

# INSTRUMENTAL VARIABLE ESTIMATION BASED ON GROUPED DATA

Paul A. Bekker and Jan van der Ploeg \*

Department of Economics

University of Groningen

1999

---

\*We would like to thank Tom Wansbeek, Ton Steerneman, Eric Zivot and anonymous referees for helpful comments.

## ABSTRACT

The paper considers the estimation of the coefficients of a single equation in the presence of dummy instruments. We derive pseudo ML and GMM estimators based on moment restrictions induced either by the structural form or by the reduced form of the model. The performance of the estimators is evaluated for the non-Gaussian case. We allow for heteroscedasticity. The asymptotic distributions are based on parameter sequences where the number of instruments increases at the same rate as the sample size. Relaxing the usual Gaussian assumption is shown to affect the normal asymptotic distributions. As a result also recently suggested new specification tests for the validity of instruments depend on Gaussianity. Monte Carlo simulations confirm the accuracy of the asymptotic approach.

**KEYWORDS:** Instrumental variable estimation, limited information maximum likelihood, generalized method of moments, linear functional relationship, group asymptotics, many-instruments asymptotics, natural experiments.

JEL Classification numbers: C1, C2, C3.

Please send correspondence to:

Paul A. Bekker  
Department of Economics  
University of Groningen  
P.O. Box 800  
9700 AV Groningen  
the Netherlands

phone: +31 50 3633817

email: P.A.Bekker@eco.rug.nl

# 1 Introduction

Over the past few years renewed<sup>1</sup> interest in the quality of instrumental variables inference has led to a significant understanding that ‘weak instruments’ can lead to problems when conventional (first-order) asymptotic inference techniques are used. That is to say, the asymptotic distributions of estimators can depart considerably from the exact finite-sample distributions even if the sample size is large. The quality of the approximations depend crucially on the relevance of the instruments, as expressed by the correlation between instruments and explanatory variables, and on the number of instruments. The departures from asymptotic normality include bimodality, bias, fat tails, and missized tests (see e.g. Nelson and Startz (1990a,b), Buse (1992), Maddala and Jeong (1992), Bekker (1994), Bound et al. (1995), and Staiger and Stock (1997)).

Hahn and Inoue (1999) distinguish, in the recent literature, two alternatives for conventional (first-order) large-sample asymptotics. The first is the ‘weak-instrument asymptotics’, or ‘local-to-zero asymptotics’, of Staiger and Stock (1997), where the correlation between instrument and endogenous regressor vanishes as a function of sample size. See also Wang and Zivot (1998). Analyses based on pretesting for weak instruments are given in Hall, Rudebush and Wilcox (1996), Shea (1997), Zivot, Startz, and Nelson (1998), and Startz, Nelson, and Zivot (1999). The second alternative is the ‘many-instrument asymptotics’, where the number of instruments grows at the same rate as the sample size. This approach was first mentioned in Anderson (1976). It was used in e.g. Kunitomo (1980, 1986, 1987) and Morimune (1983), and more recently in Bekker (1994). Hahn and Hausman (1999) show the similarity of this approach to a second-order Edgeworth expansion and use it to specify new specification tests for the validity of instrumental variables. Hahn and Inoue (1999) describe Monte Carlo experiments and conclude that reporting Bekker’s (1994) confidence interval would suffice for most microeconomic applications.

All results derived under ‘many-instruments’ methodology have been restricted to the Gaussian case. This may seem a harmless limitation since, in the model we consider, conventional first-order asymptotic approximations are not affected by departures from Gaussianity. However, the exact distributions of estimators are affected by such departures and

---

<sup>1</sup>Instrumental variables methodology has a long history in econometrics. See e.g. Anderson and Rubin (1949) and Anderson (1950). Surveys have been given by e.g. Mariano (1982) and Phillips (1983). Bowden and Turkington (1990) use it as an organizing principle. Applications can be found in the context of ‘natural experiments’ such as e.g. Angrist (1990), and Angrist and Krueger (1991, 1992).

it will be demonstrated these sensitivities are not washed out by the ‘many-instruments’ methodology. In this paper, we extend the results of Bekker (1994) by relaxing the assumption of Gaussianity. We show that the asymptotic distributions remain normal, but third- and fourth-order moments do enter the expressions. In our context, this holds in particular for the conventional instrumental variable (IV), or 2SLS, estimator. It also holds, to some extent, for the LIML estimator. However, other estimators, whose formulation is closely connected to the model specification, are not affected by such departures.

In addition to relaxing the usual distributional assumption, we also allow for heteroscedasticity, which provides an extension relevant for empirical applications. The derivation of our asymptotic results in this broader context was enabled, however, by restricting our attention to dummy instruments. We consider group indicators as instruments. This need not mean that our analysis is of limited applicability. Frequently, in the context of ‘natural experiments’, instruments do take the form of categorical variables. Such dummy instruments may come naturally, such as the lottery numbers in Angrist (1990), or the season of birth in Angrist and Krueger (1991, 1992), or they may be due to a grouping of the data based on socio-demographic variables. Such grouping, with the aim to generate instruments, has also been used in models with panel data by e.g. Deaton (1985), and Angrist (1991).

In other cases, it may be useful to replace instruments by group-indicator instruments defined as functions of the original instruments. That is to say, functions of instruments may also serve as instruments. In particular, such a reformulation can be useful in a single-equation context, where no assumptions are made about reduced form equations. In that case, first-stage regressions need not be linear in the original instruments and functions may provide additional relevant first-stage regressors.

In this paper we therefore consider the estimation of the coefficients of a single equation based on instruments generated by a grouping of the data. We derive (pseudo) ML and GMM estimators, whose performances will be evaluated by asymptotic distributions based on ‘many-instruments asymptotics’. Angrist and Krueger (1995) refer to such asymptotics as ‘group asymptotics’, which seems to be a suitable name in the present context. We show the group-asymptotic distributions to be more accurate in their approximations to the finite sample distributions of the estimators compared to large sample approximations.

We find that GMM estimators based on moment restrictions induced by the struc-

tural form of the model are inconsistent under group asymptotics. Furthermore, their asymptotic normal distributions are affected by third and fourth-order moments. We do not expect these effects to be restricted to the present context with dummy instruments. In particular, the specification tests of Hahn and Hausman (1999) can be expected to be valid only in the Gaussian case. However, (pseudo) ML and GMM estimators based on the reduced form moment restrictions are found to be group-asymptotically consistent, with simple asymptotic distributions<sup>2</sup>.

The structure of the paper is as follows. Section 2 formally introduces the model. In Section 3 we present GMM and ML estimators. We describe a class of moment estimators and show that GMM estimators based on moment conditions induced by the structural form are not within this class. Section 4 discusses group asymptotics. The asymptotic distributions of the estimators are given in Section 5, where the moment estimators from the class defined in Section 3 are found to be consistent. Section 6 describes consistent estimation of the asymptotic covariance matrices. Monte Carlo simulations are presented in Section 7. The Appendix contains the derivations of the asymptotic distributions.

## 2 The model and some reformulations

Consider the estimation of the coefficients  $\delta \in \mathbb{R}^g$  of a single equation,

$$(1) \quad y_t = x_t' \delta + \varepsilon_t, \quad t = 1, \dots, n,$$

where some of the right-hand-side variables are jointly endogenous, i.e.  $E(\varepsilon_t | x_t)$  need not equal zero. Additional information is provided by observations on instrumental variables  $z_t \in \mathbb{R}^l$  that do satisfy the mean-independence condition:

$$(2) \quad E(\varepsilon_t | z_t) = 0.$$

We consider a cross-section context, where observations on  $(y_t, x_t, z_t)$  are iid. Exogenous variables in  $x_t$  will not be partialled out. They may serve as instruments in  $z_t$ . We derive asymptotic distributions of estimators of  $\delta$ , and we estimate confidence sets, based on an alternative asymptotic approach where the number of instruments increases with the number of observations.

---

<sup>2</sup>If the LIML estimator, which is derived under the assumption of homoscedasticity, is evaluated under heteroscedasticity, it is found to be group-asymptotically inconsistent in general.

The general difficulty in using instrumental variable inference is the practical one of finding instruments. As noted by Kendall and Stuart (1979, Chapter 29.35), a solution might be based on a grouping of observations. Such instrumental group indicators may arise naturally in the context of ‘natural experiments’. Consider, for example, the random lottery numbers in Angrist (1990), the season of birth in Angrist and Krueger (1991), or the timing of mayoral and gubernatorial elections in Levitt (1997). In other cases grouping based on noncategorical instruments may provide a useful procedure to produce instruments. That is to say, let the remaining system equations be given in reduced form by:

$$(3) \quad x'_t = z'_t \Pi + v'_t,$$

where elements of  $v_t$  are zero for explanatory variables that serve as instruments. Contrary to the usual approach when specifying a system of equations—but similar to Bekker (1994)—we do not assume  $E(v_t|z_t) = 0$  in general. In other words, the linearity assumption is applied in a single-equation context and the first-stage regressions need not be linear:  $E(x'_t|z_t) \neq z'_t \Pi$ .

Of course, if the instrumental variables  $z_t$  contain only indicator variables, the linearity of the regression  $E(x_t|z_t)$  is satisfied automatically. In general, however, we might consider functions  $f(z_t) \in \mathbb{R}^m$  of the instruments  $z_t$  as new instruments to achieve identification<sup>3</sup>, or to increase the quality of the instruments. Such reformulated instruments satisfy the exogeneity condition  $E(\varepsilon_t|f(z_t)) = E(E(\varepsilon_t|z_t)|f(z_t)) = 0$ . Furthermore, if the first-stage regressions are nonlinear, the new instruments may show increased relevance, as expressed in  $R^2$  or ‘partial  $R^2$ ’, when compared to the original instrument(s)<sup>4</sup>.

In particular, we consider stepfunctions of  $z_t$  as new instruments. These stepfunctions, or indicator functions, separate the sample into  $j = 1, \dots, m$  groups. The new instruments can be formulated as  $f(z_t) = e_{jt}$ , where the vectors  $e_{jt}$  are columns of the  $m \times m$  identity matrix. As mentioned above, such a reformulation of instruments implies that the first-stage regressions become linear:  $E(x'_t|f(z_t) = e_j) = e'_j A$ , where  $A$  is an  $m \times g$ -matrix containing the  $m$  group means of  $x'_t$ . Another advantage is that heteroscedasticity can be modelled without making further assumptions, i.e.  $E(\varepsilon_t^2|f(z_t) = e_j) = \sigma_j^2, j = 1, \dots, m$ .

---

<sup>3</sup>For example, Lewbel (1996) uses functions of real income level as new instruments to analyse U.K. fuel demand data.

<sup>4</sup>For discussions of “partial  $R^2$ ” in a context of weak instruments see Shea (1997), Hall *et al.* (1996), Wang and Zivot (1998), Zivot *et al.* (1998) and Startz *et al.* (1999).

In addition to these observations, the reformulated framework enabled us to study the consequences of nonnormally distributed disturbances.

