0.037	0.073	0.112	0.152	0.187	0.112
EVE	-1.247	-0.327	0.239	0.670	1.562	0.192
UEVE	-0.023	0.110	0.245	0.402	0.691	0.275
EVE2	-0.870	-0.011	0.288	0.610	1.349	0.274
LIML	-2.072	0.859	1.401	2.430	5.103	1.601
Fw=4.667, Fy=2.992, λw=0.345, λy=0.295						
Panel C: 1000 Observations per group						
EWALD	0.068	0.101	0.139	0.175	0.211	0.139
EVE	0.104	0.235	0.363	0.556	0.835	0.403
UEVE	0.097	0.167	0.248	0.342	0.451	0.258
EVE2	0.113	0.213	0.324	0.468	0.673	0.348
LIML	0.934	1.217	1.670	2.624	4.551	2.104
Fw=8.355, Fy=4.964, λw=0.497, λy=0.451						
Panel D: 2000 Observations per group						
EWALD	0.106	0.138	0.173	0.208	0.236	0.173
EVE	0.166	0.225	0.311	0.388	0.455	0.308
UEVE	0.146	0.195	0.257	0.317	0.361	0.255
EVE2	0.159	0.215	0.293	0.363	0.422	0.290
LIML	1.213	1.416	1.752	2.417	3.204	1.952
Fw=15.820, Fy=8.958, λw=0.648, λy=0.615						

Results from 1000 random draws (with replacement) from the Full sample.
Also included are cohort and year fixed effects.

Table 6: Female Labor Supply (Income Elasticity)

	10%	25%	Median Estimate	75%	90%	Trimmed Mean
Panel A: 200 Observations per group						
EWALD	-0.149	-0.120	-0.086	-0.052	-0.021	-0.085
EVE	-0.770	-0.235	0.001	0.266	0.766	0.020
UEVE	-1.070	-0.466	-0.183	0.045	0.531	-0.220
EVE2	-0.938	-0.350	-0.038	0.309	0.996	-0.019
LIML	-3.426	-1.487	-0.708	0.495	2.803	-0.549
Fw=2.446, Fy=1.774, λw=0.183, λy=0.138						
Panel B: 500 Observations per group						
EWALD	-0.188	-0.153	-0.122	-0.088	-0.056	-0.121
EVE	-1.530	-0.679	-0.270	0.259	1.132	-0.229
UEVE	-0.706	-0.419	-0.267	-0.143	-0.013	-0.300
EVE2	-1.332	-0.632	-0.319	-0.041	0.870	-0.303
LIML	-4.896	-2.349	-1.371	-0.872	1.966	-1.569
Fw=4.667, Fy=2.992, λw=0.345, λy=0.295						
Panel C: 1000 Observations per group						
EWALD	-0.224	-0.188	-0.157	-0.122	-0.095	-0.157
EVE	-0.832	-0.574	-0.388	-0.257	-0.146	-0.428
UEVE	-0.464	-0.357	-0.278	-0.194	-0.132	-0.284
EVE2	-0.689	-0.488	-0.347	-0.236	-0.152	-0.374
LIML	-4.394	-2.554	-1.657	-1.232	-0.949	-2.067
Fw=8.355, Fy=4.964, λw=0.497, λy=0.451						
Panel D: 2000 Observations per group						
EWALD	-0.253	-0.227	-0.194	-0.162	-0.137	-0.195
EVE	-0.480	-0.407	-0.331	-0.256	-0.202	-0.333
UEVE	-0.386	-0.338	-0.279	-0.222	-0.179	-0.281
EVE2	-0.445	-0.383	-0.313	-0.245	-0.195	-0.315
LIML	-3.197	-2.351	-1.759	-1.413	-1.209	-1.925
Fw=15.820, Fy=8.958, λw=0.648, λy=0.615						

Results from 1000 random draws (with replacement) from the Full sample.
Also included are cohort and year fixed effects

Table 7: Monte Carlo for Female Labor Supply (Wage Elasticity)

