Columbia University

Department of Economics
Discussion Paper Series

Estimation of Models with Grouped and Ungrouped Data by Means of "2SLS"

Phoebus J. Dhrymes
Adriana Lleras-Muney

Discussion Paper No.: 0304-16

Department of Economics
Columbia University
New York, NY 10027

March 2004

Estimation of Models with Grouped and Ungrouped Data by Means of "2SLS"

Phoebus J. Dhrymes
Columbia University

Adriana Lleras-Muney
Princeton University

September 2001; this version March 2004*

Abstract

This paper deals with a special case of estimation with grouped data, where the dependent variable is only available for groups, whereas the endogenous regressor(s) is available at the individual level. By estimating the first stage using the available individual data, and then estimating the second stage at the aggregate level, it might be possible to gain efficiency relative to the OLS and 2SLS estimators that use only grouped data. We term this the Mixed-2SLS estimator (M2SLS). The M2SLS estimator is consistent and asymptotically normal. We also provide a test of efficiency of M2SLS relative to OLS and "2SLS" estimators.
(JEL C10, C30, C43)
Key Words: Two-stage least squares, instrumental variables, grouped data, mixed two stage least squares, test of efficiency.

1 Introduction and Review

Individual data is not always available for empirical estimation, but often grouped data can be obtained. As is well known, grouped data estimation

*Corresponding author: Adriana Lleras-Muney, 320 Wallace Hall, Princeton University, Princeton, NJ 08544 tel (609) 258-6993.

of well-specified linear models yields unbiased and consistent estimates of the parameters, see e.g. Prais and Aitchison (1954). However, it is often the case that the specified model contains one or more explanatory variables (regressors) that are correlated with the structural error term. This situation arises because the model is truly a system of simultaneous equations, because there is measurement error in one of the independent variable(s) (regressors), or because there is an omitted variable that is correlated with a regressor(s). The standard solution to this problem uses instrumental variables to obtain consistent estimates of the parameters. Instrumental variable estimation can easily be done using grouped data by the standard "2SLS" procedure common in this literature or, equivalently, estimators of the parameters of the individual model can be obtained by GLS estimation using the relevant instruments, see Angrist (1991).

This paper was motivated by, and deals with, a particular case of instrumental variables for grouped data, where the dependent variable is only available for groups, whereas the endogenous regressor(s) is available at the individual level. In general, the situation described in this paper applies to any estimation done at the aggregate level, where the first stage can potentially be estimated using disaggregated data. For example, Angrist (1990) is interested in estimating the effect of veteran status on earnings. But veteran status can be correlated with unobserved characteristics that also affect earnings: for example, unhealthy men may not be eligible to serve and may not earn as much as healthy men. So the author uses draft lottery numbers which were assigned on the basis of birth dates as instruments for veteran status. Earnings data is provided by the social security administration and due to confidentiality issues, it is only released in aggregated form. However, the first stage that predicts veteran status can be estimated using individual level data from the SIPP. Other recent papers where the data available is of this type include Pritchett and Summers(1996), Winter Ebmer and Steven (1997), Dee and Evans (1999) and Lleras-Muney (2004).

In this paper we show that when some data is available at the individual level, it may be possible to gain efficiency by estimating the first stage using the available individual data, and then estimating the second stage at the aggregate level. This estimation procedure yields a consistent and asymptotically normal estimator that we refer to as Mixed-2SLS (M2SLS). Depending on the parametric configuration of the model, the M2SLS estimator can be more or less efficient than standard 2SLS using **only** grouped data.

Previous literature on aggregation of linear models mainly explores the efficiency issues that arise when using grouped data. For example, Feser and Ronchetti (1997) and Im (1998) derive efficient estimators for grouped data. The consequences of heteroskedasticity were explored by Blackburn (1997) and Dickens (1990). Moulton (1990) discussed the problem of intra-group correlations. Shore-Sheppard (1996) looked at the implication of within-group correlation when using instrumental variables. No other paper however, has examined the special case when grouped **and** ungrouped data is available.

Finally, we should be remiss if we did not discuss the relevance of the classic paper by Wald (1940). Wald's paper, which deals with the error in variabels model, capitalizes on the simple observation that a line is determined by two points. Thus, if we group the sample in two groups and take the average of the dependent and independent variables in the two groups the slope of the line connecting these two points gives an estimate of the regression slope coefficient. This model was discussed extensively in Dhrymes (1978) and further discussion here should be superfluous beyond noting that the Wald estimator requires **extra sample** information by forming the groups in terms of the magnitude of the "true value" of the independent variable. Our paper is thus different from the Wald estimator, both in its motivation and its mechanics.

This paper is organized as follows. Section two provides the formulation for the general problem and derives the M2SLS estimator and the alternative 2SLS and OLS estimators; section three shows that the estimators under consideration are consistent and asymptotically normal; it also compares these estimators in terms of their relative efficiency; section four provides a test for the relative efficiency of the estimators and section 5 discusses a variety of issues that may arise in the empirical implementation of such estimators. Section six concludes.

2 Derivation of the M2SLS Estimator

2.1 Formulation of the Problem

Consider the model

$$y = X\beta + u, \qquad X = (X_1, x_{\cdot k}), \tag{1}$$

where y is the $n \times 1$ vector of observations on the dependent variable, X is the $n \times k$ matrix of observations on the k explanatory variables, β is a conformable vector of unknown parameters and u is the (structural) error vector whose components are asserted to be i.i.d. with mean 0 and variance $0 < \sigma_{11} < \infty$.

It is asserted that one of the variables, say x_k, is correlated with the error, while the variables in X_1 are independent of the structural error vector u.[1] It is further asserted that the observations on the correlated explanatory variable have been generated by

$$x_{\cdot k} = Z\gamma + v, \quad Z = (X_1, P) \tag{2}$$

where P is $n \times s$ and Z is the $n \times m$ ($m = k - 1 + s$) matrix of observations on the "instruments", which are assumed to be independent of v and u. By the process, inappropriately termed 2SLS, of regressing $x_{\cdot k}$ on Z and then y on $\hat{X} = (X_1, \hat{x}_{\cdot k})$, we may obtain consistent estimators of β,

$$\hat{\beta}_{i2SLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'y = \beta + (\hat{X}'\hat{X})^{-1}\hat{X}'(u + \beta_k P_z v) \tag{3}$$

where the index i indicates individual data and

$$\hat{x}_{\cdot k} = Z(Z'Z)^{-1}Z'x_{\cdot k}, \quad \hat{v} = x_{\cdot k} - \hat{x}_{\cdot k} = [I - Z(Z'Z)^{-1}Z']v = P_z v. \tag{4}$$

[1] For simplicity of exposition we present here the case where only one explanatory variable is correlated with the error term, because this is the most common case in applications. A generalization of the model is presented in Appendix C.

2.2 The M2SLS Estimator

Because in many instances individual level data are not available for y, it is desired to estimate the parameter β not as indicated above, but by means of grouped data, after the variables of Eqs. (3) and (4) have been obtained for each group. The grouping is as follows: The n observations are divided into G_n groups such that the i th group contains n_i observations and

$$\sum_{i=1}^{G_n} n_i = n. \tag{5}$$

Without loss of generality, we may rearrange the observations so that the first n_1 observations belong to group 1, the next n_2 observations belong to group 2 and so on. Grouping is effected by means of the (grouping) $G_n \times n$ matrix $H = (h_{i\cdot})$, where $h_{i\cdot}$ contains all zero elements, except for an n_i-element row vector, i.e.

$$h_{i\cdot} = (0, \frac{1}{\sqrt{n_i}} e_{i\cdot}, 0), \qquad e_{i\cdot} = (1, 1, 1, \ldots, 1) \tag{6}$$

where its nonzero components appear in the positions $n_1 + n_2 + \ldots + n_{i-1} + 1, \ldots, n_1 + n_2 + \ldots + n_{i-1} + n_i$. The grouped data are thus given by

$$Hy = (\sqrt{n_1}\bar{y}_1, \sqrt{n_2}\bar{y}_2, \ldots, \sqrt{n_{G_n}}\bar{y}_{G_n})', \;\; Hu = (\sqrt{n_1}\bar{u}_1, \sqrt{n_2}\bar{u}_2, \ldots, \sqrt{n_G}\bar{u}_{G_n})',$$

$$H\hat{X} = (\sqrt{n_1}\bar{x}_1', \sqrt{n_2}\bar{x}_2', \ldots, \sqrt{n_G}\bar{x}_{G_n}')', \;\; HZ = (\sqrt{n_1}\bar{z}_1', \sqrt{n_2}\bar{z}_2', \ldots, \sqrt{n_{G_n}}\bar{z}_G')',$$

$$Hv = (\sqrt{n_1}\bar{v}_1, \sqrt{n_2}\bar{v}_2, \ldots, \sqrt{n_{G_n}}\bar{v}_{G_n})',$$

where $\bar{u}_i$, $\bar{v}_i$, denote the (scalar) means of the corresponding variables in the i th group, and $\bar{x}_i$, $\bar{z}_i$, are k- and m-element row vectors, respectively, containing the i th group means of the x-and z-variables, respectively.

The assumptions under which we operate are

i. The matrices X and Z obey

$$\frac{1}{n}X'X \to M_{xx} > 0, \;\; \frac{1}{n}Z'Z \to M_{zz} > 0;$$

where M_{xx} and M_{zz} are positive definite matrices.

ii. $G_n \geq \max(k,m)$;[2]

iii. the random vectors $w_{s\cdot} = (u_s, v_s)$, $s = 1,2,\ldots,$ are an i.i.d. symmetric sequence with

$$Ew_{s\cdot} = 0, \quad \mathrm{Cov}(w_{s\cdot}) = \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}, \quad \sigma_{12} \neq 0,$$

and finite fourth moment.

iv. For the case of **fixed** number of groups only, $\lim_{n\to\infty} \frac{n_i}{n} = \alpha_i \in (0,1)$ such that $\alpha_1 + \alpha_2 + \ldots + \alpha_{G_n} = 1$.

v. $\lim_{n\to\infty} \bar{x}_i = \xi_i$, $\lim_{n\to\infty} \bar{z}_i = \zeta_i$, $i = 1,2,\ldots,G_n$, where the limits are to be understood as ordinary convergence if the variables are not random, or as limits in probability if they are random;

vi. $\lim_{n\to\infty}(1/n)X'H'HX = M_{\xi\xi} > 0$, $\lim_{n\to\infty}(1/n)Z'H'HZ = M_{\zeta\zeta} > 0$.

