Chapter 4

Instrumental variable estimation in the presence of weak instruments

4.1 Introduction

In the literature on instrumental variable (IV) estimation much attention is presently given to weak instruments, i.e. instruments that are only weakly correlated with the explanatory variables. For example, Nelson and Startz (1990 a,b) examine the behavior of the two-stage least squares (TSLS) estimator and show that the large sample asymptotic approximation to the finite sample distribution of TSLS breaks down in such circumstances. Similar results have been described by Buse (1992), Choi and Phillips (1992), Staiger and Stock (1994) and Bound, Jaeger and Baker (1995).

Mariano (1977) makes finite sample comparisons of the TSLS estimator and estimators of the same form but based on a subset of the instruments. Although TSLS is large sample asymptotically efficient, Mariano’s finite sample results indicate that TSLS should not be considered as efficient. This chapter presents a similar analysis for the limited information maximum likelihood (LIML) estimator and its extension to the case where the first-stage regression is nonlinear: the method of moments (MM) estimator described in Bekker (1994). That is, we assume
normality and compare the MM estimator with an MM estimator where part of the instruments are (partially) disregarded. Special attention will be given to the case where the disregarded instruments can be considered as being weak.

These comparisons will be based on asymptotic distributions which are derived for a parameter sequence in which the number of instruments increases along with the sample size. Working in the context of dummy instruments in Chapters 2 and 3, these sequences have been called group-asymptotics. The same name is adopted here, although the analysis in the present chapter is not restricted to cases where the instruments are dummy variables. Both Bekker (1994) and the results given in Chapter 2 show that the group-asymptotic approach provides approximations which are closer to the finite sample distributions of IV estimators when compared to large sample asymptotics. In this chapter both group-asymptotics and large sample asymptotics will be considered. It will be shown that both approaches imply rather different results with respect to the relative efficiency of MM-type estimators.

In particular, if the first-stage regression is linear the MM estimator is the maximum likelihood estimator known as LIML. In this case the comparisons of the distributions of the estimators can be based on known results about the distribution of the LIML estimator. From these results we conclude that the LIML estimator should not be considered as efficient. A closely related result can be found in Chapter 3, where it is shown that the LIML estimator does not reach the group-asymptotic efficiency bound.

This chapter is organized as follows. Section 4.2 introduces the model without making the assumption that the first-stage regression is linear. In Section 4.3 the MM estimator is described and the group-asymptotic and large sample asymptotic approximation of its distribution is derived. These approximations are used in Section 4.4 to compare the MM estimator with the MM estimator where some of the instruments are disregarded. Section 4.5 describes a Monte Carlo experiment where the small sample performance of the two estimators is related to approximations of the relative efficiency based on group-asymptotics and large sample asymptotics.
Chapter 4. IV Estimation in the presence of weak instruments

4.2 The model

Consider a single structural equation represented by

\[ y = X\delta + \varepsilon, \]  

(4.1)

where \( y : n \times 1 \) and \( X : n \times g \) are observed. However, it is not assumed that the elements of \( E(\varepsilon|X) \) are equal, which would lead to linear regression. So the matrix of explanatory variables may contain endogenous variables that are correlated with the \( n \times 1 \) vector of disturbances \( \varepsilon \). Aigner et al. (1984) argue that in such cases the use of additional information seems almost indispensable to make progress in practical problems.

Further information is given by observations on instrumental variables contained in a matrix \( Z : n \times K \), which is of rank \( K \). All stochastics will be conditional on \( Z \): the rows of \((y,X)\) are assumed to be independently normally distributed with identical covariance matrix \( \Omega \) and expectations satisfying \( E(y) = E(X)\delta \), where \( E(X) \) is of rank \( g \). Let \( P_Z = Z(Z'Z)^{-1}Z' \), then it is finally assumed that

\[ \frac{E(X)'P_Z E(X)}{K} > \frac{E(X)'E(X)}{n}, \]  

(4.2)

which says that the instruments do a better job in explaining \( E(X) \) than do arbitrary variables (see Bekker, 1994).

Without loss of generality, let \( E(X) = Z\Pi + R \), where \( Z\Pi = P_Z E(X) \) and \( R = (I - P_Z) E(X) \), so that \( Z'R = 0 \). In reduced form, the model can be reformulated as

\[ Y = (y,X) = Z\Pi(\delta,I) + R(\delta,I) + (u,V), \]  

(4.3)

where the rows of \((u,V)\) are independently normally distributed with zero mean and covariance matrix \( \Omega \). Note that by post-multiplying \( Z \) by the matrix \( C = (Z'Z)^{-1/2} \), \( Z \) is transformed into an isometric matrix. If, at the same time, \( \Pi \) is pre-multiplied by \( C^{-1} \), the distribution of \( Y \) remains the same. So without loss of generality it may be assumed that \( Z \) satisfies \( Z'Z = I_K \). Inequality (4.2) then becomes

\[ \Pi'\Pi > \frac{K}{n - K} R'R. \]  

(4.4)
4.3 Method of moments estimators

All the estimators that will be considered in this chapter have a method of moments interpretation, and their distributions are members of a parameterized family of distributions. It is convenient to introduce this family rather abstractly and to make the connection with the model and some of the estimators of \( \delta \) later. For that purpose consider two stochastically independent \((g + 1) \times (g + 1)\) p.s.d. matrices \( \bar{S} \) and \( S^\perp \), with (non)-central Wishart distributions given by

\[
\bar{S} \sim W \left( \Omega, \ p, (\delta, I_g)' \bar{M} (\delta, I_g) \right), \\
S^\perp \sim W \left( \Omega, \ q, (\delta, I_g)' M^\perp (\delta, I_g) \right).
\]  

(4.5)

The notation is adopted from Eaton (1983, definition 8.2), and has been explained in Chapter 1. For convenience the order of the distribution, i.e. \( g + 1 \), is suppressed.