Thus motivated, we consider many-instruments-asymptotic inference, or group-asymptotic inference, for  $\delta$  based on dummy instruments. In order not to complicate the notation unduly, we will use the notation  $z_t$  for the (reformulated) dummy instruments, i.e. the vectors  $z_t$  are assumed to be columns of the  $m \times m$  identity matrix. The model can be summarized as follows. We consider a random sample of  $n$  observations on the random vector  $(y, x', z') \in \mathbb{R}^{1+g+m}$ , with finite fourth-order moments, such that

$$\begin{aligned}
y_t &= x_t' \delta + \varepsilon_t, \\
x_t' &= z_t' \Pi + v_t', \\
(4) \quad \mathbb{E}((\varepsilon, v_t') | z_t) &= 0, \\
\mathbb{E}((\varepsilon, v_t')'(\varepsilon, v_t') | z_t = e_j) &= \Sigma_j = \begin{bmatrix} \sigma_j^2 & \sigma_{j12} \\ \sigma_{j21} & \Sigma_{j22} \end{bmatrix},
\end{aligned}$$

for  $t = 1, \dots, n$  and  $j = 1, \dots, m$ . For identification we assume:

$$(5) \quad \text{rank}(\Pi) = g.$$

In reduced form we write  $u_t = \varepsilon_t + v_t' \delta$ , and

$$\begin{aligned}
(y_t, x_t') &= \tilde{z}_t' \Pi(\delta, I_g) + (u_t, v_t'), \\
(6) \quad \mathbb{E}((u_t, v_t') | \tilde{z}_t) &= 0, \\
\mathbb{E}((u_t, v_t')'(u_t, v_t') | \tilde{z}_t = e_j) &= \Omega_j = \begin{bmatrix} \omega_j^2 & \omega_{j12} \\ \omega_{j21} & \Omega_{j22} \end{bmatrix}.
\end{aligned}$$

The relation between  $\Omega_j$  and  $\Sigma_j$  is given by

$$\Omega_j = \begin{bmatrix} 1 & \delta' \\ 0 & I_g \end{bmatrix} \Sigma_j \begin{bmatrix} 1 & 0 \\ \delta & I_g \end{bmatrix}.$$

Let the observations from the  $j$ th group, where  $z_t = e_j$ , be indexed by  $i = 1, \dots, n_j$  and let  $\Pi = (\pi_1, \dots, \pi_m)'$ , then the model equations are also given by

$$\begin{aligned}
(7) \quad y_{ij} &= \pi_j' \delta + u_{ij}, & i &= 1, \dots, n_j, \\
x_{ij} &= \pi_j + v_{ij}, & j &= 1, \dots, m.
\end{aligned}$$

The transformed model (4), (6), or (7), derived from from (1) and (2), is related to the literature in several ways. First, as mentioned above, it is related to instrumental variables models applied in a context of ‘natural experiments’. Another application is when data consist of time series of cross-sections as considered by Deaton (1985), Angrist (1991) and Verbeek and Nijman (1993). The model is also very closely related to the errors-in-variables literature. If  $n_j = 1$  and we assume homoscedasticity,  $\Omega_j = \Omega$ ,  $j = 1, \dots, m$ , and Gaussian disturbances, the model is known in the errors-in-variables literature as a linear functional relationship (Madansky, 1959, Moran, 1971, Kendall and Stuart (Chapter 29), 1979, Aigner *et al.* 1984). In fact, the origin (cf. Anderson, 1976, p.34) of the group-asymptotic approach can be found in this connection.

If  $n_j > 1$ , the model is also known as a linear functional relationship with replicated observations. Kunitomo (1986, 1987) applied group asymptotics to this model. He considered the case of a single jointly endogenous explanatory variable and assumed both homoscedasticity and Gaussianity. Here we relax these assumptions. We show that both heteroscedasticity and non-Gaussian disturbances affect the outcomes.

### 3 GMM based on structural and reduced-form moment conditions

We consider moment restrictions formulated either in terms of the transformed structural model (4), or in terms of its reduced form (6). We will argue that the latter approach is to be “strongly preferred” to the former one. The same words were used by Anderson, Kunitomo and Sawa (1982, p.1025) with regard to 2SLS and LIML. In fact, as will be shown, these are the GMM estimators under homoscedasticity, based on moment restrictions induced by the structural form (4) and the reduced form (6), respectively. Consequently, LIML can be interpreted both as a ML estimator, under a Gaussian assumption, and as a GMM estimator, which is in accordance with the minimum distance interpretation given by Goldberger and Olkin (1971), and it agrees with Pagan’s (1979) interpretation of LIML as an instrumental variable estimator using the implied reduced form coefficients to generate instruments. However, for the general heteroscedastic case we find the ML estimator, under Gaussianity, is different from the GMM estimator based on the reduced form (6).



The resulting estimators are all functions of the first two sample moments

$$(8) \quad \begin{aligned} \bar{Y}_j &= (\bar{y}_j, \bar{x}_j)' = \frac{1}{n_j} \sum_{i=1}^{n_j} (y_{ij}, x_{ij})', \\ S_j &= \frac{1}{n_j - 1} \sum_{i=1}^{n_j} \left( \begin{pmatrix} y_{ij} \\ x_{ij} \end{pmatrix} - \bar{Y}_j \right) \left( \begin{pmatrix} y_{ij} \\ x_{ij} \end{pmatrix} - \bar{Y}_j \right)'. \end{aligned}$$

In fact, the estimators are functions of  $M_j$  and  $S_j$ ,  $j = 1, \dots, m$ , where

$$(9) \quad M_j = w_j \bar{Y}_j \bar{Y}_j',$$

and

$$w_j = n_j / \sum_{j=1}^m n_j.$$

A particular function  $\hat{\delta}(M_j, S_j, j = 1, \dots, m)$  is said to be in the class of moment estimators based on the statistics  $M_j$  and  $S_j$ ,  $j = 1, \dots, m$ , if

$$(10) \quad \hat{\delta}(E(M_j), E(S_j), j = 1, \dots, m) = \delta.$$

The expectations are given by

$$(11) \quad \begin{aligned} E(M_j) &= w_j (\delta, I_g)' \pi_j \pi_j' (\delta_1 I_g) + \Omega_j / \sum_{j=1}^m n_j, \\ E(S_j) &= \Omega_j. \end{aligned}$$

It will be shown that GMM estimators based on moment restrictions formulated in terms of the structural form (4) do not satisfy the moment equations (10). As a result these estimators might be badly located. As will be shown, this is reflected by their group-asymptotic inconsistency.

### 3.1 GMM based on structural-form moment conditions

For our computations we use matrices  $\Sigma_j$ , or  $\Omega_j$ , that are nonsingular. However, our results also hold for singular matrices<sup>5</sup>. GMM estimation based on (4) can be formulated in terms of the following moment equations:

$$E [z_t'(y_t - x_t' \delta, x_t' - z_t' \Pi)] = 0.$$

---

<sup>5</sup>This is implied by a continuity argument since singular matrices can be approximated arbitrarily close by nonsingular matrices. Alternatively, one may transform the model into a form where the covariance matrices of the observations are nonsingular. Then one may perform the necessary inversions and, finally, one may retransform the outcome to find the same result.

The GMM estimator for  $\delta$  under homoscedasticity is given by

$$(12) \quad \begin{aligned} \hat{\delta}_{IV} &= \arg \min_{\delta} \sum_{j=1}^m \tilde{\delta}' M_j \tilde{\delta}, \quad \tilde{\delta} = (1, \delta')' \\ &= \left( \sum_{j=1}^m w_j \bar{x}_j \bar{x}_j' \right)^{-1} \sum_{j=1}^m w_j \bar{x}_j \bar{y}_j. \end{aligned}$$

It follows from (11) that  $\hat{\delta}_{IV}$  does not satisfy (10), which explains intuitively the bias of this estimator. The same holds for the heteroscedastic version. Let  $\hat{\sigma}_{jIV}^2$  be an estimate of  $\sigma_j^2$ , for example  $(1, -\hat{\delta}'_{IV}) S_j (1, -\hat{\delta}'_{IV})'$ , then the GMM estimator based on the structural-form moment equations is given by

$$(13) \quad \hat{\delta}_{IVGLS} = \arg \min_{\delta} \sum_{j=1}^m \frac{\tilde{\delta}' M_j \tilde{\delta}}{\hat{\sigma}_{jIV}^2}.$$

### 3.2 GMM based on reduced-form moment conditions

GMM based on (6) can be formulated in terms of the following moment restrictions:

$$E [z_t (y_t - z_t' \Pi \delta, x_t' - z_t' \Pi)] = 0.$$

The GMM estimator under homoscedasticity is given by<sup>6</sup>

$$(14) \quad \hat{\delta}_{LIML} = \arg \min_{\delta} \frac{\sum_{j=1}^m \tilde{\delta}' M_j \tilde{\delta}}{\sum_{j=1}^m (n_j - 1) \tilde{\delta}' S_j \tilde{\delta}}, \quad \tilde{\delta} = (1, \delta')'.$$

Here we used  $\frac{\sum_{j=1}^m (n_j - 1) S_j}{\sum_{j=1}^m (n_j - 1)}$  as an (unbiased) estimate for  $\Omega_j = \Omega$ . Indeed  $\hat{\delta}_{LIML}$  is the LIML estimator, i.e. the ML estimator under the assumptions of homoscedasticity and normality of the disturbances.

Under homoscedasticity the LIML estimator is a moment estimator in the sense of (10). However, in case of heteroscedasticity LIML will satisfy (10) only if the group sizes  $n_j$  are equal. As a consequence we find LIML to be group-asymptotically inconsistent in general. By making a small adaptation, we do get such a moment estimator:

$$(15) \quad \hat{\delta}_{MM} = \arg \min_{\delta} \frac{\sum_{j=1}^m \tilde{\delta}' M_j \tilde{\delta}}{\sum_{j=1}^m \tilde{\delta}' S_j \tilde{\delta}}.$$

This estimator equals the LIML estimator if the group-sizes are equal. If the group-sizes vary it still satisfies (10): it will be group-asymptotically consistent even under

---

<sup>6</sup>All derivations are available from the authors.

heteroscedasticity. The computation of both MM (15) and LIML (14) amounts to a simple eigenvalue problem.

The GMM estimator under heteroscedasticity is given by

$$(16) \quad \hat{\delta}_{GMM} = \arg \min_{\delta} \sum_{j=1}^m \frac{\tilde{\delta}' M_j \tilde{\delta}}{\tilde{\delta}' S_j \tilde{\delta}},$$

where we used  $S_j$  as an (unbiased) estimator of  $\Omega_j$ . It follows from (11) that this GMM estimator is a moment estimator in the sense of (10). However, the computation of the GMM estimator (16) cannot be reduced to a simple eigenvalue problem. Here we need an iterative optimization procedure. The MM estimator may serve as a starting value.