Remark 1. In many applied problems the matrix H is a primitive, i.e. it is suggested by the nature of the problem investigated. However, because of (i), in general, H cannot be such that $M_{\xi\xi}$ is singular, as for example in the case that all the vectors ξ_i are the same. This can occur, if the **rows** of HX are equal to ξ^* plus a $o(n^{1/2})$ term, which implies that the **means**, $\bar{x}_i$, (or their limits) are all the same, or are linearly dependent. In nearly all applications this eventuality may be safely ruled out, but the requirement in (vi) is included for the sake of completeness and rigor.

Another implicit requirement, which will be important at another stage of the argument (see below), is that the division into classes cannot be made on the basis of a(n) (explanatory) variable which is correlated with the structural error. This is so because when H is based on classifications of an explanatory variable independent of the structural error we can assert that

$$\frac{1}{\sqrt{n}} Hu \overset{\text{P or a.c.}}{\rightarrow} 0. \tag{7}$$

[2]This condition is implicit in assumption vi, but it is stated individually for clarity in applications.

Since the typical sub-vector of the left member[3] is $\left(\sqrt{n_i/n}\right)\bar{u}_i \sim \alpha_i \bar{u}_i$, and since the grouping criterion is not correlated with the structural error, the ith group mean $\bar{u}_i$, is simply the mean of a random (sub)sample from a population of i.i.d. random variables with mean zero; hence by Proposition 23, in Dhrymes (1989) p. 188, we have $\bar{u}_i \overset{\text{a.c.}}{\to} 0$. If, on the other hand, the grouping criterion is based on an explanatory variable, or any other basis, which is **correlated** with the structural error **we cannot** make the assertion above. We illustrate this in section 5.1. Precisely the same argument may be made when the number of groups is not fixed, but tends to infinity with the sample size.

From the preceding discussion we can write the grouped data variant of Eq. (1) as:

$$Hy = HX\beta + Hu = H\hat{X}\beta + (Hu + \beta_k HP_z v).$$

The M2SLS estimator whose properties we shall now establish is given by

$$\begin{aligned} \hat{\beta} &= (\hat{X}'H'H\hat{X})^{-1}\hat{X}'H'Hy = \beta + (\hat{X}'H'H\hat{X})^{-1}\hat{X}'H'H[u + \beta_k P_z v], \\ \hat{X} &= (X_1, Z\hat{\gamma}), \quad \hat{\gamma} = (Z'Z)^{-1}Z'x_{\cdot k}, \quad P_z = I - Z(Z'Z)^{-1}Z'. \end{aligned} \tag{8}$$

Intuitively, this estimator is derived by obtaining predicted values of X using individual data; upon grouping these predicted values, we then estimate the equation of interest by OLS, where HX has been replaced by $H\hat{X}$.

2.3 Alternative Estimators

Before we proceed to the limiting distribution and questions of relative efficiency, let us set forth the "2SLS" and the OLS estimators using **only** grouped data. These two are the alternative estimators one could use

[3]The notation $\sim$, means "behaves like" or, more formally, it is asymptotically equivalent, either in probability or with probability one, or in distribution, as the context requires.

instead of the M2SLS estimator we propose here. The OLS estimator is evidently given by

$$\hat{\beta}_{OLS} = (X'H'HX)^{-1}X'H'Hy = \beta + (X'H'HX)^{-1}X'H'Hu, \quad (9)$$

while the "2SLS" estimator is given by

$$\begin{aligned}\tilde{\beta}_{\text{``}2SLS\text{''}} &= [(\widetilde{HX})'(\widetilde{HX})]^{-1}(\widetilde{HX})'Hy \\ &= \beta + [(\widetilde{HX})'(\widetilde{HX})]^{-1}(\widetilde{HX})'[Hu + \beta_k P_{Hz} Hv]. \quad (10)\end{aligned}$$

There are three major issues to be discussed. First, what are the properties of the M2SLS estimator we outlined above? Second, what are the properties of the resulting estimator **if** x_k is regressed on Z **using grouped data**, i.e. if one follows the standard "2SLS" estimation procedure. Third, if only grouped data are used, is the "2SLS" significantly different from the OLS estimator.

3 Properties of Estimators

3.1 Consistency and Limiting Distribution

We shall address these issues in two contexts: (a) when the number of groups is fixed at G and (b) when the number of groups varies with the sample size, G_n, such that $\lim_{n\to\infty} G_n = \infty$.

3.1.1 Fixed Number of Groups

Given the discussion of the previous section, we have

Theorem 1. Under assumptions (i) through (vi), the following is true:

i. The M2SLS estimator defined in equation (8) is consistent and asymptotically normal. Its distribution is given by

$$\sqrt{n}(\hat{\beta} - \beta)_{\text{M2SLS}} \xrightarrow{\text{d}} N(0, \Psi), \quad (11)$$

where

$$\Psi = \eta M_{\xi\xi}^{-1} - (\eta - \sigma_{11}) M_{\xi\xi}^{-1} M_{\xi\zeta} M_{zz}^{-1} M_{\zeta\xi} M_{\xi\xi}^{-1},$$

$$\eta = (1, \beta_k) \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} (1, \beta_k)'. \quad (12)$$

ii. The "2SLS" and the OLS estimators defined in equations (9) and (10) are consistent and asymptotically normal. Their limiting distributions are identical and given by

$$\sqrt{n}(\hat{\beta} - \beta)_{\text{"}2SLS\text{"}} \sim \sqrt{n}(\hat{\beta} - \beta)_{OLS} \sim N(0, \Phi), \quad \Phi = \sigma_{11} M_{\xi\xi}^{-1}. \quad (13)$$

Proof: For a proof, See Appendix A.

Remark 2. This development makes clear that in the context of the problem as we have formulated it, and **using grouped data**, there is **no** reason to employ the "2SLS" estimator.[4]

Remark 3. The second part of Theorem 1 is a consequence of the fact that if we model **individual behavior** and group on the basis of a criterion independent of the structural error, then the group means of the correlated regressor converge to their systematic part. We conjecture that in most applications dealing with all grouped data the investigator assumes about the grouped data, what he would otherwise have assumed for individuals. In such a context OLS and "2SLS" would not necessarily be identical.

3.1.2 Inference with Fixed Number of Groups

In order to render the results above useful for purposes of inference we need to produce also an estimator of the covariance matrix Ψ. For most entities therein there is no problem, indeed the appropriate (consistent)

[4]This result may suggest to some that H is an instrumental matrix, even though its origin lies with the manner in which the data becomes available and does **not** reflect an action by the investigator to define an instrumental matrix.

estimators are obvious. However, this is not the case with σ_{11} and σ_{12}. Since the number of groups is **fixed**, the residual sum of squares of the structural equation, from which we could obtain an estimator of σ_{11} and σ_{12}, **cannot** yield consistent estimators of these parameters, **not because the estimating procedure is deficient in some way** but simply **because the number of groups cannot be expanded**. This is a problem that affects all empirical modeling that deals with a fixed number of groups, whether OLS or "2SLS" is the appropriate method of estimation. We can however produce **unbiased** estimators of these parameters, which are also **consistent** if groups were allowed to increase to ∞. We shall do so below.

Theorem 2. For suitable constants $K_i,\ i = 1, 2$, respectively, unbiased estimators of σ_{11}, σ_{12} are given by[5]

$$\tilde{\sigma}_{11} = \frac{1}{K_1}\hat{w}'\hat{w}, \quad \tilde{\sigma}_{12} = \frac{1}{K_2}\hat{w}'H\hat{v},$$

where

$$\begin{aligned} \hat{u}^* &= H(y - \hat{X}\hat{\beta}) = H(u + \beta_k\hat{v}) - H\hat{X}(\hat{\beta} - \beta) = P_{H\hat{X}}H(u + \beta_k\hat{v}), \\ \hat{w} &= \hat{u}^* - \beta_k P_{H\hat{X}}H\hat{v} = P_{H\hat{X}}Hu. \end{aligned} \tag{14}$$

Proof: Note that

$$E\hat{w}'\hat{w} = E\mathrm{tr}[H'P_{H\hat{X}}Huu'] = \sigma_{11}(G - k).$$

Thus, for $K_1 = G - k$, $\tilde{\sigma}_{11}$ is an unbiased estimator. Similarly,

$$E\hat{w}'H\hat{v} = E\mathrm{tr}[H'P_{H\hat{X}}HP_z vu'] = \sigma_{12}\mathrm{tr}[H'P_{H\hat{X}}HP_z].$$

Thus, with $K_2 = \mathrm{tr}[H'P_{H\hat{X}}HP_z]$, σ_{12} is an unbiased estimator. Since all entities involved in the calculations above are directly available (save for β_k which is strongly consistently estimable) the problem of estimating the covariance matrix of the limiting distribution is solved for the case of fixed number of groups. q.e.d.

[5]It should be noted that this is strictly true only if $\hat{X}$ is replaced by $\bar{X}$, otherwise it is an approximation owing to the fact that γ is estimated. But for large samples this is a very good approximation.