	10%	25%	Median Bias	75%	90%	Median Abs. Err.	Trimmed Mean Bias	Trimmed MAE	90% C.I. Coverage
				Panel A: 200 Observations per group					
EWALD	-0.266	-0.237	-0.203	-0.170	-0.139	0.203	-0.203	0.203	0.010
EVE	-1.103	-0.607	-0.367	-0.148	0.365	0.477	-0.374	0.477	0.671
UEVE	-0.709	-0.306	-0.116	0.109	0.561	0.294	-0.098	0.294	0.868
EVE2	-1.214	-0.600	-0.314	0.014	0.667	0.496	-0.295	0.496	0.788
LIML	-0.433	-0.201	-0.037	0.180	0.551	0.226	-0.001	0.226	0.690
Fw=2.454, Fy=1.528, λw=0.235, λy=0.140, Prop(p<0.01)=0.151									
				Panel B: 500 Observations per group					
EWALD	-0.228	-0.204	-0.176	-0.148	-0.122	0.176	-0.176	0.176	0.008
EVE	-1.669	-0.593	0.037	0.443	1.441	0.517	-0.031	0.657	0.893
UEVE	-0.187	-0.108	-0.011	0.126	0.384	0.115	0.027	0.132	0.888
EVE2	-1.102	-0.154	0.066	0.396	1.183	0.292	0.080	0.444	0.930
LIML	-0.152	-0.082	-0.002	0.105	0.245	0.090	0.017	0.097	0.767
Fw=4.648, Fy=2.325, λw=0.391, λy=0.287, Prop(p<0.01)=0.232									
				Panel C: 1000 Observations per group					
EWALD	-0.184	-0.163	-0.140	-0.116	-0.094	0.140	-0.140	0.140	0.019
EVE	-0.056	0.026	0.139	0.321	0.685	0.156	0.201	0.215	0.970
UEVE	-0.104	-0.059	-0.001	0.070	0.163	0.064	0.010	0.066	0.897
EVE2	-0.066	-0.002	0.086	0.221	0.456	0.106	0.125	0.145	0.960
LIML	-0.089	-0.047	0.001	0.058	0.121	0.052	0.007	0.053	0.805
Fw=8.309, Fy=3.664, λw=0.531, λy=0.442, Prop(p<0.01)=0.453									
				Panel D: 2000 Observations per group					
EWALD	-0.135	-0.118	-0.099	-0.079	-0.061	0.099	-0.098	0.098	0.058
EVE	-0.028	0.010	0.059	0.118	0.188	0.064	0.067	0.074	0.837
UEVE	-0.065	-0.036	-0.001	0.038	0.081	0.037	0.003	0.037	0.879
EVE2	-0.040	-0.005	0.039	0.090	0.150	0.051	0.045	0.057	0.869
LIML	-0.056	-0.030	0.000	0.035	0.069	0.032	0.003	0.032	0.833
Fw=15.623, Fy=6.344, λw=0.676, λy=0.611, Prop(p<0.01)=0.842									
				Panel E: 4500 Observations per group					
EWALD	-0.084	-0.071	-0.056	-0.042	-0.027	0.056	-0.056	0.056	0.197
EVE	-0.021	-0.001	0.024	0.049	0.074	0.029	0.025	0.031	0.828
UEVE	-0.039	-0.021	0.000	0.021	0.042	0.021	0.001	0.021	0.884
EVE2	-0.027	-0.007	0.016	0.040	0.063	0.025	0.017	0.026	0.864
LIML	-0.035	-0.019	0.000	0.020	0.041	0.020	0.001	0.019	0.893
Fw=33.924, Fy=13.029, λw=0.817, λy=0.778, Prop(p<0.01)=0.998									
				Panel F: 10000 Observations per group					
EWALD	-0.050	-0.040	-0.029	-0.018	-0.008	0.029	-0.029	0.029	0.454
EVE	-0.016	-0.004	0.010	0.024	0.038	0.015	0.010	0.016	0.860
UEVE	-0.024	-0.013	0.000	0.013	0.026	0.013	0.000	0.013	0.892
EVE2	-0.018	-0.007	0.007	0.021	0.034	0.014	0.007	0.014	0.876
LIML	-0.023	-0.012	0.000	0.013	0.026	0.013	0.001	0.012	0.900
Fw=74.142, Fy=27.756, λw=0.906, λy=0.886, Prop(p<0.01)=1.000									

True value of Wage Elasticity is 0.331. Also included are cohort and year fixed effects.
The p statistic is from the rank test (tests rank 2 versus rank 1). Results are from 2000 Monte Carlo replications.