The question is whether the GMM estimator (16) is also the ML estimator under normality and heteroscedasticity, as it is true for LIML in the homoscedastic case. We found it to be different. The maximum likelihood estimator is given by

$$(17) \quad \hat{\delta}_{ML} = \arg \min_{\delta} \sum_{j=1}^m n_j \log \left( \frac{n_j - 1}{\sum_{j=1}^m n_j} + \frac{\tilde{\delta}' M_j \tilde{\delta}}{\tilde{\delta}' S_j \tilde{\delta}} \right).$$

Just as GMM, (16), it is a moment estimator and it can be computed iteratively using the MM estimator as starting value. We also notice that in case of just-identification, when  $m = g$ , all estimators coincide and satisfy  $\sum_{j=1}^m M_j \hat{\delta} = 0$ .

## 4 Group-asymptotics

Notice that if the group sizes equal one,  $n_j = 1$ , then ‘large-sample’ asymptotics, applied to formulation (7), would amount to  $m \rightarrow \infty$ . In fact, Anderson (1976, 1984) showed that estimation in the linear functional relationship is equivalent to limited information estimation of a single equation in a simultaneous equations model. However, in the latter case the asymptotics would be different. So, the exact distribution of a single estimator would be approximated by different ‘large-sample’ distributions depending arbitrarily on how the model has been formulated. In fact, the origin of the alternative asymptotic approach can be found in this connection: group-asymptotics combines<sup>7</sup> both approaches.

---

<sup>7</sup>Such combination was first suggested in Anderson (1976, p. 34).

We use the approach described in Bekker (1994), where the number of instruments is allowed to grow as the number of observations increases. Hahn and Inoue (1999) refer to such sequences as ‘many-instruments-asymptotics’, and Hahn and Hausman (1999) interpret the resulting approximations as a convenient method of Edgeworth expansion “with wider applicability than might be thought considering Bekker-type asymptotics in isolation.”

In traditional large sample asymptotics, the first-stage regression coefficients can be consistently estimated by OLS. As a result the uncertainty about the first-stage regression coefficients does not affect the large-sample asymptotic distribution of the 2SLS estimator. Exact knowledge about these coefficients would not increase the large-sample asymptotic efficiency<sup>8</sup>. However, such additional information might be very useful in the actual finite-sample setting. Consequently, one may expect large-sample approximations to be overly optimistic in terms of efficiency. As a result, sizes of conventional large-sample tests based on 2SLS might become abysmally large (cf. Hahn and Hausman, 1999). By contrast, under group asymptotics, the number of first-stage regression coefficients increases with the number of instruments and should be viewed as incidental parameters (cf. Neyman and Scott, 1948). Consequently, these parameters cannot be consistently estimated and the uncertainty about the group means  $\pi_j$ , which is present in the finite sample, will also be present asymptotically.

In fact, we consider replicated groups, indicated by  $k = 1, \dots, K$ , such that the observations in the replications are independent with different group means  $\pi_{jk}$  and covariance matrices  $\Omega_{jk} = \Omega_j$ ,  $j = 1, \dots, K$ . The latter condition allows for consistent estimation of these covariance matrices. For the independent random  $(g + 1)$ -vectors  $(y_{ijk}, x'_{ijk})$ , where  $i = 1, \dots, n_j$ ,  $j = 1, \dots, m$ , and  $k = 1, \dots, K$ , we have

$$\begin{aligned} \text{E}(y_{ijk}, x'_{ijk}) &= \pi'_{jk}(\delta, I_g), \\ \text{Var}(y_{ijk}, x'_{ijk}) &= \Omega_j. \end{aligned} \tag{18}$$

So, the data in the sample conform to  $K = 1$ , where the index  $k$  has been suppressed.

Large-sample asymptotic theory for simultaneous equations models, where  $K$  is fixed

---

<sup>8</sup>In fact, in addition to  $\text{E}[z_t(y_t - x'_t\delta)] = 0$  the moment restrictions  $\text{E}[z_t(x'_t - z'_t\Pi)] = 0$  are irrelevant to produce (12) as a GMM estimator.

and the group sizes increase,  $n_j \rightarrow \infty$ , shows

$$(19) \quad \begin{aligned} n^{1/2}(\hat{\delta}_{IV} - \delta) &\overset{A}{\rightsquigarrow} N\left(0, \left(\sum_{j=1}^m w_j \pi_j \pi'_j\right)^{-1} \sum_{j=1}^m \sigma_j^2 w_j \pi_j \pi'_j \left(\sum_{j=1}^m w_j \pi_j \pi'_j\right)^{-1}\right), \\ n^{1/2}(\hat{\delta}_{IVGLS} - \delta) &\overset{A}{\rightsquigarrow} N\left(0, \left(\sum_{j=1}^m \sigma_j^{-2} w_j \pi_j \pi'_j\right)^{-1}\right), \end{aligned}$$

where  $w_j$  has been defined below (9). Furthermore,  $\hat{\delta}_{IV}$  may be replaced by  $\hat{\delta}_{LIML}$  or  $\hat{\delta}_{MM}$ , and  $\hat{\delta}_{IVGLS}$  may be replaced by  $\hat{\delta}_{GMM}$  or  $\hat{\delta}_{ML}$ . In case of homoscedasticity the asymptotic distributions coincide. In case of overidentification,  $m > g$ , and heteroscedasticity, the latter estimators are asymptotically more efficient. The large-sample sequence can also be represented by fixing  $n_j$  and increasing  $K$  under the restrictions

$$(20) \quad \pi_{jk} = \pi_j, \quad j = 1, \dots, m, \quad k = 1, \dots, K.$$

Notice that the number of parameters in  $\pi_{jk}$ ,  $j = 1, \dots, m$ , remains fixed in this sequence.

As an alternative, we consider sequences where the number of first-stage regression coefficients  $\pi_{jk}$  increases. That is, sequences where  $K \rightarrow \infty$  and the restrictions (20) are relaxed. We let

$$(21) \quad \begin{aligned} \frac{1}{K} \sum_{k=1}^K \pi_{jk} &\rightarrow \bar{\pi}_j, \\ \frac{1}{K} \sum_{k=1}^K \pi_{jk} \pi'_{jk} &\rightarrow \overline{\pi_j \pi'_j}. \end{aligned}$$

So, we have  $mK$  groups, or instruments, and  $K \sum_{j=1}^m n_j$  observations. Both increase as  $K \rightarrow \infty$ . In order to combine such sequences with large sample sequences, we also allow for increasing group sizes  $n_j$ :

$$(22) \quad \frac{Km}{K \sum_{j=1}^m n_j} = \frac{m}{\sum_{j=1}^m n_j} \rightarrow \alpha.$$

The large sample asymptotics is found for<sup>9</sup>  $\alpha = 0$  and  $\bar{\pi}_j = \pi_j$  and  $\overline{\pi_j \pi'_j} = \pi_j \pi'_j$ . Sequences with  $\alpha > 0$  will be referred to as group-asymptotics.

For sequences where  $K$  increases and (20) is not satisfied, the GMM and ML estimators, derived under normality, have to be found by optimization over an increasing

---

<sup>9</sup>Notice the number of instruments may grow at a rate faster than  $O(n^{\frac{1}{3}})$ , which is given as a sufficient upper bound by Koenker and Machado (1999) for the validity of conventional asymptotic inference about the GMM estimator.

number of incidental parameters. The expressions for these estimators, and also for the MM-estimator, take forms similar to the ones described in the previous section for  $K = 1$ . However, the definitions of the matrices  $S_j$  and  $M_j$  in (8) and (9), resp., should be extended as follows to cases where  $K > 1$ :

$$\begin{aligned}
\bar{Y}_{jk} &= \frac{1}{n_j} \sum_{i=1}^{n_j} (y_{ijk}, x'_{ijk})', \\
(23) \quad S_j &= \frac{1}{K} \sum_{k=1}^K \frac{1}{n_j - 1} \sum_{i=1}^{n_j} \left( \begin{pmatrix} y_{ijk} \\ x_{ijk} \end{pmatrix} - \bar{Y}_{jk} \right) \left( \begin{pmatrix} y_{ijk} \\ x_{ijk} \end{pmatrix} - \bar{Y}_{jk} \right)', \\
M_j &= \frac{1}{K} \sum_{k=1}^K w_j \bar{Y}_{jk} \bar{Y}'_{jk}.
\end{aligned}$$

The asymptotic normality of the estimators is based on the asymptotic normality of the statistics  $S_j$  and  $M_j$ . That is, let  $\eta_j = \text{vec}(S_j, M_j)$ , then for sequences satisfying (21) and (22), with  $w_j$  fixed, we find

$$\begin{aligned}
n^{\frac{1}{2}}(\eta_j - \text{E}(\eta_j)) &\overset{A}{\rightsquigarrow} N(0, \lim_{n \rightarrow \infty} n \text{Var}(\eta_j)), \\
(24) \quad n &= K \sum_{j=1}^m n_j.
\end{aligned}$$

If  $K$  is fixed and  $n_j \rightarrow \infty$ , or, if  $n_j$  is fixed and  $K \rightarrow \infty$ , while (20) is satisfied, then (24) follows from application of the Lindeberg-Lévy CLT. If  $n_j$  is fixed and  $K \rightarrow \infty$ , while (20) is not satisfied, the result follows from application of Hajèk-Sidak's Theorem (Sen and Singer, 1993, p 119). Based on (24) we find

$$\begin{aligned}
(25) \quad \text{plim}(S_j) &= \lim_{n \rightarrow \infty} \text{E}(S_j) = \Omega_j, \\
\text{plim}(M_j) &= \lim_{n \rightarrow \infty} \text{E}(M_j) = w_j(\delta, I_g)' \overline{\pi_j \pi_j'}(\delta, I_g) + \frac{\alpha}{m} \Omega_j.
\end{aligned}$$

So, for the minimum in (16) we find:

$$(26) \quad \text{plim} \left( \sum_{j=1}^m \frac{\hat{\delta}'_{GMM} M_j \hat{\delta}_{GMM}}{\hat{\delta}'_{GMM} S_j \hat{\delta}_{GMM}} \right) = \alpha.$$

## 5 The group-asymptotic distributions

The large sample asymptotic distributions, as given in (19) show that all estimators are consistent and their asymptotic distributions are not affected by deviations from Gaussianity. Of course, such deviations do affect the finite sample distributions. The difference between finite and large sample distributions is due to the approximating nature of asymptotics. Phillips (1983, p. 508) remarks that “For the process by which asymptotic machinery works inevitably washes out sensitivities that are present and important in finite samples”. This holds true for the group-asymptotic distributions as well. However, we find these approximations to be more accurate and less inclined to wash out sensitivities.

Contrary to large-sample inference, the GMM estimators based on moment conditions induced by the structural form, i.e. IV and IVGLS, are found to be inconsistent. Furthermore, we find that the group-asymptotic distributions of GMM estimators, derived under homoscedasticity, are affected by non-Gaussian deviations. That is, both IV, or 2SLS, and LIML are affected. Interestingly, under heteroscedasticity, the GMM estimator based on the reduced form and the ML estimator are consistent and have the same group-asymptotic distribution that is not affected by third and fourth-order moments.