3.1.3 Number of Groups Increases with Sample Size

We now consider the case where the number of groups varies with sample size. To stress this aspect we shall now consistently use G_n for the number of groups and H_n for the grouping matrix. To this end, let the number of groups be given by

$$G_n = n^\beta, \quad \beta \in (0, 1),$$

and the number of observations in each group be given by

$$n_i = c_i n^{1-\beta}, c_i > 0.$$

Then

$$\sum_{i=1}^{G_n} \frac{c_i n^{1-\beta}}{n} = \sum_{i=1}^{G_n} c_i n^{-\beta}, \tag{15}$$

which will satisfy the requirement that the entire sample be used, provided

$$\sum_{i=1}^{G_n} c_i = n^\beta, \quad \text{i.e. on the average the } c_i \text{ are one,} \tag{16}$$

and this specification will replace the condition in assumption iv. In fact for most of our discussion we shall take $c_i = 1$, for all i.

Before we proceed we note that the first stage results, as well as the limiting distributions of Theorem 1, remain valid with

$$M_{\xi\xi} = \lim_{n\to\infty} \frac{\bar{X}' H_n' H_n \bar{X}}{n} = \lim_{n\to\infty} \sum_{i=1}^{n^\beta} c_i \bar{x}_{i\cdot}'.$$

and similarly for the entities $M_{\xi\zeta}$ and its transpose.

Moreover, using the results of Theorem 2, we obtain

Theorem 3. The estimators,

$$\hat{\sigma_{11}} = \frac{1}{G_n} \hat{w}'\hat{w}, \quad \hat{\sigma_{12}} = \frac{1}{G_n} \hat{w}' H_n \hat{v},$$

converge with probability 1 to σ_{11}, σ_{12}, respectively.

Proof: Consider the entities of Theorem 2,

$$\hat{w}'\hat{w}, \quad \hat{w}' H_n \hat{v}.$$

We have

$$\hat{\sigma}_{11} = \frac{1}{G_n}\hat{w}'\hat{w} \sim \frac{1}{G_n}\sum_{i=1}^{G_n} n_i \bar{u}_i^2$$

$$\hat{\sigma}_{12} = \frac{1}{G_n}\hat{w}'H_n\hat{v} \sim \frac{1}{G_n}\sum_{i=1}^{G_n} n_i \bar{u}_i \bar{v}_i. \quad (17)$$

Since $n_i = n^\beta$, both equations above have rightmost members containing a sequence of i.i.d. random variables with means σ_{11}, σ_{12}, respectively. Hence, by Proposition 23, p. 188, in Dhrymes (1989)

$$\hat{\sigma}_{11} \stackrel{\text{a.c.}}{\to} \sigma_{11}, \quad \hat{\sigma_{12}} \stackrel{\text{a.c.}}{\to} \sigma_{12}, \quad (18)$$

thus completing the derivation of the inference procedure.

3.2 Relative Efficiency

Theorem 4. The M2SLS estimator is efficient relative to OLS and “2SLS” if and only if

$$\eta - \sigma_{11} = 2\beta_k\sigma_{12} + \beta_k^2\sigma_{22} < 0, \quad (19)$$

Proof: Since the three estimators examined in the previous sections are consistent, asymptotically normal and, moreover, the OLS and “2SLS” are asymptotically equivalent, the question of relative efficiency entails only a comparison of the covariance matrices of the limiting distribution of the OLS and M2SLS. Thus, consider

$$\Psi - \Phi = (\eta - \sigma_{11})[M_{\xi\xi}^{-1} - M_{\xi\xi}^{-1}M_{\xi\zeta}M_{zz}^{-1}M_{\zeta\xi}M_{\xi\xi}^{-1}], \quad (20)$$

where $\eta - \sigma_{11} = 2\beta_k\sigma_{12} + \beta_k^2\sigma_{22}$.

We show that in Eq. (20) the matrix in square brackets is positive semi-definite, using a number of results from Dhrymes (2000), chapter 3.

The matrix in question is positive semi-definite if and only if $M_{\xi\xi} - M_{\xi\zeta}M_{zz}^{-1}M_{\zeta\xi} \geq 0$. The latter, however, is the limit (after division by n), of

$$A_n = \bar{X}'H_n'H_n\bar{X} - \bar{X}'H_n'H_nZ(Z'Z)^{-1}Z'H_n'H_n\bar{X}$$

$$= \bar{X}'H_n'[H_n(I - Z(Z'Z)^{-1}Z')H_n']H_n\bar{X}. \quad (21)$$

The matrix of Eq. (21) is positive semi-definite if the matrix in square brackets is. But $P_z = I - Z(Z'Z)^{-1}Z'$ is a symmetric idempotent matrix of dimension n and rank m (the column dimension of Z). Let J be the matrix of characteristic roots of P_z (which consists of m unities and $n-m$ zeros along its main diagonal) and Q the (orthogonal) matrix of characteristic vectors. Then, we have the representation

$$A_n = \bar{X}'H_n'H_nQJQ'H_n'H_n\bar{X}, \tag{22}$$

which is evidently positive semi-definite for all $n \geq m$. Hence, the limit is also positive semi-definite, thus concluding the proof that the matrix in square brackets of Eq. (20) is positive semi-definite. q.e.d.

Remark 4. In the discussion above, we have shown that the estimator referred to in this literature as "2SLS" is, given the standard assumptions on individuals, **asymptotically equivalent** to the OLS estimator using all group data. This result is due **precisely to the fact that we have assumptions on individuals not groups**. In different contexts this may not be the case.

Remark 5. The results obtained in this section show that there are cases in which what the literature refers to as "2SLS" is inefficient relative to M2SLS. This should not be surprising given the fact that the two procedures use slightly different information.

The intuition behind the result in this section is, roughly speaking, as follows: Using individual data in the first stage utilizes more information and as such contributes to greater efficiency. However, because of subsequent grouping, the (grouped) residuals from that stage are **not necessarily orthogonal** to the grouped variables ($H_n\hat{X}$) in the second stage, so that the error term in the second stage is, in the derivation of the limiting distribution, different from the original structural error. The variance of the structural error is σ_{11} and, in a limiting sense, we may think of η as the variance of the error term in the second stage. The result then states that if $\eta - \sigma_{11} < 0$, we have efficiency for the M2SLS estimator, while if $\eta - \sigma_{11} > 0$ we do not, "because" we have added to the variability of the equation error.

4 Test for Efficiency

4.1 Derivation of a Test Statistic

Efficiency for the M2SLS estimator requires that

$$\theta = 2\beta_k\sigma_{12} + \beta_k^2\sigma_{22} \leq 0.$$

In any application, to test the hypothesis above requires a test statistic, which we shall now derive.

Theorem 5. For the case where the number of groups depends on the number of observations

i. the limiting distributions of $\sqrt{n}(\hat{\beta}_k - \beta_k)$, $\sqrt{n}(\hat{\sigma}_{22} - \sigma_{22})$, $\sqrt{G_n}(\hat{\sigma}_{12} - \sigma_{12})$ are mutually independent;

ii. they are given, respectively, by

$$N(0, \sigma^2_{\hat{\beta}_k}), \ \ N(0, \phi_{22}), \ \ N(0, \phi_{12}), \ \phi_{22} = \mu_4 - \sigma_{22}^2, \ \phi_{12} = \sigma_{12}^2 + \sigma_{11}\sigma_{22}.$$

iii. If we put

$$\zeta_n = d_1\sqrt{n}(\hat{\beta}_k - \beta_k) + d_2\sqrt{n}(\hat{\sigma}_{22} - \sigma_{22}) + d_3\sqrt{G_n}(\hat{\sigma}_{12} - \sigma_{12}),$$

we have

$$\zeta_n \xrightarrow{\mathrm{d}} N(0, \phi_\zeta), \quad \phi_\zeta = d_1^2\sigma^2_{\hat{\beta}_k} + d_2^2\phi_{22} + d_3^2\phi_{12}.$$

iv. Put

$$\lambda = \frac{\theta}{2\theta + \beta_k^2\sigma_{22}},$$

and note that, with $\xi_n = \lambda\zeta_n$, we have

$$\xi_n \xrightarrow{\mathrm{d}} N(0, \phi_\xi), \quad \phi_\xi = \lambda^2\phi_\zeta.$$

A significant detail to note is that the limiting distribution of ξ_n is **centered** on θ.

Proof: Because θ is a **nonlinear** function of the underlying parameters, β_k, σ_{12}, σ_{22} it is not routine to produce the distribution of its estimator, or the appropriate test statistic. However, following a procedure in Dhrymes (1973), we shall be able to express $\hat{\theta} - \theta$, **asymptotically**, as a linear transformation of

$$(\hat{\beta}_k - \beta_k), \quad (\hat{\sigma}_{12} - \sigma_{12}), \quad (\hat{\sigma}_{22} - \sigma_{22}).$$

Adding and subtracting appropriate entities we find

$$2\hat{\beta}_k\hat{\sigma}_{12} + \hat{\beta}_k^2\hat{\sigma}_{22} - 2\beta_k\sigma_{12} + \beta_k^2\sigma_{22} \sim (d_1, d_2, d_3)(\hat{\beta}_k - \beta, \hat{\sigma}_{22} - \sigma_{22}, \hat{\sigma}_{12} - \sigma_{12})', \tag{23}$$

where $d_1 = 2(\sigma_{12} + \beta_k\sigma_{22})$, $d_2 = \beta_k^2$, $d_3 = 2\beta_k$.

Thus, the limiting distribution of the relevant entity is given by the distribution of

$$(d_1, d_2, d_3)\begin{bmatrix} \sqrt{n}(\hat{\beta}_k - \beta_k) \\ \sqrt{n}(\hat{\sigma}_{22} - \sigma_{22}) \\ \sqrt{G_n}(\hat{\sigma}_{12} - \sigma_{12}) \end{bmatrix} \tag{24}$$

so that in order to proceed, we must find the limiting distributions of $\hat{\sigma}_{22}$ and $\hat{\sigma}_{12}$.