Table 8: Monte Carlo for Female Labor Supply (Income Elasticity)

	10%	25%	Median Bias	75%	90%	Median Abs. Err.	Trimmed Mean Bias	Trimmed MAE	90% C.I. Coverage
				Panel A: 200 Observations per group					
EWALD	0.222	0.249	0.278	0.305	0.333	0.278	0.277	0.277	0.000
EVE	-0.470	0.192	0.446	0.721	1.333	0.527	0.448	0.577	0.640
UEVE	-0.672	-0.122	0.139	0.369	0.897	0.288	0.123	0.358	0.855
EVE2	-0.791	-0.027	0.383	0.724	1.455	0.532	0.352	0.603	0.770
LIML	-0.666	-0.204	0.045	0.230	0.515	0.221	0.001	0.266	0.676
Fw=2.454, Fy=1.528, λw=0.235, λy=0.140, Prop(p<0.01)=0.151									
				Panel B: 500 Observations per group					
EWALD	0.180	0.204	0.230	0.256	0.281	0.230	0.230	0.230	0.001
EVE	-1.762	-0.557	-0.049	0.743	2.060	0.637	0.036	0.807	0.894
UEVE	-0.472	-0.150	0.018	0.129	0.220	0.136	-0.033	0.158	0.878
EVE2	-1.441	-0.490	-0.089	0.177	1.378	0.353	-0.100	0.545	0.929
LIML	-0.295	-0.120	0.003	0.094	0.176	0.104	-0.021	0.113	0.750
Fw=4.648, Fy=2.325, λw=0.391, λy=0.287, Prop(p<0.01)=0.232									
				Panel C: 1000 Observations per group					
EWALD	0.134	0.157	0.180	0.203	0.224	0.180	0.180	0.180	0.003
EVE	-0.845	-0.400	-0.171	-0.034	0.061	0.190	-0.249	0.265	0.969
UEVE	-0.197	-0.083	0.005	0.070	0.121	0.075	-0.012	0.078	0.890
EVE2	-0.554	-0.271	-0.105	0.000	0.075	0.126	-0.155	0.177	0.963
LIML	-0.144	-0.067	0.000	0.055	0.100	0.060	-0.008	0.061	0.793
Fw=8.309, Fy=3.664, λw=0.531, λy=0.442, Prop(p<0.01)=0.453									
				Panel D: 2000 Observations per group					
EWALD	0.087	0.105	0.125	0.146	0.163	0.125	0.125	0.125	0.014
EVE	-0.233	-0.144	-0.071	-0.015	0.030	0.077	-0.083	0.090	0.821
UEVE	-0.096	-0.045	0.002	0.042	0.075	0.043	-0.003	0.043	0.872
EVE2	-0.185	-0.109	-0.047	0.004	0.045	0.060	-0.055	0.069	0.861
LIML	-0.081	-0.040	0.000	0.035	0.064	0.037	-0.003	0.037	0.827
Fw=15.623, Fy=6.344, λw=0.676, λy=0.611, Prop(p<0.01)=0.842									
				Panel E: 4500 Observations per group					
EWALD	0.040	0.055	0.071	0.087	0.101	0.071	0.071	0.071	0.112
EVE	-0.090	-0.058	-0.029	-0.001	0.022	0.034	-0.031	0.037	0.811
UEVE	-0.050	-0.024	0.001	0.024	0.044	0.024	-0.001	0.024	0.877
EVE2	-0.078	-0.047	-0.019	0.006	0.029	0.029	-0.021	0.031	0.851
LIML	-0.046	-0.023	0.000	0.022	0.039	0.022	-0.001	0.022	0.891
Fw=33.924, Fy=13.029, λw=0.817, λy=0.778, Prop(p<0.01)=0.998									
				Panel F: 10000 Observations per group					
EWALD	0.013	0.024	0.036	0.048	0.059	0.036	0.036	0.036	0.358
EVE	-0.044	-0.029	-0.012	0.003	0.017	0.018	-0.013	0.018	0.844
UEVE	-0.029	-0.015	0.000	0.015	0.027	0.015	0.000	0.014	0.886
EVE2	-0.040	-0.024	-0.008	0.007	0.020	0.016	-0.009	0.017	0.870
LIML	-0.029	-0.014	0.000	0.014	0.026	0.014	-0.001	0.014	0.892
Fw=74.142, Fy=27.756, λw=0.906, λy=0.886, Prop(p<0.01)=1.000									

True Value of Income Elasticity is -0.363. Also included are cohort and year fixed effects.
The p statistic is from the rank test (tests rank 2 versus rank 1). Results are from 2000 Monte Carlo replications.