Exclusion restrictions in instrumental variables equations
Nijman, Theo; Steel, M.F.J.

Publication date:
1990

Citation for published version (APA):
Exclusion Restrictions in Instrumental Variables Equations

by

Theo E. Nijman and Mark F.J. Steel

Reprinted from Econometric Reviews, Vol. 9, No. 1, 1990

Reprint Series no. 40
Exclusion Restrictions in Instrumental Variables Equations

by Theo E. Nijman and Mark F.J. Steel

Reprinted from Econometric Reviews, Vol. 9, No. 1, 1990

Reprint Series no. 40
EXCLUSION RESTRICTIONS IN INSTRUMENTAL VARIABLES EQUATIONS

Theo E. Nijman
Mark F.J. Steel

Department of Econometrics
Tilburg University
P.O. Box 90153
5000 LE Tilburg
The Netherlands

Key Words and Phrases: two-stage estimators; incomplete models; instrumental variables; recursive exclusion restrictions; nuisance parameters; asymptotic efficiency

ABSTRACT

In this paper we consider two-stage estimators of parameters of a structural equation in a model with recursive exclusion restrictions on the instrumental variables equations. The estimators considered are simple OLS and GLS estimators after substitution of estimates of the systematic part of the IV equations for the endogenous variables. It is known in the literature that neither imposing the restrictions in the first stage nor ignoring them will in general be more efficient than the alternative. We introduce a class of mixed instrumental variables estimators (MIV) with these possibilities as special cases which yields an estimator which is not only more efficient than the two stage estimators considered in the literature but as efficient as an efficient system estimator like 3SLS.

1. INTRODUCTION

A very common phenomenon in econometrics is the necessity to estimate far more parameters than the model user is ultimately interested in. The reasons for attaching a particular importance to a small subset of the parameters, the so-called parameters of
interest, may be multiple. Their meaning may derive from economic theory and inference on them may constitute evidence in favour of or against certain theories. An alternative source of interest might be their relative stability in changing environments, leading to models which can be used for policy simulations.

Two examples in which the problem of the dimensionality of the vector of nuisance parameters is often particularly important are incomplete simultaneous equations models and models containing unobserved rational expectations. As explained in detail in Richard (1979) and Richard (1984) formulating statistical assumptions in terms of the observable endogenous variables instead of adding random error terms to an already existing set of deterministic equations will often lead to incomplete simultaneous equation models. As such an incomplete system allows an infinity of solutions, auxiliary equations have to be added, which usually express the endogenous variables as linear functions of a set of instrumental variables and introduce a number of nuisance parameters in the model. In case of a model containing unobserved rational expectations, the model has to be completed by equations that describe how these expectations are generated. The parameters in these equations will also typically be nuisance parameters only.

In this paper we assume that the parameters of interest are the coefficients of a structural equation in a simultaneous equation model. We restrict ourselves to a case of limited information, i.e. the model is completed with reduced form equations, but we assume that blocks of zero restrictions on the reduced form coefficients hold. These zero restrictions can originate e.g. from economic theory, from assumptions of strong exogeneity in the sense of Engle et.al. (1983), etc. Efficient parameter estimates can of course be obtained using a three stage least squares procedure, but this approach is computationally not very attractive if the number of nuisance parameters is large. In applications one will typically use a two stage least squares procedure either imposing the zero restrictions on the reduced
form or not. It is well known in the literature (see e.g. Turkington (1985)) that imposing the restrictions in the first stage does not necessarily yield an efficiency gain over standard (unrestricted) two stage least squares. In this paper we restrict ourselves to models where the restrictions can be arranged in a recursive pattern and show for that case how to construct simple two stage estimators which are more efficient than 2SLS. Moreover we derive a two stage estimator which is as efficient as 3SLS.

The plan of this paper is as follows. In section two we introduce the model and describe the class of mixed IV estimators to be considered. The optimal mixed IV estimator and its asymptotic efficiency are derived in section three. In section four we extend the results to a more general model. Finally, section five contains some concluding remarks.