From Theorem 1,

$$\sqrt{n}(\hat{\beta}_k - \beta_k) \xrightarrow{\text{d}} N(0, \sigma_{\hat{\beta}}^2).$$

From the discussion of the first stage estimation, Eq. (14) as well as Theorem 3, we have

$$\hat{\sigma}_{22} = \frac{1}{n}\hat{v}'\hat{v} \sim \frac{1}{n}v'v, \quad \hat{\sigma}_{12} = \frac{1}{G_n}\hat{w}'H_n'\hat{v} \sim \frac{1}{G_n}u'H_n'H_nv. \tag{25}$$

Since the expression from which we derive the limiting distribution of $\hat{\beta}_k$, involves the vectors u, v **linearly** (see Eq. 8), while for the other two entities these vectors enter **quadratically**, it is clear that by the symmetry assumption (and their asymptotic normality) the odd moments are all null, and thus the three entities are mutually independent, thus proving i. Consequently, we need only deal separately with the limiting distribution of the estimators $\hat{\sigma}_{22}$ and $\hat{\sigma}_{12}$. We have

$$\sqrt{n}(\hat{\sigma}_{22} - \sigma_{22}) \sim \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(v_i^2 - \sigma_{22}). \tag{26}$$

The summands are a sequence of independent identically distributed random variables with mean zero and variance $\mu_4 - \sigma_{22}^2 < \infty$; consequently, by Proposition 42, in Dhrymes (1989), p. 264

$$\sqrt{n}(\hat{\sigma}_{22} - \sigma_{22}) \xrightarrow{\mathrm{d}} N(0, \phi_{22}), \quad \phi_{22} = \mu_4 - \sigma_{22}^2, \tag{27}$$

where μ_4 is the fourth moment of v, whose existence is asserted in assumption iii. In addition,

$$\sqrt{G_n}(\hat{\sigma}_{12} - \sigma_{12}) = \frac{1}{\sqrt{G_n}} \hat{w}' H_n \hat{v} \sim \frac{1}{\sqrt{G_n}}(u' H_n' H_n v - \sigma_{12}) = \frac{1}{\sqrt{G_n}} \sum_{i=1}^{G_n} (n_i \bar{u}_i \bar{v}_i - \sigma_{12}). \tag{28}$$

Again, the summands are an i.i.d. symmetric sequence with

$$E(n_i \bar{u}_i \bar{v}_i - \sigma_{12}) = \frac{1}{n_i} \sum_{j, j' \in I_i} (u_j v_{j'}) - \sigma_{12} = 0, \tag{29}$$

and variance

$$E(n_i \bar{u}_i \bar{v}_i - \sigma_{12})^2 = \frac{1}{n_i^2} E \left(\sum_{j_i \in I_i} u_{j_i} \right)^2 \left(\sum_{j_i' \in I_i} v_{j_i'} \right) - \sigma_{12}^2 \sim \sigma_{12}^2 + \sigma_{11}\sigma_{22}, \tag{30}$$

where I_i is the index set relevant for $\bar{u}_i$ and $\bar{v}_i$. Hence by the same argument we conclude that

$$\sqrt{G_n}(\hat{\sigma}_{12} - \sigma_{12}) \xrightarrow{\mathrm{d}} N(0, \phi_{12}), \quad \phi_{12} = \sigma_{12}^2 + \sigma_{11}\sigma_{22}, \tag{31}$$

which completes the proof of ii.

The proofs of iii. and iv. follow from the proofs of i. and ii., q.e.d.

4.2 Form of the Test

A test for efficiency of the M2SLS estimator may be carried out using the test statistic

$$\hat{\tau} \;=\; \lambda[d_1 \sqrt{n} \hat{\beta}_k + d_2 \sqrt{n} \hat{\sigma}_{22} + \sqrt{G_n} \hat{\sigma}_{12}], \text{ and} \tag{32}$$

$$\hat{\sigma}_{\hat{\tau}} = \sqrt{\hat{\phi}_\xi}, \tag{33}$$

and amounts to a one sided test on the mean of a normal distribution (the limiting distribution of ξ_n). A uniformly most powerful (UMP) test exists, see Roussas (1997), p. 344, and is given at the 5% and 10% levels, respectively, by

$$\hat{\tau} \leq 1.64\hat{\sigma}_{\hat{\tau}}, \quad \hat{\tau} \leq 1.28\hat{\sigma}_{\hat{\tau}}, \tag{34}$$

Remark 6. Notice that in the process of expressing the difference in Eq. (23) as a **linear** function of

$$(\hat{\beta}_k - \beta_k), \quad (\hat{\sigma}_{12} - \sigma_{12}), \quad (\hat{\sigma}_{22} - \sigma_{22}),$$

the resulting entity is **centered** on

$$d_1\beta_k + d_2\sigma_{22} + d_3\sigma_{12} = 2\theta + \beta_k^2\sigma_{22};$$

consequently, we needed to make certain adjustments (multiplication by λ) in formulating the test.

4.3 Monte Carlo Results

In this section, a number of simulations are carried out to confirm our theoretical findings. Monte Carlo simulations are performed as follows. First we generate a matrix X containing 5 i.i.d normally distributed variables $X_{1,}X_2, ..., X_5$ with mean μ and covariance $\sum$; we also generate the five additional variables $X_6, X_7, X_8, X_9, X_{10}$, each of which follows a standard normal distribution. We construct the vector of instruments P, which contains $X_{1,}X_2, ..., X_5$ and $X_6X_7X_8$. We define X_{11}, the endogenous regressor, and y the dependent variable as follows:

$$X_{11} = (X, P)\gamma + v, \ v = \theta_1 X_9 + \theta_2 X_{10}$$

$$y = X^*\beta + \varepsilon_n, \ \varepsilon = \theta_3 X_9$$

where θ_1 and θ_2 and θ_3 are constants, β is a 6 by 1 vector of constants, γ is an 8 by 1 vector of constants, and X_n^* here includes $X_{1,}X_2, ..., X_5$ AND X_{11}. We therefore have 6 explanatory variables in the structural equation, and 8 instruments (5 exogenous variables and 3 excluded instruments). Our claim is that the relative efficiency of the estimators depends on the value of $-(2\beta_k\sigma_{12} + \beta_k^2\sigma_{22})$. In the simulations we can easily manipulate this expression by noting that $\sigma_{12} = \theta_1\theta_3$ and $\sigma_{22} = [\theta_1^2 + \theta_2^2]$.

Finally the matrix H is defined as a $G \times n$ matrix and can be expressed as $H = (h_{i\cdot})$, where $h_{i\cdot}$ contains all zero elements, except for an n_i element row vector, as in equation 6 in part 3. G represents the number of groups, n is the number of observations and n_i is the number of observation in group i. For the simulations we always create groupings of equal size. The estimators are then calculated using the formulas presented in the previous sections. We repeat the procedure above 1000 times.

Tables 1.1 and 1.2, in Appendix 1, report on the empirical (sampling) distribution of the M2SLS and "2SLS" estimators. Precisely, from each replication we obtain one estimate of the parameter vector β. We may think of that as one observation from the finite distribution of the two estimators, respectively. By taking their mean and standard deviation we give some information about the first two moments of the finite sample distributions. The last column gives the characteristic roots of the difference of the two empirically obtained covariance matrices. The fact that all roots are non-negative confirms the result that one of the estimators is efficient relative to the other, depending on the parametric configuration $2\beta_k\sigma_{12} + \beta_k^2\sigma_{22}$, as obtained by asymptotic theory in the discussion(s) of the previous sections. We repeated these simulations using 2,000 replications, or alternatively with a sample size of 500 (a rather small sample by the standards of the literature. The results (available upon request) are qualitatively identical to those presented here.

5 Issues arising in Empirical applications

In this section we raise and answer a number of questions of relevance in the empirical implementation of the estimator(s) discussed in this paper. First we look at the characteristics of grouping matrix H. In the first section we give a more detailed proof of the restrictions that H must satisfy for the results on consistency to hold, which we hinted at in the previous sections. Then we show that if one has a choice on how to group the data, finer groupings always increase efficiency for either estimator (although it does not affect their relative efficiency). Then we go on to address two circumstances that are often encountered in empirical applications. It is common for researchers to use "instruments" that are already defined at the aggregate level. We answer the question of whether it is still worthwhile using individual level data in the first stage even in this circumstance. Finally, we point out that the estimation procedure we labeled M2SLS can be used when matching data from different sources, as is suggested by Angrist and Krueger (1992).