The derivations of the group-asymptotic distributions are given in the Appendix. When denoting third and fourth order moments, we drop the index  $i$ , i.e.  $E(\varepsilon_j v_j v_j') \equiv E(\varepsilon_{ij} v_{ij} v_{ij}')$ ; in case of homoscedasticity we also drop the index  $j$ . If a parameter is constant over all groups, we drop the index  $j$  as well. If a parameter varies, we use a bar to denote its mean, i.e.  $\bar{\Sigma} \equiv \sum_{j=1}^m \Sigma_j / m$ . Due to its frequent occurrence we use  $\phi_j \equiv \sigma_{j21} / \sigma_j$ .

### 5.1 IV

The GMM estimators based on the structural form, i.e. IV, (12), and IVGLS, (13), are not moment estimators in the sense of the moment equations (10). As a result, these estimators are found to be inconsistent for group-asymptotic sequences if  $\alpha > 0$ . Due to this inconsistency, the group-asymptotic distributions are complicated. Furthermore, the asymptotic distributions are affected, rather strongly, by deviations from normality. So the analytical properties of these estimators are far from attractive. Here we will consider only IV. We give its group-asymptotic distribution for the general heteroscedastic case.

It follows immediately from the probability limits (25) of the matrices  $M_j$  and  $S_j$  that  $\hat{\delta}_{IV}$  is inconsistent. Let  $\delta_{IV}^* = \text{plim } \hat{\delta}_{IV}$ , then

$$\delta_{IV}^* - \delta = \alpha \left\{ \sum_{j=1}^m (w_j \overline{\pi_j \pi_j'}) + \alpha \bar{\Sigma}_{22} \right\}^{-1} \bar{\sigma}_{21}.$$

So IV is consistent only if either  $\alpha = 0$ , i.e. when large-sample asymptotics is considered, or  $\bar{\sigma}_{21} = 0$ , which amounts to exogeneity of the explanatory variables. The inconsistency might be small if the group sizes are large and the group means are well-spread, i.e. the first-stage regressions have a high  $R^2$ . Furthermore, the endogeneity of the explanatory variables should not be large. Hahn and Hausman (1999) and Hahn and Inoue (1999) discuss these matters in detail for an illustrative simple model.

In order to present the asymptotic distribution, let  $\sigma_j^{2*} = (1, -\delta_{IV}^{*'}) \Omega_j (1, -\delta_{IV}^{*'})'$  and  $\sigma_{j21}^* = (0, I_g) \Omega_j (1, -\delta_{IV}^{*'})'$ , then

$$n^{1/2}(\hat{\delta}_{IV} - \delta_{IV}^*) \overset{A}{\sim} N(0, V_{IV}),$$

$$V_{IV} = (A_1 + A_2)^{-1} (B_1 + B_2 + B_3) (A_1 + A_2)^{-1},$$

where

$$\begin{aligned} A_1 &= \sum_{j=1}^m w_j \overline{\pi_j \pi_j'}, \quad A_2 = \alpha \bar{\Sigma}_{22}, \\ B_1 &= \sum_{j=1}^m \sigma_j^{2*} w_j \overline{\pi_j \pi_j'}, \\ B_2 &= \sum_{j=1}^m \{ w_j (\delta_{IV}^* - \delta)' \overline{\pi_j \pi_j'} (\delta_{IV}^* - \delta) \Sigma_{j22} \\ &\quad - w_j \sigma_{j21}^* (\delta_{IV}^* - \delta)' \overline{\pi_j \pi_j'} - w_j \overline{\pi_j \pi_j'} (\delta_{IV}^* - \delta) \sigma_{j12}^* \\ &\quad + (\alpha/m) \sigma_j^{2*} \Sigma_{j22} + (\alpha/m) \sigma_{j21}^* \sigma_{j12}^* \}, \\ B_3 &= \sum_{j=1}^m \left[ w_j \left( \frac{\alpha}{w_j m} \right)^2 \left\{ E((\varepsilon_j - v_j'(\delta_{IV}^* - \delta))^2 v_j v_j') - \sigma_j^{2*} \Sigma_{j22} - 2\sigma_{j21}^* \sigma_{j12}^* \right\} \right. \\ &\quad - 2(\alpha/m) \bar{\pi}_j' (\delta_{IV}^* - \delta) E((\varepsilon_j - v_j'(\delta_{IV}^* - \delta)) v_j v_j') \\ &\quad + 2(\alpha/m) \bar{\pi}_j E((\varepsilon_j - v_j'(\delta_{IV}^* - \delta))^2 v_j') \\ &\quad \left. + 2(\alpha/m) E((\varepsilon_j - v_j'(\delta_{IV}^* - \delta))^2 v_j) \bar{\pi}_j' \right]. \end{aligned}$$



We find the asymptotic IV distribution is affected by deviations from normality. That is to say, the third-order moments in  $B_3$  vanish only under normality, where  $B_3 = 0$ , or if  $\alpha = 0$ , so that  $A_2 = B_2 = B_3 = 0$ . In the latter case we find the large-sample asymptotic distribution as given in (19). In general the asymptotic distribution seems rather complicated, although the expression becomes more transparent when terms of order  $\alpha^2$  are ignored.

In addition to the inconsistency, the expression for the asymptotic distribution shows the 2SLS estimator is affected rather strongly by non-Gaussianity. As this result can be expected to hold in general, i.e. not restricted to a context with indicator instruments, it shows the tests proposed in Hahn and Hausman (1999), which are based on biased-corrected 2SLS, are affected by non-Gaussianity. Whether this could have a relevant effect on the accuracy of the proposed inference procedures remains to be seen.

## 5.2 LIML

It is well-known from Neyman and Scott (1948) that ML estimators of structural parameters are not necessarily consistent in the presence of incidental parameters. However, the incidental parameters in the group-asymptotic sequences do not affect the consistency of the LIML estimator, (14). It is consistent under homoscedasticity, since in that case it is a moment estimator in the sense of the moment equations (10). Its group-asymptotic distribution has been computed under the assumption of homoscedasticity. It is given by

$$n^{1/2}(\hat{\delta}_{LIML} - \delta) \overset{A}{\rightsquigarrow} N(0, V_{LIML}),$$

$$V_{LIML} = \sigma^2 A_1^{-1} (B_1 + B_2 + B_3) A_1^{-1},$$

where

$$A_1 = B_1 = \sum_{j=1}^m w_j \overline{\pi_j \pi_j'},$$

$$B_2 = \frac{\alpha}{(1-\alpha)} (\Sigma_{22} - \phi \phi'),$$

$$\begin{aligned}
B_3 = & \left( \frac{\alpha}{1-\alpha} \right) \sum_{j=1}^m \left[ (w_j - m^{-1}) \left\{ \bar{\pi}_j (E(\varepsilon/\sigma)^3 \phi' + E((\varepsilon/\sigma)^2 v')) \right. \right. \\
& \left. \left. + (E(\varepsilon/\sigma)^3 \phi + E((\varepsilon/\sigma)^2 v)) \bar{\pi}'_j \right\} \right] + \\
& \left( \frac{\alpha}{1-\alpha} \right)^2 \sum_{j=1}^m \left[ \frac{(w_j - m^{-1})^2}{w_j} \left\{ E(\varepsilon/\sigma)^4 \phi \phi' + E((\varepsilon/\sigma)^2 v v') \right. \right. \\
& \left. \left. - \phi E((\varepsilon/\sigma)^3 v') - E((\varepsilon/\sigma)^3 v) \phi' - \Sigma_{22} + \phi \phi' \right\} \right].
\end{aligned}$$

Again we find the group-asymptotic distribution is affected by non-Gaussian disturbances. The third- and fourth-order moments in  $B_3$  vanish only under normality, where  $B_3 = 0$ , or if either  $\alpha = 0$  so that  $B_2 = B_3 = 0$ , or group-sizes are equal:  $w_j = m^{-1}$ . Especially in the latter case the asymptotic distribution of LIML is more simple than the one found for 2SLS if  $\alpha > 0$ . Of course, if  $\alpha = 0$  the asymptotic distribution boils down to (19) again.

We also find, under normality and homoscedasticity, that LIML is group-asymptotically efficient among the other consistent estimators considered in this paper. That is, under normality and homoscedasticity.  $V_{LIML} \leq V_{MM} = V_{GMM}$ , where the difference is zero if the group sizes are equal or  $\alpha = 0$ .

These results for LIML have been derived under the assumption of homoscedasticity. We found LIML to be consistent over this restricted parameter space. However, in case of heteroscedasticity, LIML is no longer a moment estimator in the sense of (10). In general, under heteroscedasticity (when LIML is not an ML estimator even under Gaussianity), LIML is inconsistent. However, using a small adaptation we were able to formulate the MM estimator (15). If the group sizes are equal, LIML is numerically equivalent to the MM estimator. The latter estimator is a moment estimator, in the sense of (10), even if the disturbances vary heteroscedastically. Its group-asymptotic distribution has been computed for the general heteroscedastic case.

### 5.3 MM

The moment estimator (15) is consistent under heteroscedasticity and its group-asymptotic distribution is given by

$$n^{1/2}(\hat{\delta}_{MM} - \delta) \overset{A}{\rightsquigarrow} N(0, V_{MM}),$$

$$V_{MM} = A_1^{-1}(B_1 + B_2)A_1^{-1},$$

where

$$\begin{aligned}
A_1 &= \sum_{j=1}^m w_j \overline{\pi_j \pi_j'}, \\
B_1 &= \sum_{j=1}^m \sigma_j^2 w_j \overline{\pi_j \pi_j'}, \\
B_2 &= \alpha \sum_{j=1}^m \frac{w_j \sigma_j^2}{(w_j m - \alpha)} \left\{ \Sigma_{j22} - \phi_j \phi_j' + 2 \left( \phi_j - \frac{\sigma_j \bar{\sigma}_{21}}{\sigma^2} \right) \left( \phi_j - \frac{\sigma_j \bar{\sigma}_{21}}{\sigma^2} \right)' \right\}.
\end{aligned}$$

Its large sample asymptotic distribution, found for  $\alpha = 0$ , is the same as for LIML. Remarkably, and contrary to LIML, the group-asymptotic distribution of the MM estimator is not affected by third and fourth order moments. In addition to its consistency, the simple asymptotic distribution of the MM estimator is an attractive property. This holds in particular for the construction of confidence intervals based on an estimation of  $V_{MM}$ . Due to its consistency under heteroscedasticity the MM estimator may also serve as an initial value to compute the GMM and ML estimators iteratively.