2. THE MODEL AND THE ESTIMATORS CONSIDERED

Consider the following model:

\[ y' = Y\beta_1 + \tilde{Z}\beta_2 + W\beta_3 + \epsilon \]  
\[ Y = Z\Pi_1 + U \]  

where \( y \) is a \( T \) dimensional vector, while \( Y, \tilde{Z}, W \) and \( Z \) are matrices of dimension \( T \times k, T \times h_1, T \times h_2, T \times n \) respectively, and assume that

\[ V(\epsilon: U) = \Sigma \otimes I_T, \quad \Sigma = \begin{bmatrix} \sigma_{00} & \sigma_{01} \\ \sigma_{10} & \sigma_{11} \end{bmatrix}. \]  

where \( \Sigma \) is P.D.S. The variables in \( Z, \tilde{Z} \) and \( W \) are assumed to be weakly exogenous for the parameters in (1) and (2). The matrix \( \tilde{Z} \) is assumed to be a sub-matrix of \( Z \), which defines \( W \) as the set of variables that do not appear in the instrumental variables equation (2).
The model outlined above is sufficiently general to illustrate the argument and to be of relevance in applications but not too general to cloud the essentials by tedious mathematical details. The model is a straightforward extension of the case where \( k = 1 \), which has been extensively discussed in the literature on the estimation of models with unobserved rational expectations (see e.g. Pagan (1984), Turkington (1985) and Pesaran (1987)).

The model in (1) to (3) can be written in short as

\[
\begin{align*}
y &= X \beta + e \\
X &= Z_s \Pi_s + U_s
\end{align*}
\]  

where

\[
\begin{align*}
X &= (Y : Z : W), \\
Z_s &= (Z : W), \\
U_s &= (U : 0 : 0), \\
\beta' &= (\beta_1' : \beta_2' : \beta_3'), \\
\Pi_s &= \begin{pmatrix} \Pi_1 & S & 0 \\ \Pi_2 & 0 & I \end{pmatrix}, \text{ with } \Pi_2 = 0
\end{align*}
\]

and where \( S \) is a selection matrix. Throughout we assume that \( \text{plim} \ T^{-1}Z_s'Z_s \) is finite and non-singular.

As already noted in the introduction, the parameter vector \( \beta \) in (1) is assumed to consist of the parameters of interest of our analysis as this is the part of the system that originated from economic theory conveying relevance and interpretability to its parameters. All other parameters will be treated as nuisance parameters. As has often been stressed in the literature there is
no need to estimate all parameters of such a system jointly. This has prompted the wide-spread use of two-stage estimators, the general theory of which is discussed in Pagan (1986). The two-step estimators of $\beta$ in (1) which have been proposed in the literature can be written as

$$
\hat{\beta}_{OLS} = \left( \tilde{\Pi} \tilde{Z}' \tilde{Z} \tilde{\Pi} \right)^{-1} \tilde{\Pi} \tilde{Z}' \tilde{y}
$$

where $\tilde{\Pi}$ is a consistent estimator of $\Pi$. It can easily be checked that $\hat{\beta}_{OLS}$ is consistent. All estimators $\tilde{\Pi}$ that will be considered in this paper are of the form

$$
\tilde{\Pi} = \begin{bmatrix}
\tilde{\Pi} & S & 0 \\
0 & 0 & I
\end{bmatrix}
$$

(7)

If $\tilde{\Pi}$ is chosen to be the unrestricted OLS estimator $\Pi$ of $\Pi$,

$$
\hat{\Pi} = \left( \tilde{Z}' \tilde{Z} \right)^{-1} \tilde{Z}' \tilde{y}
$$

(8)

the estimator $\hat{\beta}_{OLS}$ coincides with 2SLS. The knowledge that $\Pi_2 = 0$ is not imposed in (8). Because every endogenous variable is explained by the same set of instruments this procedure will be denoted by "common blocks of instrumental variables" or CIV following Richard (1987). In the literature on rational expectations the CIV or 2SLS estimator is usually referred to as the "errors in variables method", following Wickens (1982).