5.1 Choice of the grouping matrix H

As we noted in Remark 1, in most applied problems the matrix H is a primitive, i.e. it is suggested by the nature of the problem investigated. For example, data is sometimes only available at the aggregate level due to confidentiality concerns. In the discussion above we have asserted that, on the assumption

$$n, \quad n_i \longrightarrow \infty \quad \text{such that} \quad \frac{n_i}{n} \longrightarrow \alpha_i > 0,$$

$$\frac{1}{\sqrt{n}} Hu \xrightarrow{\text{P}} 0 \tag{35}$$

where u is a vector of i.i.d. random variables with mean zero and variance $\sigma_{11} > 0$. Can the grouping matrix, otherwise, be chosen arbitrarily? The answer is generally yes, provided the grouping **is not chosen on the basis of a variable that is correlated with u.** So an obvious implication is that the grouping cannot be done on the basis of the endogenous variable x_{k}, nor can it be done on the basis of the outcome of interest y. For example in the paper by Angrist (1990) that, it would be inappropriate to use income levels as a basis for grouping. Although this is generally acknowledged in the oral tradition of this literature (see Feige and Watts, 1972), no rigorous derivation of this result is available. We provide a suitable argument to that effect. Let $u_{\cdot i} = (u_{i1}, u_{i2})'$, $i = 1, 2, \ldots, n$ be a sequence of independent identically distributed random vectors with

$$Eu_{\cdot i} = \mu, \qquad \mathrm{Cov}(u_{\cdot i}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} > 0. \tag{36}$$

If we group on the basis of u_2, it means that observations in group i have the property $u_{s2} \in (k_{1i}, k_{2i}]$, for some constants k_{1i}, k_{2i} and $s = n^*_{i-1} + 1, n^*_{i-1} + 2, \ldots, n^*_i$, where $n^*_i = \sum_{j=1}^{i} n_j$; moreover, this holds for all i. To answer the question posed we need to determine the conditional mean of u_1 **given that** $u_2 \in (k_1, k_2]$. Although the argument may be made for an arbitrary distribution, the exposition can be considerably simplified if we assume normality, in which case we have a readily available expression for the conditional distribution. Thus, consider

$$I(k_1, k_2) = \frac{1}{F(k_1, k_2)} \int_{k_1}^{k_2} f_2(u_2) \left(\int_{-\infty}^{\infty} u_1 f(u_1|u_2) du_1 \right) du_2,$$

$$I(k_1, k_2) = E(u_1 | k_1 < u_2 < k_2), \quad F(k_1, k_2) = F_2(k_2) - F_2(k_1) \tag{37}$$

where $f(u_1|u_2)$ is the conditional density of u_1 given u_2, and F_2 is the marginal cdf of u_2. Carrying out the remaining integration we have to evaluate

$$\begin{aligned} I(k_1, k_2) &= \frac{1}{F(k_1, k_2)} \frac{1}{\sqrt{2\pi\sigma_{22}}} \int_{k_1}^{k_2} \left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}} (u_2 - \mu_2) \right) e^{-\frac{1}{2\sigma_{22}}(u_2 - \mu_2)^2} du_2 \\ &= \mu_1 + \frac{\sigma_{12}}{\sigma_{22}} \frac{[f_2(k_1) - f_2(k_2)]}{F(k_1, k_2)}, \end{aligned}$$

where f_2 is the density of a normal variable with mean μ_2 and variance σ_{22}. It is, thus, quite evident that unless $\sigma_{12} = 0$, the rightmost member of the equation above cannot possibly be μ_1 for all groups. Therefore the group means of u_1 will not converge to μ_1 if σ_{12} is different than zero.

In Table 2.1, Appendix 2, we verify empirically the results given in this section. The two tables refer to two sets of simulations as follows: In the first table we obtain 1,000 samples (replications) of 10,000 observations each, on the bivariate vector (x_1, x_2), such that their mean is **zero**, and the covariance between them (the parameter σ_{12}) is about .89. The observations in each replication are first ranked on the basis of the **magnitude of** x_2, and divided into 10 groups each containing 1,000 observations. Then we compute the group means for the two variables. The results speak for themselves; even with such great number of observations, the group means of x_1 are "significantly" different from zero for **all** groups.

In second table, the covariance between the two variables is much smaller, $\sigma_{12} \approx .1$. While many of the group means for x_1 are still significantly different from zero, **some are not**. This implies that the inconsistency entailed by grouping based on an "endogenous" variable tends to be less significant the lower the correlation between this variable and the structural error term.

5.2 Is finer or coarser grouping more efficient?

In this section we answer the question: if the problem and the data permit multiple groupings, i.e. if we can define groups equally well so as to contain more or fewer of the "individual" observations, does it make a difference, in terms of asymptotic efficiency, which is being chosen? For example if one of the exogenous variables in the equation of interest is years of education, one could group by single years of schooling or one could use two groups: less than high school and high school or more.

Without loss of generality let us pose the problem as one in which we consolidate two adjacent groups to form a new, larger group. Thus, suppose the initial grouping matrix is H as defined by the discussion surrounding Eq. (5), while after consolidation it is given by

$$H_2 = DH, \quad D = \begin{bmatrix} d_1 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & d_g \end{bmatrix},$$

$$d_{2s+1\cdot} = \left(\sqrt{\frac{n_{2s+1}}{n_{2s+1} + n_{2(s+1)}}}, \sqrt{\frac{n_{2(s+1)}}{n_{2s+1} + n_{2(s+1)}}} \right), \quad g = \frac{G}{2}, \quad (38)$$

for $s = 0, 1, \ldots, g-1$, on the assumption that G is even. The limiting distribution of the estimator in the two cases is given, respectively, by

$$\sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{\mathrm{d}} N(0, \Psi_1), \tag{39}$$

$$\sqrt{n}(\tilde{\gamma} - \gamma) \xrightarrow{\mathrm{d}} N(0, \Psi_2)$$

$$\Psi_1 = \sigma_{22} \operatorname*{plim}_{n\to\infty} \left(\frac{W_1' H' H W_1}{n} \right)^{-1}, \quad \Psi_2 = \sigma_{22} \operatorname*{plim}_{n\to\infty} \left(\frac{W_1' H_2' H_2 W_1}{n} \right)^{-1}.$$

To show that the estimator using finer groups is efficient we need to show that $\Psi_2 - \Psi_1 \geq 0$. Using the results in Dhrymes (2000), chapter 3, it is sufficient to show that

$$W_1' H' H W_1 - W_1' H_2' H_2 W_1 \geq 0, \text{ or alternatively that } H'H - H_2' H_2 \geq 0.$$

The last matrix difference is block diagonal and its sth diagonal block, $s = 0, 1, \ldots, G-1$, is given by

$$A_s = \begin{bmatrix} c\zeta e_{2s+1}' e_{2s+1} & -c e_{2s+1}' e_{2(s+1)} \\ -c e_{2(s+1)}' e_{2s+1} & \frac{c}{\zeta} e_{2(s+1)}' e_{2s+1} \end{bmatrix}, \tag{40}$$

where

$$c = \frac{1}{n_{2s+1} + n_{2(s+1)}}, \quad \zeta = \frac{n_{2(s+1)}}{n_{2s+1}}.$$

Thus, for every s, we can write the matrix A_s above as

$$A_s = c\zeta \left(e_{2s+1\cdot}, \quad -\frac{1}{\zeta} e_{2(s+1)\cdot} \right)' \left(e_{2s+1\cdot}, \quad -\frac{1}{\zeta} e_{2(s+1)\cdot} \right) \geq 0, \tag{41}$$

which is evidently positive semi-definite. The limiting distribution of the estimator of β with coarse groups is normal with mean zero and covariance matrix

$$\sigma_{11} \operatorname*{plim}_{T\to\infty} \left(\frac{1}{n} \bar{X}' H_2' H_2 \bar{X} \right)^{-1}.$$

The corresponding entity with finer groups is given by

$$\sigma_{11} \operatorname*{plim}_{T\to\infty} \left(\frac{1}{n} \bar{X}' H' H \bar{X} \right)^{-1}.$$

The matrix difference

$$J = \sigma_{11} \bar{X}' [H'H - H_2' H_2] \bar{X}$$

is positive semi-definite if the matrix in square brackets is, which was shown this to be so in an earlier discussion. Thus, finer groups always yield more efficient estimators of the structural parameters of interest when only grouped data is used.[6] The same is true for the mixed estimator, but the demonstration of this is too complex to discuss here.

Table 2.2 (Appendix 2) illustrates and confirms, using Monte Carlo simulations, the results obtained by asymptotic theory. We see in particular that with samples of 10,000 observations, increasing the number of groups from 100 to 200 results in increase in precision (lower MSE) for both estimators as the theory predicts although we note that the gain is small.

5.3 Instruments available only at the aggregate level

In this part we analyze the following problem: in the first stage we need to estimate the relationship

$$x_{\cdot k} = Z\gamma + v.$$

Let $Z = (X_1, P^*)$, where P^* is a matrix containing only **exogenous** variables. The problem is that P^* is not available. What we do have is P, which refers to all the exogenous variables **at the aggregate (group)** level. This is the case for example in the paper of Dee and Evans (1999) that uses state level drinking age policies as instruments: even though the instruments are at the aggregate level (in their case the state level), the first stage can estimated at the individual level and individual covariates can be included. Do we gain efficiency by "blowing up" such variables to the individual level, and if so how should this be done? Since

$$Hx_{\cdot k} = (HX_1, P)\gamma + Hv$$

[6] Prais and Aitchison (1954), on entirely intuitive grounds, argue that efficiency increases when observations are grouped as to maximize the between group variance. This is a special case of the result proved here which shows that *ipso facto* finer grouping is more efficient than coarser grouping, without provisos. Feige and Watts (1972), working with bank data from the Federal Reserve System noted that in their application coarser aggregation results in a significant loss of efficiency.

is the correct representation of the model in aggregate form, we must define the variables in P at the individual level as $H_1 P = P^*$, so that $HP^* = P$. This implies that we should take

$$H_1 = H' \quad \text{because then} \quad HH_1 = HH' = I_G. \tag{42}$$

In such cases, we take the individual data based model to be

$$x_{\cdot k} = W_1 H_1^* \gamma + v, \quad H_1^* = (I_n, H_1), \quad W_1 = \begin{bmatrix} X_1 & 0 \\ 0 & P \end{bmatrix} \tag{43}$$

By the arguments given earlier, the OLS estimator, $\hat{\gamma}$, for the model in Eq. (43) has the limiting distribution

$$\sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{\mathrm{d}} N(0, \Psi_1), \quad \Psi_1 = \sigma_{22} \operatorname*{plim}_{n\to\infty} \left(\frac{W_1' H_1^{*\prime} H_1^* W_1}{n} \right)^{-1}. \tag{44}$$

If we use the aggregate version of the model, viz.

$$Hx_{\cdot k} = (HX_1, P)\gamma + Hv = W_1 H^* \gamma + Hv, \quad H^* = (H, I_G), \tag{45}$$

the OLS estimator from this model, $\tilde{\gamma}$, has the limiting distribution

$$\sqrt{n}(\tilde{\gamma} - \gamma) \xrightarrow{\mathrm{d}} N(0, \Psi_2), \quad \Psi_2 = \sigma_{22} \operatorname*{plim}_{n\to\infty} \left(\frac{W_1' H^{*\prime} H^* W}{n} \right)^{-1}.$$