#### 5.4 GMM and ML

The GMM estimator (16) based on the heteroscedastic reduced form is numerically different from the ML-estimator under heteroscedasticity (17), but their group-asymptotic distributions are equal. Contrary to the estimators derived under homoscedasticity, they are consistent in general. The asymptotic distribution is given by

$$n^{1/2}(\hat{\delta}_{GMM} - \delta) \overset{A}{\sim} N(0, V_{GMM}),$$

$$V_{GMM} = A_1^{-1}(B_1 + B_2)A_1^{-1},$$

where

$$\begin{aligned}
A_1 = B_1 &= \sum_{j=1}^m \sigma_j^{-2} w_j \overline{\pi_j \pi_j'}, \\
B_2 &= \alpha \sum_{j=1}^m \frac{w_j \sigma_j^{-2}}{(w_j m - \alpha)} \left( \Sigma_{j22} - \phi_j \phi_j' \right).
\end{aligned}$$

Here we find a relatively simple group-asymptotic distribution that is not affected by deviations from Gaussianity. The large sample asymptotic distribution, found for  $\alpha = 0$ , is efficient under both hetero and homoscedasticity. Also, if  $\alpha \neq 0$  and  $\sigma_j^2 = \sigma^2$ , we find  $V_{GMM} \leq V_{MM}$ , although the difference is zero in case of homoscedasticity, i.e. when  $\Sigma_j = \Sigma$ . However, it cannot be established in general that  $V_{GMM} \leq V_{MM}$ . If the group

sizes are equal,  $w_j = 1/m$ , and  $\sigma_j^2 = \sigma^2$  then we find  $V_{GMM} = V_{LIML}$ . Consequently, the GMM estimator has many attractive properties, including a remarkably simply asymptotic distribution, which is not affected by third and fourth-order moments.

## 6 Consistent estimation of the group-asymptotic distributions.

In order to compute confidence intervals based on the asymptotic distributions, by inverting Wald-tests, the group-asymptotic covariance matrices should be consistently estimated. Under Gaussianity and homoscedasticity a consistent estimator of  $V_{LIML}$  has been given in Bekker (1994, (4.11)). In the present context, the covariance matrix  $V_{LIML}$  is affected, to some extent, by third and fourth-order moments. A stronger dependence, which is present even when group sizes are equal, is present in the asymptotic distributions of IV and IVGLS. However, we make no attempt at estimating third and fourth-order moments in order to estimate asymptotic covariance matrices of inconsistent estimators. We only consider consistent estimation of  $V_{MM}$  and  $V_{GMM}$ . Such estimators can be easily formulated based on (25) and (26). In fact, for the simulations we used estimators of  $\alpha$  for each group separately.

$$(27) \quad \hat{\alpha}_j = mw_j \frac{(1, -\hat{\delta}')M_j(1, -\hat{\delta}')'}{(1, -\hat{\delta}')N_j(1, -\hat{\delta}')'}$$

where

$$(28) \quad N_j = \frac{n_j - 1}{\sum_{j=1}^m n_j} S_j + M_j,$$

and  $\hat{\delta}$  is a consistent estimator, such as  $\hat{\delta}_{MM}$  or  $\hat{\delta}_{GMM}$ . The matrices  $\overline{\pi_j \pi_j'}$  were estimated by  $M_{j22} - \frac{\hat{\alpha}_j}{m} S_{j22}$ .

## 7 Monte Carlo simulations

### 7.1 The design.

A Monte Carlo study by Hahn and Inoue (1999) confirms that confidence sets based on Bekker's (1994) estimator of  $V_{LIML}$  perform very well, even when the correlation between the endogenous regressor and the instruments is very small. The present Monte Carlo

experiment focusses on the quality of group-asymptotic approximations to the exact distributions of the estimators and on the performance of group-asymptotic confidence sets in a heteroscedastic context.

In the simulations the group sizes were taken equal,  $n_j = n/m$ , so that the LIML and MM estimators were numerically equivalent. We considered a single explanatory variable and used, without loss of generality,  $\delta = 0$ . The  $2 \times 2$  matrices  $\Sigma_j$  have been parameterized as

$$\Sigma_j = \begin{bmatrix} \sigma_j^2 & r\sigma_j\sigma_{vj} \\ r\sigma_j\sigma_{vj} & \sigma_{vj}^2 \end{bmatrix},$$

so that the correlation,  $r$ , is fixed. The values for  $\sigma_j$ ,  $\sigma_{vj}$  and the group means  $\pi_j$  were found by independent drawings from uniform distributions. That is,  $\pi_j$  was drawn from  $U(\frac{1}{2}p, 1\frac{1}{2}p)$  and both  $\sigma_j$  and  $\sigma_{vj}$  were drawn from  $U(\frac{1}{2}, 1\frac{1}{2})$ . Consequently, the input parameters for the simulations were given by:  $n, m, p$  and  $r$ .

We used three values of  $n$  and for each such value we used three values of  $p$ . That is, we used:

$n = 120$ :  $p = .7, 1, 3$ ;

$n = 600$ :  $p = .3, .7, 1$ ;

$n = 3000$ :  $p = .1, .3, .5$ .

We considered all parameter combinations with  $m = 3, 10, 30$ , and  $r = .1, .5, .7$ . Using these 81 parameter combinations we generated Gaussian observations with 20,000 replications.

The smaller values of  $p$  correspond to cases with instruments that are not strong. However, these values are not that extreme to make the bimodality of the IV distribution (Nelson and Startz, 1990; Maddala and Jeong, 1992) manifest.

## 7.2 Approximations.

Let  $M_1$  and  $M_2$  be given by the maximum absolute difference between the simulated distributions<sup>10</sup> and the asymptotic approximations for  $\alpha = 0$  and  $\alpha = \alpha^* = (m-1)/(n-1)$ ,

---

<sup>10</sup>The distributions of the estimators have been estimated at the points  $P(\hat{\delta} \leq x(Asd1))$ ,  $x = -3.0(0.5) - 2.0(0.6) - 1.4(0.4) - 1.0(0.2)1.0(0.4)1.4(0.6)2.0(0.5)3.0$ .

respectively. The latter value for  $\alpha$  equals the ratio of the degree of overidentification and the total number of degrees of freedom<sup>11</sup> (cf. Bekker, 1994). We computed the following estimators: IV, MM (= LIML), GMM and ML. The asymptotic standard deviation for IV is denoted by  $Asd1_{IV} = (V_{IV}/n)^{1/2}$  if  $\alpha = 0$ , and by  $Asd2_{IV}$  if  $\alpha = \alpha^*$ . We use similar expressions for the remaining estimators.

We find for all estimators and virtually all parameter combinations  $M_1 > M_2$ . For IV, a plot of  $M_1$  and  $M_2$  against the normalized bias and, for the remaining estimators, plots of  $M_1 - M_2$  against the normalized standard deviations,  $Asd2/Asd1$ , provide pictures very similar to the ones given in Bekker (1994) for the homoscedastic case.

TABLE 1  
Maximum Absolute Differences between  
Empirical and Asymptotic  
Cumulative Distribution Functions

	Mean		Maximum	
	$M_1$	$M_2$	$M_1$	$M_2$
IV	.146	.010	.708	.037
MM	.025	.016	.122	.065
GMM	.039	.030	.136	.093
ML	.045	.037	.182	.146

Table 1 gives for each estimator the mean and maximum over the 81 values of  $M_1$  and  $M_2$ . It shows that the improvement in fit achieved by the alternative approximation,  $\alpha = \alpha^*$ , compared to  $\alpha = 0$  is most dramatic for IV. The fit of IV is rather good. For the other estimators the deviations between the simulated distribution and their approximations can be explained in part by the skewness of the distributions (cf. Bekker, 1994). In addition, the simulated distributions are more spread out compared to their approximations. This effect is rather mild for MM, but it gets stronger in the sequence MM, GMM, ML. We find that ML does not perform well for 9 extreme parameter points where the group sizes equal 4.

<sup>11</sup>Alternatively,  $\alpha^*$  is the ratio of the number of incidental parameters in  $\Pi$ ,  $(mg)$ , minus the minimum number needed for identification,  $(g^2)$ , and the number of observations in  $x_{ij}$ ,  $(ng)$ , minus the number needed for identification,  $(g^2)$ .

### 7.3 Means and MSE's.

Here we consider the means and MSE's of the simulated distributions as informative summary statistics<sup>12</sup>. For IV we find that the means and MSE's behave very much like their asymptotic ( $\alpha = \alpha^*$ ) counterparts. When plotted against each other we find almost perfect straight lines with slope 1.

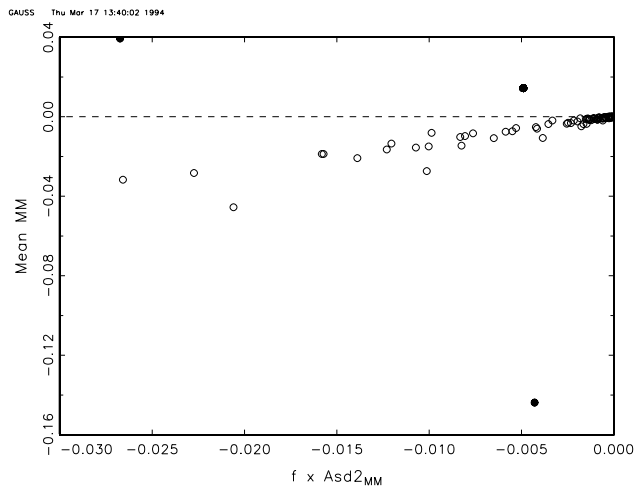


Figure 1: The mean of MM.

Figure 1 gives the mean of MM against  $f \times Asd2_{MM}$ , where

$$f = -r / \sqrt{\sum_{j=1}^m n_j \pi_j^2 / \sigma_{v_j}^2},$$

is a measure of skewness similar to the one used by Bekker (1994). Indeed we find the mean of the simulations is affected by the skewness. The three differently indicated points<sup>13</sup> correspond to the largest values of  $Asd2_{MM}$ . For GMM we find very similar plots. For ML, the number and extremeness of the outliers is much larger.

Figure 2 gives the MSE of MM against  $(Asd2_{MM})^2$ . However, we excluded 6 outliers with relatively weak instruments<sup>14</sup> and 2 outliers<sup>15</sup> where the group sizes equal 4. We find that the MSE is bounded by its asymptotic counterpart, indicating that the simulated

<sup>12</sup>The first two moments of the exact distributions do not exist for the LIML-like estimators. So the results should be interpreted with some care. Outliers may be the result of fat tails.

<sup>13</sup>Found for  $p = .1$ ,  $r = .1$ ,  $m = 10, 30$ ;  $p = .1$ ,  $r = .5$ ,  $m = 30$ .

<sup>14</sup>Found for  $p = .1$ ,  $r = .1, .5$ .

<sup>15</sup>Found for  $p = .7$ ,  $r = .1, .5$ .

distribution is more spread out than its asymptotic approximation.