Based on these statistics one can formulate an estimator of the parameter vector of interest \( \delta \) by considering the expectations of \( \bar{S} \) and \( S^\perp \) (Eaton, 1982, p.317):

\[
E(\bar{S}) = (\delta, I_g)' \bar{M} (\delta, I_g) + p\Omega, \\
E(S^\perp) = (\delta, I_g)' M^\perp (\delta, I_g) + q\Omega.
\]  

(4.6)

If the p.s.d. matrices \( \bar{M} \) and \( M^\perp \) are ordered according to

\[
\bar{M} > \frac{p}{q} M^\perp,
\]  

(4.7)
the parameter $\delta$ can be solved from the moment restrictions given in (4.6). That is, a solution of $\delta$ is given implicitly by

$$\left( E(\bar{S}) - \lambda E(S^\perp) \right) \begin{pmatrix} 1 \\ -\delta \end{pmatrix} = 0,$$

(4.8)

where $\lambda = p/q$, which is the smallest value of $\lambda$ for which $E(\bar{S}) - \lambda E(S^\perp)$ is singular. So, a method of moments (MM) estimator is found by solving the sample analogue

$$\left( \bar{S} - \hat{\lambda}S^\perp \right) \begin{pmatrix} 1 \\ -\hat{\delta} \end{pmatrix} = 0,$$

(4.9)

where $\hat{\lambda}$ is minimal.

Since $\bar{S}$ and $S^\perp$ are independent, the distribution of $\hat{\delta}$ can be characterized by the parameters of the Wishart distributions given in (4.5). For parameter points that satisfy (4.7), the following representation of the distribution of $\hat{\delta}$ is used:

$$\hat{\delta} \sim F_{\text{MM}}(\Omega, \delta; \bar{M}, p; M^\perp, q),$$

(4.10)

which will henceforward be referred to as an MM distribution.

In the model the statistics $\bar{S}$ and $S^\perp$ may have the form $\bar{S} = Y'PY$, and $S^\perp = Y'QY$, where $P$ is a projection matrix of rank $p$, say, and $Q$ a projection matrix of rank $q$. These statistics are independent if $PQ = QP = 0$ and it follows from proposition 8.13, Eaton 1983, that the distributions of $\bar{S}$ and $S^\perp$ are given in (4.5), where

$$\bar{M} = (ZII + R)'P(ZII + R),$$

$$M^\perp = (ZII + R)'Q(ZII + R).$$

In general, without further restrictions on $P$ and $Q$, condition (4.7) will not be satisfied. This condition is satisfied if $P = P_Z$ and $Q = I - P_Z$. For this choice, $p = K$, $q = n - K$, $\bar{M} = \Pi\Pi$ and $M^\perp = \Pi'\Pi$ and condition (4.7) is equivalent to (4.4). Now (4.9) provides a meaningful MM estimator $\hat{\delta}_F$:

$$\left( Y'P_ZY - \hat{\lambda}_F Y'(I - P_Z)Y \right) \begin{pmatrix} 1 \\ -\hat{\delta}_F \end{pmatrix} = 0,$$

(4.11)
where \( \hat{\lambda}_F \) is minimal. This estimator has been previously described by Bekker (1994). In case \( R = 0 \), \( \hat{\delta}_F \) is the maximum likelihood (ML) estimator known as LIML.

### 4.3.1 The MM distribution

The MM estimator \( \hat{\delta}_F \) will be compared with another MM estimator where the choice of the projection matrix \( P \) is different from \( P_Z \). To make this comparison there is no need to develop new theory concerning the distribution of these estimators. Indeed, any statistic \( \hat{\delta} \) whose distribution is given by \( F_{MM}(\Omega, \delta; \bar{M}, p; M^\perp, q) \) can, from a distributional point of view, be identified with an estimator \( \hat{\delta}_F \) in a model given by (4.3). That is, let \( Z \) be some isometric \((p + q) \times p\) matrix and let \( R \) and \( \Pi \) be such that \( R'R = M^\perp \) and \( \Pi'\Pi = \bar{M} \). Then, with this choice of the parameters, the distribution of \( \hat{\delta}_F \) is the same as that of \( \hat{\delta} \). So, the existing statements in the literature on the distribution of \( \hat{\delta}_F \) can be used to evaluate the MM distribution.

In particular, if \( M^\perp = 0 \), the MM distribution coincides with the distribution of the LIML estimator. This estimator has been intensively studied in the context of simultaneous equations, using various approaches. Phillips (1984,1985) derived an exact analytical expression for this distribution, but it is usually too complicated to permit meaningful general conclusions. Another approach has been to approximate the distribution by one or more terms in an asymptotic expansion. Such expansions have been given for a variety of parameter sequences; see, for example, Anderson (1977). Tabulations of an approximate distribution of the LIML estimator can be found in Anderson, Kunitomo and Sawa (1982).

Bekker (1994) derived the asymptotic distribution of \( \hat{\delta}_F \), without the assumption that \( R = 0 \), under asymptotic sequences where the number of instruments, \( K \), increases along with the sample size. In Chapters 2 and 3 it is shown that the incidental parameters involved in such sequences may affect the consistency and efficiency of IV estimators. In particular, the LIML estimator was found to be inefficient under group-asymptotics. In this chapter we show that the same result holds true for general MM estimators of the form \( \hat{\delta}_F \).
4.3.2 An asymptotic approximation of the MM distribution

To find an approximation to the distribution $F_{\text{MM}}$, consider Bekker’s (1994) formulation of the group-asymptotic sequences. Let $\tau = p + q - g$, then, as $\tau \to \infty$, consider a sequence for which

\[
p/\tau \to \alpha < 1, \quad \tau^{-1} M \to \bar{H}, \quad \tau^{-1} M^\perp \to H^\perp, \tag{4.12}\]

where $\bar{H}$ and $H^\perp$ satisfy, similar to (4.7),

\[
\bar{H} > \frac{\alpha}{1 - \alpha} H^\perp.
\]