To determine whether the OLS estimator from the individual model is efficient relative to the one from the aggregate model it is sufficient to establish that

$$J = W_1' H_1^{*\prime} H_1^* W_1 - W_1' H^{*\prime} H^* W_1 \geq 0,$$

see Dhrymes (2000), pp. 89. But

$$J = W_1' \begin{bmatrix} I_n - H'H & 0 \\ 0 & 0 \end{bmatrix} W_1,$$

which is positive semi-definite if

$$I_n - H'H \geq 0. \tag{46}$$

It is easily shown by direct multiplication that

$$I_n - H'H = \operatorname{diag}\left(I_{n_1} - \frac{e_{1\cdot}' e_{1\cdot}}{n_1}, \;\; I_{n_2} - \frac{e_{2\cdot}' e_{2\cdot}}{n_2}, \ldots, I_{n_G} - \frac{e_{G\cdot}' e_{G\cdot}}{n_G} \right), \tag{47}$$

where $e_{i\cdot}$ is an n_i-element row vector of unities. Consequently, for **every** i,

$$I_{n_i} - \frac{e_{i\cdot}' e_{i\cdot}}{n_i} \geq 0, \tag{48}$$

which shows that the estimator based on **individual data** is efficient, even though the information in the matrix P is only available at the aggregate level. This is due to the presence of actual individual information as contained in the matrix X_1. As pointed out earlier efficiency (i.e. a smaller limiting covariance matrix) in the first stage implies efficiency in the second stage for both the "2SLS" and M2SLS estimators.

Evidently in the absence of individual level information, beyond $x_{\cdot k}$, "individual"-based estimators will be identical to aggregate-based estimators! A claim to that effect (without demonstration) is also noted in Prais and Aitchison (1954). We give a formal demonstration in the remark below.

The model in question is

$$y = x_{\cdot k}\beta_k + u, \quad x_{\cdot k} = H_1 P\gamma + v, \tag{49}$$

where $x_{\cdot k}$ is available in individual data form, while y is only available in group form, i.e. we only have the observations Hy. If we follow the procedures just mentioned the OLS estimator of γ is given by

$$\hat{\gamma} = (P'H_1'H_1P)^{-1}P'H_1'x_{\cdot k}. \tag{50}$$

Noting that $H_1 = H'$, we see that $H_1'H_1 = HH' = I_G$, so that

$$\hat{\gamma} = (P'P)^{-1}P'Hx_{\cdot k}, \tag{51}$$

which is **precisely the estimator that would have been obtained** had we implemented the first stage using grouped data.

5.4 Data available from different sources

In some recent empirical applications, it has been noted that it is sometimes not possible to find one data set that contains all the variables of interest. For example, Dee and Evans use drinking laws at the state level to produce instrumental variables estimates of the impact of education on drinking behavior. The authors combine data from Monitoring the Future which contains information drinking behavior and drinking laws at the state level, and data from the census PUMS, which contains data on educational attainment and state drinking laws. In case all requisite data are not available from the same source, can we combine data from different sources to estimate the parameters of the problem using the mixed 2SLS estimator? The answer is yes, provided these diverse sources pertain to the same universe, i.e. the data generating function for all relevant sources pertains to the same model. To see why this is so, revert to the equation defining the Mixed 2SLS estimator, i.e.

$$\hat{\beta} = (\hat{X}'H'H\hat{X}')^{-1}\hat{X}'H'Hy$$

Thus, for example, if the constituent data in matrix $\hat{X}$ are available from one source, and Hy **only** is available from another source, combining these two sources enables us to obtain the mixed 2SLS, as we did earlier, provided the data in Hy refer to the same universe, or data generating function. Indeed, since the properties of estimators depend on the limits of data moment matrices, it matters little in principle, whether all moments come from the same sample, or from different samples, provided the constituent data in the (sample) moments refer to the same process. Angrist and Krueger (1992) refer to this estimator as a **two sample IV estimator**. Since, asymptotically, it is precisely the same estimator we would have gotten were it possible to use only one sample, a separate name for such estimators is not required. In a sense, we are not really producing a different estimator, we are merely obtaining the moments required by this (same) estimator from two different sources.

Note that one can think of the mixed 2SLS estimator as a two-sample IV estimator where the first stage uses individual level data and the second stage uses aggregate data. One of the contributiona of this paper is to have shown that even when all the data are available from the same sample at the aggregate level and some are available also at the individual level, one might gain efficiency by utilizing the more dissagregated data in the first stage; the same will hold true even if the data in the first stage come from a different sample.

6 Conclusions

This paper has derived the properties of an IV estimator that can be obtained when the dependent variable is only available for groups, whereas the endogenous regressor(s) and other exogenous variables are available at the individual level. In this situation it might be possible to gain efficiency by estimating the first stage using the available individual data, and then estimating the second stage at the aggregate level. This estimation procedure yields a consistent and asymptotically normal estimator that we refer to as M2SLS. Depending on the parametric configuration of the model, the M2SLS estimator can be more or less efficient than standard "2SLS", which uses only aggregate data. In fact, given the standard assumptions on an individual based model, the "2SLS" using only aggregate (group) data is asymptotically equivalent to the OLS estimator (using only grouped data). Simulation results confirm our theoretical findings.

Acknowledgements: We wish to acknowledge useful comments from the participants of the econometrics lunch seminar at Princeton. We are especially grateful to Bo Honoré for his suggestions, and to Debopam Bhattacharya for excellent research assistance. We are also indebted to two anonymous referees for their suggestions.

REFERENCES

Angrist, J. D. (1991), "Grouped Data Estimation and Testing in Simple Labor Supply Models," *Journal of Econometrics*, vol. 47, pp. 243-66

Angrist, J. D. (1990),"Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," *American Economic Review*, vol. 80, Issue 3, pp. 313-36

Angrist, J. D. and A. B. Krueger (1992), "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples," *Journal of the American Statistical Association*, vol. 87, pp. 328-36.

Blackburn, M. (1997), "Misspecified skedastic functions in grouped-data models," *Economic Letters* vol. 55, pp. 1-8

Cramer, J.S. (1964), "Efficient Grouping, Regression and Correlation in Engel Curve Analysis," *Journal of the American Statistical Association*, vol. 59, pp. 233-250.

Dee, T. S. and W. N. Evans (1997), "Teen Drinking and Education Attainment: Evidence From Two-Sample Instrumental Variables (TSIV) Estimates," National Bureau of Economic Research, Working Paper #6082.

Dhrymes, P. J. (1970) *Econometrics: Statistical Foundations and Applications*, New York, Harper and Row.

Dhrymes, P. J. (1973), "Restricted and Unrestricted Reduced Forms: Asymptotic Distribution and Relative Efficiency," *Econometrica*, vol. 41, pp. 119-134.

Dhrymes, P.J. (1978), *Introductory Econometrics*, New York: Springer Verlag.

Dhrymes, P.J. (1989), *Topics in Advanced Econometrics: vol. I, Probability Foundations*, New York, Springer-Verlag.

Dhrymes, P.J. (1994), *Topics in Advanced Econometrics: vol. II Linear and Nonlinear Simultaneous Equations*, New York, Springer-Verlag.

Dhrymes, P. J. (2000), *Mathematics for Econometrics*, third edition, New York, Springer-Verlag.

Dickens, W. T. (1990), "Error Component in grouped data: Is it ever Worth Weighting?," *Review of Economics and Statistics*, vol. 72, pp. 328-333.

Im, K. S. (1998), "Efficient Estimation with Grouped Data", *Economics Letters*, vol. 59, pp. 169-74.

Lleras-Muney, A. (2004),"The effect of Education on Adult Mortality in the US," *Review of Economic Studies*, forthcoming.

Moulton, B. R. (1990), "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables in Micro Units," The Review of Economic and Statistics, Volume 72, Issue 2, pp. 334-338

Prais S.J. and J. Aitchison (1954), "The Grouping of Observations in Regression Analysis," *Journal of the International Statistical Institute*, No. 22.

Pritchett, L. and L. H. Summers (1996), "Wealthier is Healthier," *Journal of Human Resources*, 31(4) pp. 841-68.

Roussas, G. G. (1997), *A Course in Mathematical Statistics*, second edition, San Diego, Academic Press.

Shore-Sheppard, L. D. (1996), "The Precision of Instrumental Variables Estimates with Grouped Data," IRS Working Paper No. 374, Princeton University.

Victoria-Feser, M. and E. Rochetti (1997), "Robust Estimation for Grouped Data," *Journal of the American Statistical Association*, vol. 92 pp. 333-40

Wald, A. (1940), "The fitting of straight lines if both variables are subject to error," *Annals of Mathematical Statistics*, vol. 11, pp. 284-300.

Winter Ebmer, R. and R. Steven (1999), "Identifying the Effect of Unemployment on Crime," Center for Economic Policy Research Discussion Paper 2129.

Appendix A: Proof of Theorem 1

Recall that the M2SLS is given by:

$$\hat{\beta} = (\hat{X}'H'H\hat{X})^{-1}\hat{X}'H'Hy = \beta + (\hat{X}'H'H\hat{X})^{-1}\hat{X}'H'H[u + \beta_k P_z v].$$

For consistency we need to show that 1) $(\hat{X}'H'H\hat{X}/n)^{-1}$ converges to a positive definite matrix and, 2) $\frac{1}{n}\hat{X}'H'H[u + \beta_k P_z v]$ converges to 0.