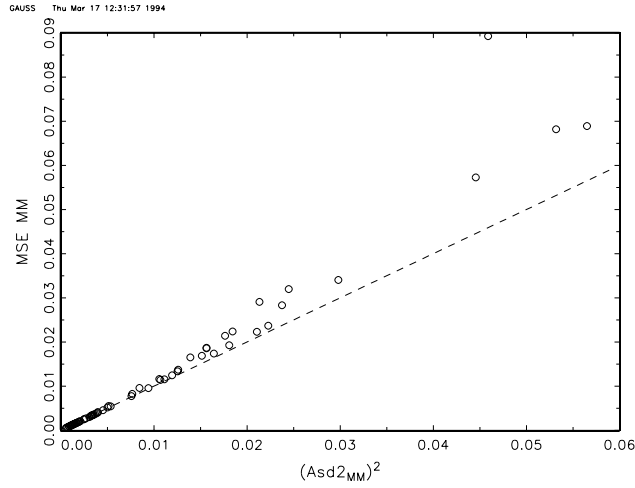


Figure 2: MSE of MM.

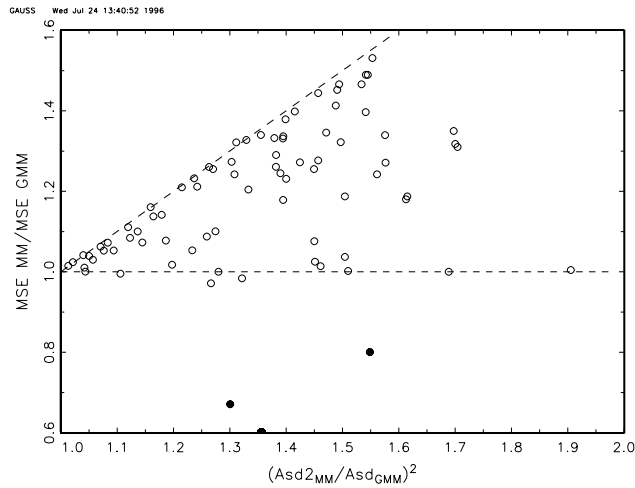


Figure 3: The ratio of MSE's of MM and MD.

Figure 3 gives a comparison of the MSE's of MM and GMM in relation to their asymptotic counterparts. That is, we plotted  $MSE_{MM}/MSE_{GMM}$  against  $(Asd2_{MM}/Asd2_{GMM})^2$ . The three differently indicated points<sup>16</sup> are cases where the fit of the asymptotic approximation is much better for MM ( $M_2 < 0.01$ ) than for GMM ( $M_2 > 0.07$ ). We find that the ratio of the MSE's is bounded by its asymptotic counterpart. Furthermore, there is a clear tendency of GMM to be more efficient than MM.

<sup>16</sup>Found for  $n = 120$ ,  $m = 30$ ,  $p = 3$ ,  $r = .1, .5, .7$ .



## 7.4 Confidence intervals.

Based on the estimates of the asymptotic variances, as described in Section 6, 95% confidence intervals have been computed both for  $\alpha = 0$  and for the alternative asymptotics, where  $\alpha$  is estimated. Let the proportion of the 20,000 replications where the true value  $\delta = 0$  was covered by these intervals be given by  $P_1$  and  $P_2$ , respectively.

TABLE 2  
The level of 95% confidence intervals.

	Mean		Minimum	
	$P_1$	$P_2$	$P_1$	$P_2$
IV	.888	-	.257	-
MM	.916	.953	.633	.943
GMM	.871	.919	.526	.627
ML	.851	.897	.404	.604

Table 2 gives the mean and the minimum over the 81 values of  $P_1$  and  $P_2$ . It shows that the confidence intervals based on an estimate of  $\alpha$  are indeed more accurate. For IV there is only one confidence interval, based on  $\alpha = 0$ , whose performance is frequently abysmally poor. The accuracy of the alternative intervals is good for MM and moderate for GMM and ML, although GMM is better than ML. However, for GMM and ML, there is a clear group of outliers. These are the 9 parameter points with group size 4. If we exclude this group, then the mean and minimum of  $P_2$  for GMM become .942 and .899, respectively. If we also exclude group size 12, then these quantities become .946 and .927, resp. So, as long as the group sizes are not too small ( $\leq 12$ ), GMM performs well.

Figure 4 gives  $P_1$  and  $P_2$  for MM against  $(Asd2_{MM}/Asd1_{MM})^2$ . It shows a systematic improvement in accuracy when the intervals are based on estimates of  $\alpha$ , instead of  $\alpha = 0$ . Although there is a slight tendency to overestimate the exact interval.

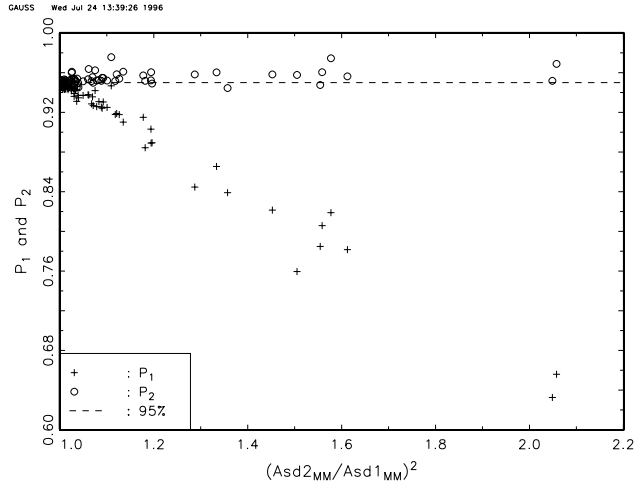


Figure 4: Confidence intervals of MM.

## Appendix

In order to derive the group-asymptotic distributions of the estimators, we first give a matrix formulation of the model. Let the  $mK \times g$  matrix  $\Pi$  have rows  $(e'_j \otimes e'_k)\Pi = \pi'_{jk}$ , where  $e_j$  and  $e_k$  are the  $j$ -th and  $k$ -th column of  $I_m$  and  $I_K$ , resp. So, we consider data conforming to

$$(y_{ijk}, x'_{ijk}) = (e'_j \otimes e'_k)\Pi(\delta, I_g) + (u_{ijk}, v'_{ijk}),$$

where the vectors of  $(u_{ijk}, v'_{ijk})$  are independent with zero means and covariance matrices  $\Omega_j$ , for  $i = 1, \dots, n_j$ ,  $j = 1, \dots, m$  and  $k = 1, \dots, K$ .

Collect the data  $(y_{ijk}, x'_{ijk}, e'_j \otimes e'_k)$  in a matrix  $(Y, Z)$ , where  $Y(0, I_g)' = X$  contains the observations  $x'_{ijk}$ . Furthermore, define the projection matrices

$$Q_j = \text{Diag}\{Z(e_j \otimes \iota_K)\},$$

$$P_j = Z(e_j e'_j \otimes I_K)Z'/n_j,$$

where  $\iota_K$  is a vector of  $K$  ones. So  $Q_j$  and  $P_j$  have ranks equal to  $Kn_j$  and  $K$ , resp. The matrices in (23) can now be written as  $M_j = Y'P_jY/n$ , and  $S_j = Y'(Q_j - P_j)Y/\{(K(n_j - 1)\}$ . We will also use  $N_j$  as given in (28):  $N_j = Y'Q_jY/n$ .

The results are based on the asymptotic properties of extremum estimators as described in Amemiya (1985). For each estimator  $\hat{\delta} = \arg \min_{\delta} (L)$ , we give  $\delta^* = \text{plim}_{\delta} \hat{\delta} = \arg \min_{\delta} (\text{plim}_{\delta} (L))$  and matrices  $A$  and  $B$  satisfying

$$A = \text{plim} \left( \frac{\partial^2 L}{\partial \delta \partial \delta'} \right) \Big|_{\delta^*}, \quad n^{1/2} \frac{\partial L}{\partial \delta} \Big|_{\delta^*} \overset{A}{\rightsquigarrow} N(0, B).$$

For the probability limits we use (25) and we notice that

$$\text{plim} (N_j) = w_j(\delta, I_g)' \overline{\pi_j \pi_j'}(\delta, I_g) + w_j \Omega_j.$$

The asymptotic normality is based on (24). We find

$$n^{1/2}(\hat{\delta} - \delta^*) \overset{A}{\rightsquigarrow} N(0, A^{-1} B A^{-1}).$$

For the computations we use the following lemma, which is a generalization of Lemma 1 in Bekker (1994), which has also been applied by Hahn and Hausman (1999). Consider the expectation and covariance matrix of vectors of the form  $x = (M+U)'C(M+U)a$ , where  $a$ ,  $M$  and  $C$  are nonstochastic;  $a$  is a vector and  $C$  is a symmetric matrix. The rows of  $U$  are i.i.d. with zero expectation and covariance matrix  $\Omega$ . Let  $u'$  be such a row and let  $d$  be the vector consisting of the diagonal elements of  $C$ .

**Lemma 1** *The expectation and variance of  $x$  are given by*

$$\begin{aligned} E(x) &= M' C M a + \text{tr}(C) \Omega a, \\ \text{Var}(x) &= a' \Omega a M' C^2 M + a' M' C^2 M a \Omega \\ &\quad + \Omega a a' M' C^2 M + M C^2 M a a' \Omega \\ &\quad + \text{tr}(C^2) a' \Omega a \Omega + \text{tr}(C^2) \Omega a a' \Omega \\ &\quad + d' d \{ E((a'u)^2 u u') - a' \Omega a \Omega - 2 \Omega a a' \Omega \} \\ &\quad + 2 d' C M a E((a'u) u u') \\ &\quad + M' C d E((a'u)^2 u') + E((a'u)^2 u) d' C M. \end{aligned}$$

*If  $U$  is Gaussian, the terms in which  $d$  appears are equal to zero.*

## A. IV

Here

$$L = 1/2 \sum_{j=1}^m \tilde{\delta}' M_j \tilde{\delta},$$

and

$$2 \text{plim } L = \tilde{\delta}' \left[ (\delta_0, I_g)' \sum_{j=1}^m (w_j \pi_j \pi_j') (\delta_0, I_g) + \alpha \bar{\Omega} \right] \tilde{\delta},$$

so  $\delta_{IV}^*$  is easily found as the one given in Section 5.1. The first derivative is given by  $\partial L / \partial \delta = -(\delta_0, I_g)' \sum_{j=1}^m M_j \tilde{\delta}$ . Under the additional condition  $n^{1/2}(K/n - \alpha/m) = o(1)$ , we find

$$n^{1/2} E \left( \frac{\partial L}{\partial \delta} \right) \Big|_{\delta_{IV}^*} = n^{1/2} \text{plim} \left( \frac{\partial L}{\partial \delta} \right) \Big|_{\delta_{IV}^*} + o(1) = o(1).$$

So,

$$n^{1/2} \frac{\partial L}{\partial \delta} \Big|_{\delta_{IV}^*} = n^{1/2} \left\{ \frac{\partial L}{\partial \delta} \Big|_{\delta_{IV}^*} - E \left( \frac{\partial L}{\partial \delta} \right) \Big|_{\delta_{IV}^*} \right\} + o(1),$$

which has asymptotic covariance matrix

$$B_{IV} = \lim_{n \rightarrow \infty} \sum_{j=1}^m (\delta_0, I_g)' \text{Var} (Y' P_j Y \tilde{\delta}_{IV}^*) (\delta_0, I_g)' / n.$$