Using the result derived by Bekker (1994) we find that the group-asymptotic distribution of an estimator $\hat{\delta} \sim F_{\text{MM}} (\Omega, \delta; \bar{M}, p; M^\perp, q)$ is given by

\[
\sqrt{\tau}(\hat{\delta} - \delta) \sim N \left( 0, \sigma^2 B^{-1} C B^{-1} \right), \tag{4.13}
\]

where

\[
B = \bar{H} - \frac{\alpha}{1 - \alpha} H^\perp, \quad C = \bar{H} + \frac{\alpha}{1 - \alpha} (\Sigma_{22} - \phi \phi') + \left( \frac{\alpha}{1 - \alpha} \right)^2 H^\perp, \]

and

\[
\Sigma = \begin{pmatrix}
\sigma^2 & \sigma \phi \\
\sigma \phi' & \Sigma_{22}
\end{pmatrix} = \begin{pmatrix}
1 & -\delta' \\
0 & I_g
\end{pmatrix} \Omega \begin{pmatrix}
1 & 0 \\
-\delta & I_g
\end{pmatrix}.
\]

For an approximation to the finite sample distribution of $\hat{\delta}$, we use

\[
\bar{H} = \bar{M}/(p + q - g), \quad H^\perp = M^\perp/(p + q - g),
\]

and $\alpha^* = (p - g)/(p + q - g)$. There is good reason for choosing $\alpha$ in this way, rather than $\alpha = p/(p + q - g)$. If $p = g$ the matrix $\bar{S}$ is singular so $\lambda = 0$. In that case the distribution of the MM estimator does not
depend on the distribution of the statistic $S^\perp$. Choosing $\alpha = \alpha^*$ implies the same holds true for the approximation.

Hence the MM distribution $F_{MM}(\Omega, \delta; \bar{M}, p; M^\perp, q)$ is approximated by a normal distribution with mean $\delta$ and covariance matrix

$$V^* = \sigma^2 \left( \bar{M} - \frac{p-g}{q} M^\perp \right)^{-1} \times$$

$$\left( \bar{M} + \frac{p-g}{q} (p+q-g)(\Sigma_{22} - \phi \phi') + \left( \frac{p-g}{q} \right)^2 M^\perp \right) \times$$

$$\left( \bar{M} - \frac{p-g}{q} M^\perp \right)^{-1}.$$  \hfill (4.14)

Large sample asymptotics is described by the sequences (4.12) where $\alpha = 0$. Hence, the large sample approximation of $F_{MM}(\Omega, \delta; \bar{M}, p; M^\perp, q)$ is a normal distribution with mean $\delta$ and covariance matrix

$$V^0 = \sigma^2 \bar{M}^{-1}.$$  \hfill (4.15)

### 4.4 A comparison of some MM estimators

Recently, considerable attention has been given to instruments that are only weakly correlated with the explanatory variables, i.e. weak instruments. Staiger and Stock (1994) model the matrix of first stage regression coefficients $\Pi$ as being in a small neighborhood of zero and call it the "weakly correlated" case. Here situations are considered where some of the instruments are weak.

In order to give a more precise meaning to the phrase that some of the instruments are weak consider a decomposition of the projection matrix $P_Z$:

$$P_Z = P_1 + P_2,$$

$$P_1 P_2 = P_2 P_1 = 0,$$  \hfill (4.16)

where the rank of $P_i$ is $K_i$, $i = 1, 2$. Assume furthermore that $K_1 \geq g$ and $K_2 \geq 1$, so that the model is overidentified.
\[ K = K_1 + K_2 > g. \]

Such a decomposition of \( P_Z \) can be thought of as the result of a partitioning of the instruments \( Z \). That is, \( P_i \) can be written as \( P_i = Z_i Z_i' \), where \( Z_i \) is an \( n \times K_i \) matrix of full column rank and \( (Z_1, Z_2) = Z O \) for some orthogonal matrix \( O \). Model (4.3) may now be reformulated as

\[
Y = (Z_1, Z_2) \begin{pmatrix} \Pi_1 \\ \Pi_2 \end{pmatrix} (\delta, I_g) + R(\delta, I_g) + (u, V), \tag{4.17}
\]

where \( \Pi_1 \) is \( K_1 \times g \), \( \Pi_2 \) is \( K_2 \times g \) and \( (\Pi_1', \Pi_2')' = O^{-1} \Pi \). The instruments \( Z_2 \) are said to be weak if \( \Pi_2 \) is located in a small neighborhood of zero.

In the next subsection an MM estimator will be formulated where the instruments \( Z_2 \) are disregarded. This estimator is compared with the estimator \( \hat{\delta}_F \) using the group and large sample asymptotic approximation of \( F_{MM} \), which has been given in Subsection 4.3.2. Such a comparison is most easy to make if \( \Omega \) is nonsingular. This case will be considered in Subsection 4.4.1. The singular case, where some of the explanatory variables are exogenous, will be considered in Subsection 4.4.3.

### 4.4.1 An alternative MM estimator

Suppose that in addition to (4.4) the following restriction holds:

\[
E(X)'P_1 E(X) = \Pi_1' \Pi_1 > \frac{K_1}{n - K} R'R. \tag{4.18}
\]

Then \( \delta \) can also be estimated by the MM estimator \( \hat{\delta}_R \) which solves

\[
\begin{pmatrix} Y'P_1 Y - \hat{\lambda}_R (I - P_Z)Y \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ \hat{\delta}_R \end{pmatrix} = 0, \tag{4.19}
\]

for \( \hat{\lambda}_R \) minimal. This estimator is different from the estimator \( \hat{\delta}_F \), as given in (4.11), in the sense that for \( \hat{\delta}_F \), \( \bar{S} = Y'P_2 Y = Y'P_1 Y + Y'P_2 Y \), whereas for \( \hat{\delta}_R \), \( \bar{S} = Y'P_1 Y \). So, the latter estimator completely disregards the statistic \( Y'P_2 Y \), which is independent from \( Y'P_1 Y \) and \( Y'(I - P_2)Y \).