That these two conditions are satisfied follows almost immediately from assumptions (i) through (vi) in section 3 and the content of Remark 1, **provided** that grouping was not done on a basis that is correlated with the structural error. Notice that

$$\frac{1}{n}\hat{X}'H'H\hat{X} = \sum_{i=1}^{G}\left(\frac{n_i}{n}\right)\bar{x}_i'\bar{x}_i \rightarrow \sum_{i=1}^{G}\alpha_i\xi_i'\xi_i = M_{\xi\xi} > 0 \qquad (52)$$

and moreover, because of Eq. (2), the last diagonal element of M_{xx} is given by

$$m_{(xx),kk} = \gamma' M_{zz}\gamma + \sigma_{22}$$

To be more explicit about requirement 2, the consistency of the estimator (in the sense of, at least, convergence in probability) will be established if we prove that

$$\frac{1}{n}\hat{X}'H'H[u + \beta_k P_z v] \xrightarrow{\mathrm{P}} 0. \qquad (53)$$

To show this we first note that

$$\frac{1}{n}\hat{X}'H'H[u + \beta_k v] = \sum_{j=1}^{G}\left(\frac{n_j}{n}\right)\bar{x}_j'[\bar{u}_j + \beta_k\bar{v}_j] \xrightarrow{\mathrm{P}} 0$$

due to the fact that $w_{s\cdot}$ is a sequence of i.i.d. random vectors with mean zero; in fact,

$$\bar{u}_j + \beta_k\bar{v}_j \xrightarrow{\mathrm{a.c.}} 0$$

by Kolmogorov's strong law of large numbers, see Dhrymes (1989), p. 188, and Remark 1. Next consider

$$\frac{1}{n}\hat{X}'H'HZ\left(\frac{1}{n}Z'Z\right)^{-1}\frac{1}{n}Z'v.$$

The last term converges to zero, at least in probability; thus the proof of consistency will be complete if we can show that

$$\frac{1}{n}\hat{X}'H'HZ \rightarrow \sum_{i,j=1}^{G} \alpha_j \xi_j' \zeta_j = M_{\xi\zeta},$$

and the last matrix is well defined, i.e. it has finite elements. But this is evident by the Cauchy inequality and the fact that, due to the assumptions made,

$$\frac{1}{n}Z'H'HZ \rightarrow \sum_{j=1}^{G} \alpha_j \zeta_j' \zeta_j = M_{\zeta\zeta} > 0,$$

and it, as well as $M_{\xi\xi}$ have finite elements.

Alternative Estimators

That the two alternative estimators are **consistent** may be established as follows. In the case of the OLS estimator we have

$$\frac{1}{n}X'H'HX \stackrel{\mathrm{P}}{\rightarrow} M_{\xi\xi}, \quad \frac{1}{n}X'H'Hu \stackrel{\mathrm{a.c.}}{\rightarrow} 0,$$

which shows consistency. The argument for the consistency of the "2SLS" estimator is essentially the same as that for the M2SLS. This is so because the "2SLS" estimator is given by

$$\tilde{\beta} = \beta + [(\widetilde{HX})'(\widetilde{HX})]^{-1}(\widetilde{HX})'[Hu + \beta_k P_{Hz} Hv], \tag{54}$$

and $\widetilde{HX} = (HX_1, \widetilde{Hx}_{\cdot k})$, where

$$\widetilde{Hx}_{\cdot k} = (Z'H'HZ)^{-1} Z'H'Hx_{\cdot k}. \tag{55}$$

To see this more clearly, observe that both procedures go through the intermediate step of estimating the vector γ, one using ungrouped data, the other using grouped data. In either case the resulting estimator (of γ) is consistent.

Limiting Distributions

Since there is a great deal of similarity in the arguments establishing the limiting distribution of all three estimators we shall deal with them simultaneously. Thus,

$$
\begin{aligned}
\sqrt{n}(\hat{\beta}-\beta)_{OLS} &= \left(\frac{X'H'HX}{n}\right)^{-1}\frac{1}{\sqrt{n}}X'H'Hu;\\
\sqrt{n}(\hat{\beta}-\beta)_{\text{“}2SLS\text{”}} &= \left(\frac{(\widetilde{HX})'(\widetilde{HX})}{n}\right)^{-1}\frac{1}{\sqrt{n}}(\widetilde{HX})'Hu;\\
\sqrt{n}(\hat{\beta}-\beta)_{\text{2SLS}} &= \left(\frac{\hat{X}'H'H\hat{X}}{n}\right)^{-1}\frac{1}{\sqrt{n}}\hat{X}'H'H(u+\beta_k Pzv);
\end{aligned}
$$

To simplify the discussion of limiting distribution and relative efficiency issues regarding these estimators we argue as follows: suppose the matrix of “instruments”, Z, contains X_1 as a sub-matrix, i.e. suppose

$$Z=(X_1,P),\quad P \text{ is n } \times(m-k+1) \text{ and independent of } u. \tag{56}$$

Consequently, we may write

$$\widetilde{HX}=(HX_1,(HZ)\tilde{\gamma})=HZ(I^*_{k-1},\tilde{\gamma}),\quad (\widetilde{HX})'P_{Hz}=0. \tag{57}$$

When this is the case, the “2SLS” **is asymptotically equivalent to the OLS estimator**, which implies that the widespread empirical practice of including X_1 as part of the instrumental matrix Z renders the “2SLS” estimator superfluous in the case we are considering. To elucidate the comment made above note that

$$\frac{1}{n}x'_{\cdot k}H'Hx_{\cdot k}=\sum_{i=1}^{G}\left(\frac{n_i}{n}\right)[\gamma'\bar{z}_i\bar{z}_i'\gamma+\bar{v}_i\bar{v}_i].$$

Letting $n\rightarrow\infty$ (and consequently $n_i\rightarrow\infty$, for all i), we find by the strong or weak law of large numbers invoked in connection with Eq. (7) that

$$\operatorname*{plim}_{n\rightarrow\infty}\frac{1}{n}x'_{\cdot k}H'Hx_{\cdot k}=\sum_{i=1}^{G}\alpha_i\gamma'\bar{z}_i\bar{z}_i'\gamma,$$

so that

$$\operatorname*{plim}_{n\rightarrow\infty}\frac{1}{n}[\widetilde{HX}'\widetilde{HX}-X'H'HX]=0.$$

To further facilitate discussion, introduce the notation

$$\bar{X}=(X_1,Z\gamma) \tag{58}$$

and note that

$$\frac{1}{\sqrt{n}} H(\bar{X} - X) \xrightarrow{\text{a.c.}} 0 \tag{59}$$

owing to the fact that

$$\frac{1}{\sqrt{n}} H(\bar{X} - X) = (0, -p), \quad p = \left[(n_1/n)^{1/2}\bar{v}_1, (n_2/n)^{1/2}\bar{v}_2, \cdots, (n_G/n)^{1/2}\bar{v}_G\right]', \tag{60}$$

and the group means converge to zero by Kolmogorov's strong law of large numbers. Similarly,

$$\frac{1}{\sqrt{n}} H(\hat{X} - X) = (0, p^*) \xrightarrow{\text{P}} 0, \tag{61}$$

because $p^* = \left(\frac{1}{\sqrt{n}}\right) HZ(\hat{\gamma} - \gamma) - p$. Consequently, by the consistency of the estimator of γ in both M2SLS and "2SLS" we need only deal with the relations above where X or $\hat{X}$ is replaced by $\bar{X}$. Moreover, since

$$\frac{1}{\sqrt{n}} X'H'Hu = \frac{1}{\sqrt{n}} \bar{X}'H'Hu + \frac{1}{\sqrt{n}} (0, v)'H'Hu, \quad \frac{1}{\sqrt{n}} (0, v)'H'Hu \xrightarrow{\text{P}} 0, \tag{62}$$

we need only deal with

$$\begin{aligned}
\sqrt{n}(\hat{\beta} - \beta)_{OLS} &\sim \left(\frac{(H\bar{X})'(H\bar{X})}{n}\right)^{-1} \frac{1}{\sqrt{n}} (H\bar{X})'Hu; \\
\sqrt{n}(\hat{\beta} - \beta)_{\text{``}2SLS\text{''}} &\sim \left(\frac{(\bar{X}H)'H\bar{X}}{n}\right)^{-1} \frac{1}{\sqrt{n}} (H\bar{X})'Hu; \\
\sqrt{n}(\hat{\beta} - \beta)_{\text{M2SLS}} &\sim \left(\frac{\bar{X}'H'H\bar{X}}{n}\right)^{-1} \frac{1}{\sqrt{n}} \bar{X}'H'H(u + \beta_k P_z v).
\end{aligned}$$

As noted earlier, the simplification in the "2SLS" estimator is occasioned by the fact that the matrix of "instruments", Z, contains, as a sub-matrix, X_1. An earlier version of the paper dealt with the case in which Z is not so restricted, and showed that when we restrict it as in the discussion above we obtain the result just given. Since most practitioners routinely include X_1, we chose to retain only the simplified discussion.

Although evidently the limiting distribution is slightly different for the two cases, the results of the comparison with the M2SLS estimator are, in substance, precisely the same whether one includes or does not include X_1 as a sub-matrix of Z.

For the first two estimators (OLS and "2SLS"), the limiting distribution is determined by the behavior of

$$\frac{1}{\sqrt{n}}(H\bar{X})'Hu \sim (\alpha_1^{1/2}\xi_1', \alpha_2^{1/2}\xi_2', \cdots, \alpha_G^{1/2}\xi_G)'Hu \xrightarrow{\mathrm{d}} N(0, \sigma_{11}M_{\xi\xi}), \quad (63)$$

using the central limit theorem for i.i.d. random variables, see Dhrymes (1988) p. 264.