Using Lemma 1 we find  $B_{IV} = B_1 + B_2 + B_3$  as in Section 5.1. The second derivative is given by  $\partial^2 L / \partial \delta \partial \delta' = (\delta_0, I_g)' \sum_{j=1}^m M_j (\delta_0, I_g)'$ . So

$$A_{IV} = \sum_{j=1}^m (w_j \overline{\pi_j \pi_j'}) + \alpha \bar{\Omega}_{22}.$$

## B. LIML

The group-asymptotic distribution of the LIML estimator will be derived under the assumption of homoscedasticity. Equivalent to (14), we may consider

$$L = 1/2 \log \frac{\sum_{j=1}^m \tilde{\delta}' N_j \tilde{\delta}}{\sum_{j=1}^m \tilde{\delta}' (N_j - M_j) \tilde{\delta}}.$$

As  $\text{plim}(L)$ , with  $\Omega_j = \Omega$ , is minimized by the true value  $\delta$ ,  $\hat{\delta}_{LIML}$  is consistent. The first derivative is given by

$$\begin{aligned}\frac{\partial L}{\partial \delta} &= Rx, \\ R &= \frac{(0, I_g) \sum_{j=1}^m (\tilde{\delta}' N_j \tilde{\delta} I_{g+1} - N_j \tilde{\delta} \tilde{\delta}')}{\sum_{j=1}^m (\tilde{\delta}' N_j \tilde{\delta}) \sum_{j=1}^m (\tilde{\delta}' (N_j - M_j) \tilde{\delta})}, \\ x &= \sum_{j=1}^m \{(N_j - M_j) \tilde{\delta} + a N_j \tilde{\delta}\},\end{aligned}$$

where  $a = 0$ . However, we may also choose  $a = mK/n - 1$ , so that  $E(x)|_\delta = 0$ . That is, let  $C_j = n^{-1}\{(mK/n)Q_j - P_j\}$ , then  $\text{tr} \sum_{j=1}^m C_j = 0$  and  $x = \sum_{j=1}^m Y' C_j Y \tilde{\delta}$ . So, by Lemma 1,  $x$  has zero expectation. However, the diagonal elements of the matrices  $C_j$  are equal to zero only in case of equal group sizes. We find, by application of Lemma 1

$$\begin{aligned}n^{1/2}x|_\delta &\overset{A}{\rightsquigarrow} N(0, V_x), \\ V_x &= \lim_{n \rightarrow \infty} \sum_{j=1}^m n \text{Var}(Y' C_j Y \tilde{\delta}) \\ &= \sum_{j=1}^m \left[ \sigma^2 (1 - \alpha)^2 w_j(\delta, I_g) \overline{\pi_j \pi_j'}(\delta, I_g) \right. \\ &\quad + (\alpha^2 w_j + (1 - 2\alpha)\alpha/m)(\sigma^2 \Omega + \Omega \tilde{\delta} \tilde{\delta}' \Omega) \\ &\quad + w_j \left( \alpha - \frac{\alpha}{w_j m} \right)^2 \left\{ E \left( \varepsilon^2 \begin{pmatrix} u \\ v \end{pmatrix} (u, v') \right) - \sigma^2 \Omega - 2\Omega \tilde{\delta} \tilde{\delta}' \Omega \right\} \\ &\quad + w_j (\alpha - 1) \left( \alpha - \frac{\alpha}{w_j m} \right) \left\{ (\delta, I_g)' \bar{\pi}_j E \left( \varepsilon^2 (u, v') \right) \right. \\ &\quad \quad \left. + E \left( \varepsilon^2 \begin{pmatrix} u \\ v \end{pmatrix} \right) \bar{\pi}_j'(\delta, I_g) \right\} \left. \right].\end{aligned}$$

As,

$$\text{plim}(R)|_\delta = \frac{1}{\sigma^2(1 - \alpha)} \left( (0, I_g) - \frac{\phi \tilde{\delta}'}{\sigma} \right) = D,$$

we find  $B_{LIML} = DV_x D' = \sigma^{-2}(B_1 + B_2 + B_3)$ , where  $B_1, B_2$  and  $B_3$  are given in Section

5.2. Finally, the second derivative is given by

$$\frac{\partial^2 L}{\partial \delta \partial \delta'} = (0, I_g) \left\{ \frac{\sum_{j=1}^m N_j}{\sum_{j=1}^m \tilde{\delta}' N_j \tilde{\delta}} - 2 \frac{\left( \sum_{j=1}^m N_j \tilde{\delta} \right) \left( \tilde{\delta}' \sum_{j=1}^m N_j \right)}{\left( \sum_{j=1}^m \tilde{\delta}' N_j \tilde{\delta} \right)^2} - \frac{\sum_{j=1}^m (N_j - M_j)}{\sum_{j=1}^m \tilde{\delta}' (N_j - M_j) \tilde{\delta}} + 2 \frac{\left( \sum_{j=1}^m (N_j - M_j) \tilde{\delta} \right) \left( \tilde{\delta}' \sum_{j=1}^m (N_j - M_j) \right)}{\left( \sum_{j=1}^m \tilde{\delta}' (N_j - M_j) \tilde{\delta} \right)^2} \right\} (0, I_g)'.$$

So,  $A_{LIML} = \sigma^{-2} \sum_{j=1}^m w_j \overline{\pi_j \pi_j'}$ .

### C. MM

Here

$$L = 1/2 \log \frac{\sum_{j=1}^m \tilde{\delta}' M_j \tilde{\delta}}{\sum_{j=1}^m \tilde{\delta}' S_j \tilde{\delta}}.$$

As  $\text{plim}(L)$  is minimized by the true value  $\delta$ ,  $\hat{\delta}_{MM}$  is consistent. The first derivative is given by  $\partial L / \partial \delta = Rx$ , where

$$R = (0, I_g) \frac{\left( \sum_{j=1}^m S_j \tilde{\delta} \tilde{\delta}' - \sum_{j=1}^m \tilde{\delta}' S_j \tilde{\delta} I_{g+1} \right)}{\sum_{j=1}^m (\tilde{\delta}' S_j \tilde{\delta}) \sum_{j=1}^m (\tilde{\delta}' M_j \tilde{\delta})},$$

$$x = \sum_{j=1}^m M_j \tilde{\delta} - a \sum_{j=1}^m S_j \tilde{\delta},$$

and  $a = 0$ . However,  $Rx$  is invariant under different choices for  $a$ . Therefore, we use  $a = K/n$ , so that  $E(x)|_{\delta} = 0$ . That is, let  $x = \sum_{j=1}^m x_j$  with

$$\begin{aligned} \text{(A1)} \quad x_j &= M_j \tilde{\delta} - (K/n) S_j \tilde{\delta} \\ &= n^{-1} Y' \frac{(n_j P_j - Q_j)}{n_j - 1} Y \tilde{\delta} = Y' C Y \tilde{\delta}, \end{aligned}$$

then the implicitly defined matrix  $C$  has diagonal elements equal to zero. Applying Lemma 1 we find

$$(A2) \quad n^{1/2}x_j \Big|_{\delta} \overset{A}{\rightsquigarrow} N(0, V_{x_j})$$

$$\begin{aligned} V_{x_j} &= \lim_{n \rightarrow \infty} n \text{Var}(x_j) \\ &= \sigma_j^2 w_j(\delta, I_g)' \overline{\pi_j \pi_j'}(\delta, I_g) + \frac{\alpha w_j}{m w_j - \alpha} (\sigma_j^2 \Omega_j + \Omega_j \tilde{\delta} \tilde{\delta}' \Omega_j). \end{aligned}$$

Furthermore, we have

$$\text{plim}(R)|_{\delta} = (\alpha \overline{\sigma^2})^{-2} \{(\overline{\sigma^2})(0, I_g) - (0, I_g) \overline{\Omega} \tilde{\delta} \tilde{\delta}'\} = D.$$

So

$$\begin{aligned} B_{MM} &= D \sum_{j=1}^m V_{x_j} D' \\ &= (\alpha \overline{\sigma^2})^{-2} \left[ \sum_{j=1}^m \sigma_j^2 w_j \overline{\pi_j \pi_j'} + \frac{\alpha w_j}{m w_j - \alpha} \left\{ \sigma_j^2 \Omega_{j22} - (0, I_g) \Omega_j \tilde{\delta} \tilde{\delta}' \Omega_j (0, I_g)' \right. \right. \\ &\quad \left. \left. + 2(0, I_g) \left( \Omega_j - \frac{\sigma_j^2 \overline{\Omega}}{\sigma^2} \right) \tilde{\delta} \tilde{\delta}' \left( \Omega_j - \frac{\sigma_j^2 \overline{\Omega}}{\sigma^2} \right) (0, I_g)' \right\} \right]. \end{aligned}$$

The second derivative is given by

$$\begin{aligned} \frac{\partial^2 L}{\partial \delta \partial \delta'} &= (0, I_g) \left\{ \frac{\sum_{j=1}^m M_j}{\sum_{j=1}^m \tilde{\delta}' M_j \tilde{\delta}} - 2 \frac{(\sum_{j=1}^m M_j \tilde{\delta})(\sum_{j=1}^m \tilde{\delta}' M_j)}{(\sum_{j=1}^m \tilde{\delta}' M_j \tilde{\delta})^2} \right. \\ &\quad \left. - \frac{\sum_{j=1}^m S_j}{\sum_{j=1}^m \tilde{\delta}' S_j \tilde{\delta}} + 2 \frac{(\sum_{j=1}^m S_j \tilde{\delta})(\sum_{j=1}^m \tilde{\delta}' S_j)}{(\sum_{j=1}^m \tilde{\delta}' S_j \tilde{\delta})^2} \right\} (0, I_g)'. \end{aligned}$$

So  $A_{MM} = (\alpha \overline{\sigma^2})^{-1} \sum_{j=1}^m w_j \overline{\pi_j \pi_j'}$ .

#### D. GMM

Here

$$L = 1/2 \sum_{j=1}^m \frac{\tilde{\delta}' M_j \tilde{\delta}}{\tilde{\delta}' S_j \tilde{\delta}}.$$

As  $\text{plim}(L)$  is minimized by the true value  $\delta$ ,  $\hat{\delta}_{GMM}$  is consistent. The first derivative is given by

$$\begin{aligned} \frac{\partial L}{\partial \delta} &= \sum_{j=1}^m R_j x_j, \\ R_j &= (0, I_g) \frac{(S_j \tilde{\delta} \tilde{\delta}' - \tilde{\delta}' S_j \tilde{\delta} I_{g+1})}{(\tilde{\delta}' S_j \tilde{\delta})^2}, \\ x_j &= M_j \tilde{\delta} - a S_j \tilde{\delta}, \end{aligned}$$

where  $a = 0$ . However, we may also use  $a = K/n$  so that  $x_j$  is the same as in (A1).