For a comparison of the estimators \( \hat{\delta}_F \) and \( \hat{\delta}_R \) consider their finite sample distributions, which are given by
\[ \hat{\delta}_F \sim F_{MM} (\Omega, \delta; \Pi'\Pi, K; R'R, n - K), \]
\[ \hat{\delta}_R \sim F_{MM} (\Omega, \delta; \Pi'_1\Pi_1, K_1; R'R, n - K). \]

The distributions only differ with respect to the parameters \( p \) and \( \bar{M} \). Based on the large sample asymptotic approximations of these distributions, one is led to believe that \( \hat{\delta}_F \) is always to be preferred to \( \hat{\delta}_R \). Indeed, the large sample asymptotic covariance matrix (4.15) of \( \hat{\delta}_F \) is given by
\[ V_0^F = \sigma^2 (\Pi'\Pi)^{-1}, \]
whereas the large sample asymptotic covariance matrix of \( \hat{\delta}_R \) is given by
\[ V_0^R = \sigma^2 (\Pi'_1\Pi_1)^{-1}. \]

Now since
\[ \Pi'\Pi = \Pi'_1\Pi_1 + \Pi'_2\Pi_2 \geq \Pi'_1\Pi_1, \]
it follows that \( V_0^F < V_0^R \) if \( \Pi'_2\Pi_2 > 0 \). In a sense this fits the intuition that ignoring statistics that contain information about \( \delta \) is not a good idea if \( \delta \) is to be estimated. Moreover, if \( R = 0 \), \( V_0^F \) is the inverse of the Fisher information matrix. So \( \hat{\delta}_F \) is (large sample) asymptotically efficient, whereas \( \hat{\delta}_R \) reaches this bound if and only if \( \Pi_2 = 0 \).

Evaluations based on group-asymptotics tell a very different story. The group-asymptotic covariance matrix (4.14) of \( \hat{\delta}_F \) is given by
\[ V_*^F = \sigma^2 \left( \Pi'\Pi - \frac{K - g}{n - K} R'R \right)^{-1} \times \]
\[ \left( \Pi'\Pi + \frac{K - g}{n - K} (n - g)(\Sigma_{22} - \phi\phi') + \left( \frac{K - g}{n - K} \right)^2 R'R \right) \times \]
\[ \left( \Pi'\Pi - \frac{K - g}{n - K} R'R \right)^{-1}. \]

The group-asymptotic covariance matrix of \( \hat{\delta}_R \) is given by
Chapter 4. IV Estimation in the presence of weak instruments

\[ V^*_R = \sigma^2 \left( \Pi_1' \Pi_1 - \frac{K_1 - g}{n-K} R'R \right)^{-1} \times \]

\[ \left( \Pi_1' \Pi_1 + \frac{K_1 - g}{n-K} (n-K_2 - g)(\Sigma_{22} - \phi \phi') + \left( \frac{K_1 - g}{n-K} \right)^2 R'R \right) \times \]

\[ \left( \Pi_1' \Pi_1 - \frac{K_1 - g}{n-K} R'R \right)^{-1}. \]  

(4.22)

Now consider the case where \( R = 0 \) and \( \Pi_2 = 0 \). The difference between \( V^*_F \) and \( V^*_R \) is then given by

\[ V^*_F - V^*_R = l\sigma^2 \left( \Pi_1' \Pi_1 \right)^{-1} (\Sigma_{22} - \phi \phi') \left( \Pi_1' \Pi_1 \right)^{-1}, \]  

(4.23)

where

\[ l = \frac{(K-g)(n-g) - (K_1 - g)(n-K_2 - g)}{n-K} > \frac{(n-g)K_2}{n-K} > 0. \]

As the nonsingularity of \( \Omega \) implies the nonsingularity of \( \Sigma_{22} - \phi \phi' \), the difference given in (4.23) is positive definite. Moreover, since \( V^*_F - V^*_R \) is continuous in its arguments, it remains positive definite if \( R \) and \( \Pi_2 \) are unequal to zero but small. Contrary to evaluations based on large sample asymptotics, the group-asymptotic approximations suggest that the spread of the distribution of \( \hat{\delta}_R \) may be smaller than the spread of the distribution of \( \hat{\delta}_F \) if the instruments \( Z_2 \) are weak, i.e. if \( \Pi_2 \) is small.

This makes sense, which can be shown by considering the expectations of \( Y'P_2Y \) and \( Y'P_1Y \):

\[ E(Y'P_2Y) = (\delta, I_g)' \Pi \Pi (\delta, I_g) + K\Omega, \]

\[ E(Y'P_1Y) = (\delta, I_g)' \Pi_1' \Pi_1 (\delta, I_g) + K_1\Omega. \]

The structural part of \( E(Y'P_2Y) \) is, in general, larger than the structural part of \( E(Y'P_1Y) \): \( \Pi'\Pi \geq \Pi_1'\Pi_1 \). On the other hand, the error part of \( E(Y'P_2Y) \) is also larger than the error part of \( E(Y'P_1Y) \): \( K > K_1 \). In case \( \Pi_2 = 0 \), a projection of the observations on \( Z \) only captures more noise when compared to projecting the observations on \( Z_1 \). This additional error in the estimator \( \hat{\delta}_F \) explains the ordering \( V^*_F > V^*_R \). If \( \Pi_2 \)
is not equal to zero it may still be too small to outweigh this additional error.

4.4.2 The tables of Anderson, Kunitomo and Sawa

If \( R = 0 \), the possible efficiency gain of \( \hat{\delta}_R \) over \( \hat{\delta}_F \) could possibly have been inferred from results in the literature. For \( g = 1 \) Anderson, Kunitomo and Sawa (1982) give tables of an approximate distribution of LIML. That is, tables of an approximation of the MM distribution where \( M^\perp = 0 \):

\[
F_{MM}(\Omega, \delta; \bar{M}, p; 0, q), \quad (4.24)
\]

The tables contain, what they expect to be ",.enough values of the parameters to cover all cases of interest." Based on these tables they conclude that the spread of the distribution of LIML (measured in terms of percentiles) is decreasing in \( \bar{M} \) (or \( \Pi'\Pi \)) but increasing in \( p \) (or \( K \): the number of instruments). This observation is in accordance with a second order approximation of the mean squared error of the LIML distribution, reported in the same paper for \( g = 1 \) and for the multivariate case in Rothenberg (1983).