To deal with the M2SLS estimator, given the preceding discussion, we need only deal with the term

$$\begin{aligned} \frac{1}{\sqrt{n}}\bar{X}'H'H(u + \beta_k P_z v) &= \left(\frac{1}{\sqrt{n}}\bar{X}'H'\right) H(u + \beta_k v) \\ &\quad - \beta_k \left(\frac{\bar{X}'H'HZ}{n}\right)\left(\frac{Z'Z}{n}\right)^{-1} \frac{1}{\sqrt{n}}Z'v. \quad (64) \end{aligned}$$

The equation above makes clear that we are dealing with a sequence of independent non-identically distributed random vectors obeying the Lindeberg condition, see Dhrymes (1989) p. 265; thus, we conclude that[7]

$$\sqrt{n}(\hat{\beta} - \beta)_{\mathrm{M2SLS}} \xrightarrow{\mathrm{d}} N(0, \Psi), \quad (65)$$

where

$$\begin{aligned} \Psi &= \eta M_{\xi\xi}^{-1} - (\eta - \sigma_{11})M_{\xi\xi}^{-1}M_{\xi\zeta}M_{zz}^{-1}M_{\zeta\xi}M_{\xi\xi}^{-1}, \\ \eta &= (1, \beta_k)\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}(1, \beta_k)'. \quad (66) \end{aligned}$$

[7]Detailed calcualtions available upon request.

Appendix B: Generalization of Model

Here we provide a generalization to the models examined earlier and show that no essential difference in the properties of the M2SLS estimator is found.

We alter Eq. (1) of the model so that

$$y = X\beta + u, \quad X = (X_1, X_2), \quad \beta = (\beta'_{(1)}, \beta'_{(2)})', \tag{67}$$

X_2 contains $r - k + 1$, $r < m$, and thus X is $n \times r$. Eq.(2) is also changed so that

$$X_2 = Z\Gamma + V, \quad V = (v_{i\cdot}), \quad i = 1, 2, \ldots, n \tag{68}$$

and assumption iii. is changed so that $w_{i\cdot} = (u_i, v_{i\cdot})$ is a sequence of i.i.d. vectors such that

$$\Sigma = \begin{bmatrix} \sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \tag{69}$$

Σ_{12} being the **row** vector containing the covariances between u_i and $v_{i\cdot}$, and Σ_{22} being the covariance matrix of $v_{i\cdot}$. In assumption i. we note that the sub-matrix corresponding to X_2 is altered so that it now becomes

$$M_{x_2x_2} = \Gamma' M_{zz} \Gamma + \Sigma_{22}, \tag{70}$$

but we need not change assumption vi. because the v-components of the elements of X_2 will converge to zero when we are dealing with group means.

For the sake of simplicity and maximal correspondence with the development of the main discussion we shall assume that all variables appear in all equations so that we need only apply OLS methods, thus obtaining

$$\hat{X}_2 = Z\hat{\Gamma} = Z(Z'Z)^{-1}Z'X_2, \quad X_2 - \hat{X}_2 = P_z V, \tag{71}$$

and writing the structural equation in grouped form as

$$Hy = H\hat{X}\beta + Hu + HP_z V\beta_{(2)}. \tag{72}$$

The M2SLS estimator then becomes

$$\hat{\beta}_{2SLS} = (\hat{X}'H'H\hat{X})^{-1}\hat{X}'H'Hy = \beta + (\hat{X}'H'H\hat{X})^{-1}\hat{X}'H'[Hu + HP_zV\beta_{(2)}]. \tag{73}$$

That this is a consistent estimator is quite evident, particularly so if we note that

$$V\beta_{(2)} = \sum_{j=1}^{m-k+1} v_{\cdot j}\beta_{k-1+j}, \tag{74}$$

so that the entity in the square brackets becomes

$$Hu + HP_zV\beta_{(2)} = Hu + \sum_{j=1}^{m-k+1} \beta_{k-1+j}HP_zv_{\cdot j}, \tag{75}$$

which, with the exception of the addition of finitely many terms, is **identical** with what appears in the first equation of Eq. (8). In the equation above $v_{\cdot j}$ is simply the n-element jth column vector of V.

Consequently, the addition of more variables that may, in fact, be correlated with the structural error makes the discussion of the problem more complex but **does not add** novel features to the required argument.

If one works through the application of an appropriate central limit theorem one establishes that the limiting distribution of the M2SLS estimator is given by

$$\sqrt{n}(\hat{\beta} - \beta)_{\mathrm{M2SLS}} \xrightarrow{\mathrm{d}} N(0, \Psi), \tag{76}$$

where

$$\begin{aligned} \Psi &= \eta M_{\xi\xi}^{-1} - (\eta - \sigma_{11})M_{\xi\xi}^{-1}M_{\xi\zeta}M_{zz}^{-1}M_{\zeta\xi}M_{\xi\xi}^{-1}, \\ \eta &= (1, \beta'_{(2)}) \begin{bmatrix} \sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} (1, \beta'_{(2)})'. \end{aligned} \tag{77}$$

As is quite apparent from the preceding it is only the definition of

$$\eta = \sigma_{11} + 2\Sigma_{12}\beta_{(2)} + \beta'_{(2)}\Sigma_{22}\beta_{(2)}, \tag{78}$$

that differentiates the covariance matrix Ψ above, from the one exhibited in the second equation of Eq. (52), which deals with the standard model containing only one potentially endogenous explanatory variable.

APPENDIX 1

TABLE 1.1

M2SLS efficient ($2\beta_k\sigma_{12} + \beta_k^2\sigma_{22}) = -33.75 < 0$)

Number of observations is 50,000; number of replications is 1,000.

β	Mixed ($\widehat{\beta}$)		2SLS ($\tilde{\beta}$)		λ
	Mean	sd	Mean	sd	V($\tilde{\beta}$)-V($\widehat{\beta}$)
0.03	0.033	0.043	0.033	0.046	0.0007
0.02	0.022	0.037	0.024	0.045	0.0007
1	0.999	0.031	0.999	0.035	0.0002
-0.5	-0.496	0.048	-0.495	0.054	0.00006
-0.8	-0.797	0.048	-0.797	0.053	0.0004
0.5	0.499	0.007	0.499	0.008	0.0000

TABLE 1.2

2SLS efficient ($2\beta_k\sigma_{12} + \beta_k^2\sigma_{22}) = 44.75 > 0$)

Number of observations is 50,000; number of replications is 1,000.

β	Mixed ($\widehat{\beta}$)		2SLS ($\tilde{\beta}$)		λ
	Mean	sd	Mean	sd	V($\widehat{\beta}$)-V($\tilde{\beta}$)
0.03	0.027	0.053	0.027	0.049	0.0009
0.02	0.0021	0.052	0.020	0.044	0.0008
1	0.999	0.039	0.999	0.034	0.0004
-0.5	-0.500	0.062	-0.500	0.055	0.00008
-0.8	-0.796	0.056	-0.796	0.051	0.00000
0.5	0.499	0.009	0.499	0.008	0.0006

APPENDIX 2

TABLE 2.1

Group Means of x_1. Observations grouped by the magnitude of x_2.
Number of observations is 10,000; number of replications is 1,000.

High Correlation Case, i.e. $Ex_1 = Ex_2 = 0$, $Ex_1x_2 \simeq .89$

	x_2			x_1		
Group	Means	sd	t	Means	sd	t
1	-7.849	0.084	-92.968	-1.570	0.023	-67.219
2	-4.673	0.066	-70.427	-0.935	0.019	-49.021
3	-3.031	0.060	-50.835	-0.606	0.019	-32.466
4	-1.731	0.056	-30.642	-0.346	0.019	-18.450
5	-0.565	0.057	-9.994	-0.113	0.019	-6.086
6	0.562	0.057	0.057	9.914	0.112	6.131
7	1.725	0.057	30.464	0.346	0.018	19.123
8	3.027	0.059	51.205	0.605	0.019	32.284
9	4.669	0.067	69.900	0.933	0.020	46.676
10	7.845	0.088	88.809	1.569	0.023	68.216

Low Correlation Case, i.e. $Ex_1 = Ex_2 = 0$, $Ex_1x_2 \simeq .1$

	x_2			x_1		
Group	Means	sd	t	Means	sd	t
1	-17.636	0.198	-89.290	-0.175	0.032	-5.533
2	-10.499	0.144	-73.089	-0.105	0.032	-3.232
3	-6.806	0.133	-51.230	-0.067	0.032	-2.095
4	-3.87	0.129	-30.046	-0.037	0.032	-1.128
5	-1.259	0.122	-10.324	-0.016	0.032	-0.510
6	1.263	0.121	10.431	0.014	0.030	0.476
7	3.883	0.124	31.344	0.041	0.033	1.254
8	6.807	0.131	51.920	0.067	0.030	2.197
9	10.497	0.143	73.599	0.103	0.032	3.220
10	17.629	0.193	91.199	0.175	0.032	5.470

TABLE 2.2

Aggregation Trade-off:
Group size versus number of groups

1,000 replications-M2SLS is efficient

M2SLS Estimator

β	n=10,000 g=100			n=10,000 g=200		
	Mean	sd	MSE	Mean	sd	MSE
0.03	0.032	0.102	0.010	0.037	0.099	0.010
0.02	0.021	0.084	0.007	0.024	0.078	0.006
1	0.998	0.068	0.005	0.997	0.063	0.004
-0.5	-0.496	0.111	0.012	-0.496	0.109	0.012
-0.8	-0.8	0.107	0.012	-0.791	0.103	0.011
0.5	0.499	0.017	0.000	0.499	0.016	0.000

2SLS Estimator

β	n=10,000 g=100			n=10,000 g=200		
	Mean	sd	MSE	Mean	sd	MSE
0.03	0.032	0.107	0.012	0.033	0.105	0.011
0.02	0.032	0.098	0.010	0.025	0.091	0.008
1	0.998	0.077	0.006	0.997	0.071	0.005
-0.5	-0.495	0.124	0.016	-0.495	0.121	0.015
-0.8	-0.800	0.117	0.014	-0.791	0.112	0.013
0.5	0.499	0.019	0.000	0.499	0.018	0.000