Furthermore,

$$(A3) \quad \text{plim}(R_j)|_\delta = (0, I_g) \frac{(\Omega_j \tilde{\delta} \tilde{\delta}' - \sigma_j^2 I_{g+1})}{\sigma_j^4} = D_j.$$

So

$$\begin{aligned} B_{GMM} &= \sum_{j=1}^m D_j V_{x_j} D_j' \\ &= \sum_{j=1}^m \sigma_j^{-2} \left\{ w_j \overline{\pi_j \pi_j'} + \frac{\alpha w_j}{m w_j - \alpha} \left( \Omega_{j22} - (0, I_g) \frac{\Omega_j \tilde{\delta} \tilde{\delta}' \Omega_j}{\sigma_j^2} (0, I_g)' \right) \right\}. \end{aligned}$$

The second derivative, which can be used in a Newton-Raphson procedure, is given by

$$\begin{aligned} \frac{\partial^2 L}{\partial \delta \partial \delta'} &= (0, I_g) \sum_{j=1}^m \left\{ \frac{M_j}{\tilde{\delta}' S_j \tilde{\delta}} - 2 \frac{M_j \tilde{\delta} \tilde{\delta}' S_j}{(\tilde{\delta}' S_j \tilde{\delta})^2} - 2 \frac{S_j \tilde{\delta} \tilde{\delta}' M_j}{(\tilde{\delta}' S_j \tilde{\delta})^2} - \frac{(\tilde{\delta}' M_j \tilde{\delta}) S_j}{(\tilde{\delta}' S_j \tilde{\delta})^2} \right. \\ &\quad \left. + 4 \frac{(\tilde{\delta}' M_j \tilde{\delta}) S_j \tilde{\delta} \tilde{\delta}' S_j}{(\tilde{\delta}' S_j \tilde{\delta})^3} \right\} (0, I_g)'. \end{aligned}$$

So,  $A_{GMM} = \sum_{j=1}^m \sigma_j^{-2} w_j \overline{\pi_j \pi_j'}$ .

### E. ML

Here

$$L = 1/2 \sum_{j=1}^m w_j \log \frac{\tilde{\delta}' N_j \tilde{\delta}}{\tilde{\delta}' S_j \tilde{\delta}}.$$

As  $\text{plim}(L)$  is minimized by the true value  $\delta$ ,  $\hat{\delta}_{ML1}$  is consistent. The first derivative is given by

$$\begin{aligned} \frac{\partial L}{\partial \delta} &= \sum_{j=1}^m R_j x_j, \\ R_j &= (0, I_g) \frac{(N_j \tilde{\delta} \tilde{\delta}' - \tilde{\delta}' N_j \tilde{\delta} I_{g+1})}{\tilde{\delta}' N_j \tilde{\delta} \tilde{\delta}' S_j \tilde{\delta}}, \\ x_j &= a N_j \tilde{\delta} - w_j S_j \tilde{\delta}, \end{aligned}$$

where  $a = 0$ . However, we may also choose  $a = 1$ , so that  $x_j$  is the same as in (A1). As

$\text{plim}(R_j)|_\delta$  is as in (A3), we find  $B_{ML} = B_{GMM}$ . The second derivative is given by

$$\frac{\partial^2 L}{\partial \delta \partial \delta'} = (0, I_g) \sum_{j=1}^m w_j \left\{ \frac{N_j}{\tilde{\delta}' N_j \tilde{\delta}} - 2 \frac{N_j \tilde{\delta} \tilde{\delta}' N_j}{(\tilde{\delta}' N_j \tilde{\delta})^2} - \frac{S_j}{\tilde{\delta}' S_j \tilde{\delta}} + 2 \frac{S_j \tilde{\delta} \tilde{\delta}' S_j}{(\tilde{\delta}' S_j \tilde{\delta})^2} \right\} (0, I_g)'.$$

So,  $A_{ML} = A_{GMM}$ .



## REFERENCES

- Aigner, D.J., C. Hsiao, A. Kapteyn and T.J. Wansbeek (1984): "Latent Variable Models in Econometrics", in *Handbook of Econometrics*, Vol. 2, ed. by Z. Griliches and M.D. Intriligator, Amsterdam: North-Holland.
- Anderson, T.W. (1950): "Estimation of the Parameters of a Single Equation by the Limited-Information-Maximum-Likelihood Method" *Statistical Inference in Dynamic Economic Models*, ed. by T.C. Koopmans, New York: Wiley.
- Anderson, T.W. (1976): "Estimation of Linear Functional Relationships: Approximate Distributions and Connections with Simultaneous Equations in Econometrics", *Journal of the Royal Statistical Society, Series B*, 38, 1-36
- Anderson, T.W. (1984): "Estimating Linear Functional Relationships", *Annals of Statistics*, 12, 1-45
- Anderson, T.W., and H. Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations", *The Annals of Mathematical Statistics*, 20, 46-63.
- Anderson, T.W., N. Kunitomo and T. Sawa (1982): "Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator", *Econometrica*, 50, 1009-1027.
- Angrist, J.D. (1990): "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records", *The American Economic Review*, 80, 313-336.
- Angrist, J.D. (1991): "Grouped-data Estimation and Testing in Simple Labor-Supply Models", *Journal of Econometrics*, 47, 243-266.
- Angrist, J.D., and A.B. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?", *Quarterly Journal of Economics*, 106, 979-1014.
- 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", *Quarterly Journal of Economics*, 56, 979-1014.
- Angrist, J.D., and A.B. Krueger (1995): "Split Sample Instrumental Variables Estimates of the Return to Schooling", *Journal of Business and Economic Statistics*, 13, 225-235.
- Amemiya, T. (1985): *Advanced Econometrics*. Cambridge: Harvard University Press.

- Bekker, P.A. (1994): "Alternative Approximations to the Distributions of Instrumental Variable Estimators", *Econometrica*, 62, 657-681.
- Bound, J., D.A. Jaeger, and R.M. Baker (1995): "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Explanatory Variable is Weak", *Journal of the American Statistical Society*, 90, 443-450.
- Bowden, R.J. and D.A. Turkington (1984): *Instrumental Variables*, Cambridge, U.K.: Cambridge University Press.
- Buse, A. (1992): "The Bias of Instrumental Variable Estimators", *Econometrica*, 60, 173-180.
- Deaton, A. (1985): "Panel Data from Time Series of Cross-Sections", *Journal of Econometrics*, 30, 109-126.
- Goldberger, A.S. and I. Olkin (1971): "A Minimum-Distance Interpretation of Limited-Information Estimation", *Econometrica*, 39, 635-939.
- Hahn, J., and J. Hausman (1999): "A New Specification Test for the Validity of Instrumental Variables", unpublished manuscript, MIT.
- Hahn, J., and A. Inoue (1999): "A Monte Carlo Comparison of Various Asymptotic Approximations to the Distribution of Instrumental Variables Estimators", unpublished manuscript, University of Pennsylvania.
- Hall, A.R., G.D. Rudebusch, and D.W. Wilcox (1996): "Judging Instrument Relevance in Instrumental Variables Estimation", *International Economic Review*, 37, 283-289.
- Kendall, M.G. and A. Stuart (1979): *The Advanced Theory of Statistics*, Vol. 2, Fourth Edition. New York: Macmillan.
- Koenker, R., and J.A.F. Machado (1999): "GMM Inference when the Number of Moment Conditions is Large", *Journal of Econometrics*, 93, 327-344.
- Kunitomo, N. (1980): "Asymptotic Expansions of the Distributions of Estimators in a Linear Functional Relationship and Simultaneous Equations", *Journal of the American Statistical Society*, 75, 693-700.
- Kunitomo, N. (1986): "Comparing Some Modified Maximum Likelihood Estimators of a Slope Coefficient in a Linear Functional Relationship", *J. Japan Statist. Soc.* 16, 173-185.
- Kunitomo, N. (1987): "A Third Order Optimum Property of the ML Estimator in a Linear Functional Relationship and Simultaneous Equation System in Econometrics",

*Annals of the Institute of Statistical Mathematics* 39, 575-591.

- Levitt, S.D. (1997): "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime", *The American Economic Review*, 87, 270-290.
- Lewbel, A. (1996): "Demand Estimation with Expenditure Measurement Errors on the Left and Right Hand Side", *The Review of Economics and Statistics*, 78, 718-725.
- Madansky, A. (1959): "The Fitting of Straight Lines when both Variables are Subject to Error", *Journal of the American Statistical Association*, 54, 193-205.
- Maddala, G.S. and J. Jeong (1992): "On the Exact Small Sample Distribution of the Instrumental Variable Estimator", *Econometrica* 60, 181-183.
- Mariano, R.S. (1982): "Analytical Small-Sample Distribution Theory in Econometrics: The Simultaneous-Equations Case", *International Economic Review*, 23, 503-533.
- Moran, P.A.P. (1971): "Estimating Structural and Functional Relationships", *Journal of Multivariate Analysis*, 1, 232-255.
- Morimune, K. (1983): "Approximate Distributions of  $k$ -Class Estimators when the Degree of Overidentification is Large Compared with the Sample Size", *Econometrica*, 51, 821-842.
- Nelson, C.R., and R. Startz (1990a): "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator", *Econometrica*, 58, 967-976.
- Nelson, C.R., and R. Startz (1990b): "The Distribution of the Instrumental Variables Estimator and Its  $t$ -ratio when the Instrument is a Poor One", *Journal of Business*, 63, S125-S140.
- Neyman, J., and E. Scott (1948): "Consistent Estimates Based on Partially Consistent Observations", *Econometrica*, 16, 1-32.
- Pagan, A. (1979): "Some Consequences of Viewing LIML as an Iterated Aitken Estimator", *Economics Letters*, 3, 369-372.
- Phillips, P.C.B. (1983): "Exact Small Sample Theory in the Simultaneous Equations Model", in *Handbook of Econometrics*, Vol. I, ed. by M.D. Intriligator, and Z. Griliches. Amsterdam: North-Holland.
- Sen, P.K. and J.M. Singer (1993): *Large Sample Methods in Statistics*, Chapman and Hall, London.
- Shea, J. (1997): "Instrument Relevance in Multivariate Linear Models: A Simple Measure", *The Review of Economics and Statistics*, 79, 348-352.

- Staiger, D. and J.H. Stock (1997): “Instrumental Variables Regression with Weak Instruments”, *Econometrica*, 65, 557-586.
- Startz, R., C.R. Nelson, and E. Zivot (1999): “Improved Inference for the Instrumental Variable Estimator”, unpublished manuscript, University of Washington.
- Verbeek, M., and T. Nijman (1993): “Minimum MSE Estimation of a Regression Model with Fixed Effects from a Series of Cross-Sections”, *Journal of Econometrics*, 59, 125-136.
- Wang, J., and E. Zivot (1998): “Inference on Structural Parameters in Instrumental Variables Regressions with Weak Instruments”, *Econometrica*, 66, 1389-1404.
- Zivot, E., R. Startz, and C.R. Nelson (1998): “Valid Confidence Intervals and Inference in the Presence of Weak Instruments”, *International Economic Review*, 39, 1119-1144.
-