These observations indeed suggest that the distribution of \( \hat{\delta}_R \) may be less spread out than the distribution of the LIML estimator \( \hat{\delta}_F \). Comparing the parameters of the finite sample distributions of \( \hat{\delta}_F \) and \( \hat{\delta}_R \), as given in (4.20) where \( R = 0 \), one finds that

\[
\Pi'_1\Pi_1 \leq \Pi'\Pi.
\]

At the same time, however,

\[
K_1 \leq K.
\]

The net effect may be that the spread of \( \hat{\delta}_R \) is smaller than the spread of \( \hat{\delta}_F \).

Unfortunately, the selected parameter points in Anderson et al. (1982) exclude cases where for some choice of the parameters, \( p_0 \) and \( M_0 \) say, the spread of the distribution is smaller than the spread for another choice, \( p \) and \( M \) say, where \( M > M_0 \) and \( p > p_0 \). An evaluation of such cases is interesting because of their relation with the relative efficiency.
of $\hat{\delta}_R$ and $\hat{\delta}_F$. Using Monte Carlo approximations we describe such cases in Section 4.5.

4.4.3 Exogenous explanatory variables

If some of the explanatory variables are exogenous, $\Omega$ is a singular matrix. This complicates the comparison between the group-asymptotic covariances of $\hat{\delta}_F$ and $\hat{\delta}_R$, as described in the previous subsection. The difference given in (4.23) between $V^*_F$ and $V^*_R$ for parameter points where $R = 0$ and $\Pi_2 = 0$ is now positive semidefinite, instead of positive definite. This does not necessarily mean that $V^*_F - V^*_R$ remains positive semidefinite if $R$ and $\Pi_2$ are not equal to zero but small. An extension of the results given in the previous subsection to the singular case is developed here.

Consider an equation with endogenous variables $X : n \times g$ and exogenous variables collected in the matrix $X : n \times G$

$$y = X\beta + X\gamma + \varepsilon,$$

where $\beta : g \times 1$ and $\gamma : G \times 1$. Apart from being explanatory variables, the variables in $X$ serve as instruments. Additional instruments are collected in the matrix $Z : n \times K$, which may be assumed, without loss of generality, to satisfy $Z'X = 0$ and $Z'Z = I_K$. So $(X, Z)$ is the full matrix of instruments and equation (4.3) becomes

$$Y = (y, X, X) = (X, Z) \begin{bmatrix} \bar{\Pi} & I_G \\ \Pi & 0 \end{bmatrix} \begin{bmatrix} \beta' \\ \gamma' \end{bmatrix}, I_{G+g}$$

$$+ (R, 0) \begin{bmatrix} \beta' \\ \gamma' \end{bmatrix}, I_{G+g} + (u, V, 0),$$

where $\bar{\Pi} : G \times g$, $\Pi : K \times g$ and $R : n \times g$. The rows of $(u, V)$ are normally distributed with identical covariance matrix $\tilde{\Omega}$. So $\Omega$, the covariance matrix of any row of $(u, V, 0)$, is given by

$$\Omega = \begin{bmatrix} \tilde{\Omega} & 0 \\ 0 & 0 \end{bmatrix}.$$

The estimator $\hat{\delta}_F = (\hat{\beta}_F', \hat{\gamma}_F')'$ of $\delta = (\beta', \gamma')'$ is found by solving
4.4. A comparison of some MM estimators

\[(Y'P_{(X,Z)}Y - \hat{\lambda}_F Y'(I - P_{(X,Z)})Y) \begin{pmatrix} 1 \\ -\hat{\delta}_F \end{pmatrix} = 0, \quad (4.25)\]

where \( \hat{\lambda}_F \) is minimal. Let \( Y_1 = (y, X) \), then, as is shown in Appendix 4.A, \( \hat{\delta}_F = (\hat{\beta}_F', \hat{\gamma}_F')' \) satisfies

\[(Y'P_{Z}Y_1 - \hat{\lambda}_F Y_1'(I - P_{(X,Z)})Y_1) \begin{pmatrix} 1 \\ -\hat{\beta}_F \end{pmatrix} = 0, \quad (4.26)\]

where \( \hat{\lambda}_F \) is minimal, and

\[\hat{\gamma}_F = (X'X)^{-1}X'(y - X\hat{\beta}_F). \quad (4.27)\]

In the literature it is usual to concentrate on the estimation of \( \beta \) (Anderson, Kunitomo and Sawa (1982), Kunitomo (1980, 1981, 1986, 1987)). If one is only interested in this parameter vector, it suffices to consider the marginal distribution of \( \hat{\beta}_F \), which is given by

\[\hat{\beta}_F \sim F_{\text{MM}}(\tilde{\Omega}, \beta; \Pi'\Pi, K; R'R, n - G - K).\]

Since \( \tilde{\Omega} > 0 \), the arguments given in the previous subsection now simply apply to the subvector \( \beta_F \). That is, if the instruments \( Z \) are partitioned as

\[Z = (Z_1, Z_2)O,\]

for some orthogonal matrix \( O \), and \( P_1 = Z_1Z_1' \), we may consider the estimator \( \hat{\delta}_R \) which solves

\[(Y'P_1Y_1 - \hat{\lambda}_R Y_1'(I - P_{(X,Z)})Y_1) \begin{pmatrix} 1 \\ -\hat{\beta}_R \end{pmatrix} = 0,\]

for \( \hat{\lambda}_R \) minimal. Just as in Subsection 4.4.1, the group-asymptotic covariance matrix of the estimator \( \hat{\beta}_R \) may be smaller than the group-asymptotic covariance matrix of the estimator \( \hat{\delta}_F \) if the disregarded instruments \( Z_2 \) are weak.
4.5 Monte Carlo simulations

In order to verify that $\hat{\delta}_R$ may be preferred to $\hat{\delta}_F$, as suggested by the group-asymptotic approximations of their distributions, we present the results of a Monte Carlo experiment in which their finite sample performances are compared. In view of the reduction of the singular case to the nonsingular case, as discussed in the previous section, the design is restricted to the case where $\Omega$ is nonsingular. A further reduction is implied by the following lemma, which presents a canonical representation of the MM distribution. Such a canonical form can also be found in Anderson et al. (1982) and Bekker (1994). For the formulation of this lemma let $\Omega$ be partitioned as

$$
\Omega = \begin{bmatrix}
\omega_{11} & \omega_{12} \\
\omega_{21} & \Omega_{22}
\end{bmatrix},
$$

where $\omega_{11}$ is a scalar, $\omega_{21} = \omega_{12}' : g \times 1$ and $\Omega_{22} : g \times g$.

**Lemma 4.4** Let the distribution of $\hat{\delta}$ be given by

$$
\hat{\delta} \sim F_{MM} (\Omega, \delta; \bar{M}, p; M^\perp, q),
$$

where $\Omega$ is nonsingular. The distribution of

$$
\hat{\delta}^* = \left( \frac{|\Omega_{22}|}{|\Omega|} \right)^{1/2} \Omega_{22}^{1/2} (\delta - \Omega_{22}^{-1} w_{21}),
$$

(4.28)

is then given by $\hat{\delta}^* \sim F_{MM} (I_{g+1}, \delta^*; \bar{M}^*, p; M^\perp^*, q)$, where

$$
\bar{M}^* = \Omega_{22}^{-1/2} \bar{M} \Omega_{22}^{-1/2},
$$

$$
M^\perp^* = \Omega_{22}^{-1/2} M^\perp \Omega_{22}^{-1/2},
$$

$$
\delta^* = \left( \frac{|\Omega_{22}|}{|\Omega|} \right)^{1/2} \Omega_{22}^{1/2} (\delta - \Omega_{22}^{-1} w_{21}).
$$

**Proof.** The proof is given in Appendix 4.B. □
As (4.28) is an invertible linear transformation, efficiency comparisons between \( \hat{\delta}_F \) and \( \hat{\delta}_R \) are not affected by this transformation. In other words, it may be assumed, without loss of generality, that \( \Omega \) equals the identity matrix. In order not to complicate the comparisons, the analysis is restricted to the univariate case \( g = 1 \). Furthermore, we assume that \( R = 0 \), so \( \hat{\delta}_F \) is the LIML estimator. Equation (4.17) then becomes

\[
Y = Z_1 \pi_1(\delta, 1) + Z_2 \pi_2(\delta, 1) + (u, V),
\]

(4.29)

where \( \pi_i : K_i \times 1 \), \( \Omega = I_2 \) and \( (Z_1, Z_2)'(Z_1, Z_2) = I_K = I_{K_1+K_2} \).

Let \( \mu_i^2 = \pi_i' \pi_i, i = 1, 2 \) and \( \mu^2 = \mu_1^2 + \mu_2^2 \), which is the noncentrality or concentration parameter (Anderson et al., 1982; Phillips, 1983) in case \( \Omega \) is the identity matrix. In the Monte Carlo experiment the distribution of \( \hat{\delta}_F \) and \( \hat{\delta}_R \) are simulated for cases where the instruments \( Z_2 \) are weak in comparison with the instruments \( Z_1 \): \( \mu_2^2 = 0.1 * \mu^2 \).

In the simulations, 20,000 replications were drawn from the distributions of \( \hat{\delta}_F \) and \( \hat{\delta}_R \):

\[
\hat{\delta}_F \sim F_{\text{MM}} (I_2, \delta; \mu^2, K; 0, n - K),
\]

\[
\hat{\delta}_R \sim F_{\text{MM}} (I_2, \delta; \mu_1^2, K_1; 0, n - K);
\]

for the following parameter combinations:

- \( K = 40, \mu^2 = 50, 75, 100, 400 \),
- \( K = 20, \mu^2 = 25, 50, 75, 100 \),
- \( K = 10, \mu^2 = 15, 25, 50, 75 \),

and \( \delta = -0.1, -0.3, -0.5, -1 \) and \( n = 1200 \). For these 48 combinations two choices of \( K_1 \) are used: \( K_1 = 0.5K \) and \( K_1 = 0.2K \); and in all 96 cases: \( \mu_1^2 = 0.9\mu^2 \).

To get a first impression of the results, Figure 4.1 gives the simulated distributions of \( \hat{\delta}_F \) and \( \hat{\delta}_R \) for the parameter combination \( \mu^2 = 25, K = 20, K_1 = 4, \delta = -0.1 \). The distributions are found to be correctly centered around the true value of \( \delta \). Moreover, the distribution of \( \hat{\delta}_R \) is less spread out than the distribution of the LIML estimator \( \hat{\delta}_F \).
Figure 4.1: Simulated distributions of $\hat{\delta}_R$ and LIML.

As a measure of the relative efficiency of $\hat{\delta}_R$ with respect to $\hat{\delta}_F$ one might consider

$$e(\hat{\delta}_F, \hat{\delta}_R) = \frac{E \left\{ (\hat{\delta}_F - \delta)^2 \right\}}{E \left\{ (\hat{\delta}_R - \delta)^2 \right\}}.$$ 

However, the variances of the estimators do not exist (Mariano and Sawa (1972)). Therefore, we use another measure of the spread of the distributions. Let $\Delta$ be such that 90% of the observations are larger than $\delta - \Delta$ and smaller than $\delta + \Delta$. For the relative efficiency we use

$$e(\hat{\delta}_F, \hat{\delta}_R) = \frac{\Delta_F}{\Delta_R}. \quad (4.30)$$

This measure takes into account possible differences in the central location of the distributions of $\hat{\delta}_F$ and $\hat{\delta}_R$. It is worth noting, however, that their central locations were found to be almost the same. The maximum of the absolute difference between the median and the true value of $\delta$ over all parameter points was found to be equal to 0.005 for $\hat{\delta}_F$ and 0.004 for $\hat{\delta}_R$. So the distributions are almost median unbiased.
for the selected parameter points. This is in accordance with the median unbiasedness of an $o(n^{-1})$ Edgeworth approximation of the LIML distribution as given in Rothenberg (1983).

In Figure 4.1, $\Delta_F$ is roughly 0.54 and $\Delta_R$ is roughly 0.41. So $e(\hat{\delta}_F, \hat{\delta}_R) \approx 1.32$ for the parameter combination $\mu^2 = 25, K = 20, K_1 = 4, \delta = -0.1$. Instead of tabulating all 96 values of $e$, these values are compared with approximations of $e$ based on group-asymptotics and large sample asymptotics. That is, if the large sample asymptotic distributions of $\hat{\delta}_F$ and $\hat{\delta}_R$ were exact, the estimators would be normally distributed and $\Delta_F$ and $\Delta_R$ would be proportional to the standard deviations $V_{F}^{0 \ 1/2}$ and $V_{R}^{0 \ 1/2}$, resp. So an approximation of $e$ based on large sample asymptotics is given by

$$e^0 = \frac{V_{F}^{0 \ 1/2}}{V_{R}^{0 \ 1/2}} = \sqrt{\frac{\mu_1^2}{\mu^2}}.$$

Similarly, if the group-asymptotic distributions of $\hat{\delta}_F$ and $\hat{\delta}_R$ were exact, $\Delta_F$ and $\Delta_R$ would be proportional to $V_{F}^{* \ 1/2}$ and $V_{R}^{* \ 1/2}$, resp. So an approximation of $e$ based on group-asymptotics is given by

$$e^* = \frac{V_{F}^{* \ 1/2}}{V_{R}^{* \ 1/2}}.$$

Since $\mu_1^2 = 0.9\mu^2$, $e^0$ is constant over all 96 parameter combinations:

$$e^0 = \sqrt{0.9} \approx 0.95.$$

In particular, $e^0 \approx 0.95$ is the large sample asymptotic approximation of $e \approx 1.32$ found for the parameter combination $\mu^2 = 25, K = 20, K_1 = 4, \delta = -0.1$. Most importantly, the large sample asymptotic approximation gives the false impression that the spread of the distribution of $\hat{\delta}_F$ is smaller than the spread of the distribution of $\hat{\delta}_R$. The group-asymptotic approximation $e^*$, on the other hand, equals 1.22 in this case. Although this is still too small, it correctly suggest that the distribution of $\hat{\delta}_F$ is more spread out than the distribution of $\hat{\delta}_R$.

In order to make such comparisons for all 96 parameter combinations, in Figure 4.2 the values of $e$ are plotted against their group-asymptotic approximation $e^*$. The horizontal line corresponds with the large sample
Figure 4.2: Relative efficiency of $\hat{\delta}_R$ and LIIML.

asymptotic approximation of $e$: $e^0$. The plus signs indicate points for which $K_1 = 0.2K$, and the circles are points for which $K_1 = 0.5K$. As could be expected from the group-asymptotic approximations, the gain in efficiency is more extreme when more instruments are disregarded. However, $e^*$ is, in most cases, too large if $K_1 = 0.5K$ and, in most cases, too small if $K_1 = 0.2K$.

On the other hand, the large sample asymptotic approximation $e^0$ is systematically too small for the selected parameter combinations. It suggests that in all cases the spread of the distribution of $\hat{\delta}_F$ is smaller than the spread of the distribution of $\hat{\delta}_R$. This is clearly not the case. Disregarding the weak instruments $Z_2$ may indeed improve the MM estimator. For a measure of the relative efficiency of the two MM estimators, the group-asymptotic approximations should be preferred.

4.6 Conclusion

In this chapter the MM, or LIIML, estimator is compared with an estimator where some of the instruments are disregarded. Such comparisons
are made using group-asymptotic and large sample asymptotic approximations to the finite sample distributions of the estimators. Whereas the large sample asymptotic approximations suggest that disregarding instruments always leads to a loss in efficiency, the group-asymptotic approximations suggest that disregarding instruments may improve the MM estimator if the disregarded instruments are weak. For the case where the first-stage regression is linear, the different implications are put to the test by simulating the finite sample distributions of the two estimators. The Monte Carlo simulations show that a group-asymptotic approximation of the relative efficiency of the two estimators is more accurate when compared to large sample approximations. Disregarding weak instruments may indeed improve the performance of MM-type estimators.

It should be emphasized, however, that the results given in this chapter are only useful in practice if there exists a priori evidence that some of the instruments are weak. When weak instruments are not identified a priori it is tempting to use the data to locate them and then to use the estimator that disregards these instruments. Similar attempts have been undertaken in the context of the TSLS estimator. For instance, Hall, Rudebusch and Wilcox (1994) examined the effects of pre-testing instrument relevance and concluded that this can in fact exacerbate problems. Similar problems may be expected in the present context.

Nevertheless, the results are interesting from a theoretical point of view. In particular, the results put statements that say that the LIML estimator is asymptotically efficient in another perspective. For example, in Anderson et al. (1982, p1025) it is said that Kunitomo (1980), Morimune and Kunitomo (1980) and Kunitomo (1982) have shown that the LIML estimator is asymptotically efficient under a parameter sequence where the number of instruments increases with the number of observations. However, in this chapter it is shown, based on results similar to the tables of Anderson et al. (1982), that LIML should not be considered as (asymptotically) efficient.
4.A A reformulation of the MM estimator

In Subsection 4.4.3 it is stated that the MM estimator \( \hat{\delta}_F = (\hat{\beta}'_F, \hat{\gamma}'_F)' \), which solves

\[
\begin{pmatrix}
Y' P_{(X,Z)} Y - \hat{\lambda}_F Y' (I - P_{(X,Z)}) Y
\end{pmatrix} \begin{pmatrix}
1 \\
-\hat{\delta}_F
\end{pmatrix} = 0,
\]

(4A.1)

where \( \hat{\lambda}_F \) is minimal, satisfies

\[
\begin{pmatrix}
Y' P_Z Y_1 - \hat{\lambda}_F Y_1' (I - P_{(X,Z)}) Y_1
\end{pmatrix} \begin{pmatrix}
1 \\
-\hat{\beta}_F
\end{pmatrix} = 0,
\]

(4A.2)

and

\[
\hat{\gamma}_F = (X' X)^{-1} X' (y - X \hat{\beta}_F),
\]

(4A.3)

where \( Y_1 = (y, X) \). To prove this statement, note that \( X' Z = 0 \) implies \( P_{(X,Z)} X = X \) and

\[
P_{(X,Z)} = P_Z + P_X.
\]

So, \( Y' P_{(X,Z)} Y \) can be written as

\[
Y' P_{(X,Z)} Y = \begin{bmatrix}
Y_1' P_{(X,Z)} Y_1 & Y_1' X \\
X' Y_1 & X' X
\end{bmatrix}
\]

\[
= \begin{bmatrix}
Y_1' P_Z Y_1 & 0 \\
0 & 0
\end{bmatrix} + (Y_1, X)' P_X (Y_1, X),
\]

and \( Y' (I - P_{(X,Z)}) Y \) can be written as

\[
Y' (I - P_{(X,Z)}) Y = \begin{bmatrix}
Y_1' (I - P_{(X,Z)}) Y_1 & 0 \\
0 & 0
\end{bmatrix}.
\]

With this new representation of the matrices \( Y' P_{(X,Z)} Y \) and \( Y' (I - P_{(X,Z)}) Y \) we find from (4A.1) that
4.B. The proof of Lemma 4.4

\[(1, -\hat{\beta}'_F) \left( Y_1' P_Z Y_1 - \hat{\lambda}_F Y_1'(I - P_{(X,Z)}) Y_1 \right) \begin{pmatrix} 1 \\ -\hat{\beta}'_F \end{pmatrix} + \]
\[\left( Y_1 \begin{pmatrix} 1 \\ -\hat{\beta}'_F \end{pmatrix} - X \hat{\gamma}'_F \right)' P_X \left( Y_1 \begin{pmatrix} 1 \\ -\hat{\beta}'_F \end{pmatrix} - X \hat{\gamma}'_F \right) = 0. \] (4A.4)

The matrix \( Y_1' P_Z Y_1 - \hat{\lambda}_F Y_1'(I - P_{(X,Z)}) Y_1 \) is the Schur complement of \( X'X \) in \( Y'P_{(X,Z)} Y - \hat{\lambda}_F Y'(I - P_{(X,Z)}) Y \). Since the latter matrix is positive semidefinite, the same holds true for \( Y_1' P_Z Y_1 - \hat{\lambda}_F Y_1'(I - P_{(X,Z)}) Y_1 \) (Ouellette (1981), Corollary 3.1). Hence, both terms in (4A.4) are nonnegative, which implies that

\[\left( Y_1' P_Z Y_1 - \hat{\lambda}_F Y_1'(I - P_{(X,Z)}) Y_1 \right) \begin{pmatrix} 1 \\ -\hat{\beta}'_F \end{pmatrix} = 0, \] (4A.5)

\[P_X \left( Y_1 \begin{pmatrix} 1 \\ -\hat{\beta}'_F \end{pmatrix} - X \hat{\gamma}'_F \right) = 0. \] (4A.6)

Now (4A.5) is (4A.2) and solving \( \hat{\gamma}'_F \) from (4A.6) shows the validity of (4A.3).

4.B The proof of Lemma 4.4

Since the distribution of \( \hat{\delta} \) is \( F_{MM} (\Omega, \delta; \bar{M}, p; M^\perp, q) \), \( \hat{\delta} \) can be seen as the solution of

\[(\bar{S} - \hat{\lambda} S^\perp) \begin{pmatrix} 1 \\ -\hat{\delta} \end{pmatrix} = 0, \]

where \( \hat{\lambda} \) is minimal and \( \bar{S} \) and \( S^\perp \) are independent with Wishart distributions as given in (4.5). Let the \( (g+1) \times (g+1) \) matrix \( Q \) be defined as

\[Q = \begin{bmatrix}
(\Omega_{22}/|\Omega|)^{1/2} & 0 \\
-(\Omega_{22}/|\Omega|)^{1/2} \Omega_{22}^{-1} \omega_{21} & \Omega_{22}^{-1/2}
\end{bmatrix}, \]

where \( |\Omega_{22}| \) and \( |\Omega| \) denote the determinant of the matrix \( \Omega_{22} \) and \( \Omega \), resp. The estimator \( \hat{\delta}^* \) satisfies
\[ Q\left( \frac{1}{-\hat{\delta}^*} \right) = (|\Omega_{22}|/|\Omega|)^{1/2} \left( \frac{1}{-\hat{\delta}} \right). \]

So \( \hat{\delta}^* \) is the solution of

\[ (Q'SQ - \hat{\lambda}Q'S^\perp Q)\left( \frac{1}{-\hat{\delta}^*} \right) = 0, \]

where \( \hat{\lambda} \) is minimal. The statistics \( Q'SQ \) and \( Q'S^\perp Q \) are independent with Wishart distributions given by

\[ Q'SQ \sim W\left(Q'\Omega Q, p, Q'(\delta, I_g)' \bar{M} (\delta, I_g) Q \right), \]
\[ Q'S^\perp Q \sim W\left(Q'\Omega Q, q, Q'(\delta, I_g)' M^\perp (\delta, I_g) Q \right). \]

Since \( Q'\Omega Q = I_{g+1} \) and \( Q'(\delta, I_g)' = (\delta^*, I_g)' \Omega_{22}^{-1/2} \), the distribution of \( \hat{\delta}^* \) is \( F_{MM} (I_{g+1}, \delta^*; \bar{M}^*, p; M^\perp*; q) \), which proves Lemma 4.4.
4.B. The proof of Lemma 4.4