A choice of $\tilde{\Pi}$ in (7) which imposes the zero restrictions in the IV equations is

$$
\hat{\Pi} = \begin{bmatrix}
(Z'Z)^{-1} Z'Y \\
0
\end{bmatrix}
$$

(9)

This procedure will be labelled "individual blocks of instrumental variables" (IIV) because no longer the same set of instruments is used for all regressors. This procedure is known as the the substitution approach in the literature on rational expectations.
It has been established in the literature (see e.g. Pagan (1984) and Turkington (1985)) that neither CIV nor IIV will in general be more efficient than the other. A natural question therefore is whether we can find a way in between the two extremes considered here which will dominate both in terms of the asymptotic covariance matrix. In this paper we propose to use the OLS estimator defined in (6) with

$$\tilde{\Pi} = \tilde{\Pi} \Gamma + \tilde{\Pi} (I - \Gamma)$$

where $\Gamma$ is a $k \times k$ weighting matrix. This approach will be denoted by mixed instrumental variables (MIV) estimation. Note that it reduces to CIV for $\Gamma = 0$ and to IIV for $\Gamma = I$. We will show that the MIV estimator with a suitable chosen weighting matrix $\Gamma$ is at least as efficient as both CIV and IIV. In Nijman (1985) a related estimator was proposed for the case where $k = 1$.

Besides the OLS estimator in (6) we will consider in this paper the GLS estimator

$$\hat{\beta}_{\text{GLS}} = (\tilde{\Pi}^* Z_\Omega^{-1} Z_\Omega^* \tilde{\Pi}^*)^{-1} \tilde{\Pi}^* Z_\Omega^{-1} y$$

where $\tilde{\Pi}^*$ is again generated by (7) and (10), $\Omega$ is the $T \times T$ covariance matrix of the structural disturbance term after substituting for the endogenous variables and the hat (\hat{} ) denotes that unknown parameters are replaced by consistent estimates. Convenient expressions for $\Omega$ and its inverse will be given in the next section. A similar GLS estimator has recently been proposed by Pesaran (1987, p. 168) for the case where $k = 1$. Hoffman (1987) proposed a GLS estimator in a non-recursive model similar to the model to be discussed in section 4. Both authors considered IIV procedures only. The GLS estimator which is proposed in this paper is also related to the modified nonlinear 2SLS procedure proposed in Amemiya (1975), who does not consider exclusion restrictions on the IV equations but the case where nonlinearity of the structural equation prevents efficiency of the usual 2SLS estimator.
In section 3 we show that the two-stage GLS procedure in (11) yields an efficient estimator. Note that several authors (see Wickens (1982), Turkington (1985)) have suggested that this would require system methods. Efficient estimates of $\beta$ can, of course, also be obtained using such a full system method like 3SLS\(^1\). However, 3SLS will require matrix inversions of a dimension equal to the total amount of free parameters in $\beta$ and $\Pi_*$. The dimension of the matrices to be inverted in two-stage approaches will usually be far smaller.

3. THE CHOICE OF THE WEIGHTING MATRIX

In this section we will derive the weighting matrix $\Gamma$ which minimizes the asymptotic variance covariance matrix (in the PDS sense) of the OLS and GLS estimators introduced in the previous section.

The first estimator to be considered is the OLS estimator of $\beta$ in

$$y = Z_* \hat{\Pi}_* \beta + w,$$

with

$$w = e + U \beta_1 + Z_* (\Pi - \hat{\Pi}) \beta_1 + Z_* (\Pi - \hat{\Pi}) \Gamma \beta_1.$$

Now assume for simplicity that $e$ and $U$ are independent of $Z$. The variance covariance matrix of $w$ conditional on $Z_*$ can be shown (see appendix) to equal

$$E[ww' | Z_*] = \Omega = \sigma_{00} I + b_1 M_{Z_*} + b_2 (P_{Z_*} - P_Z).$$

---

\(^1\) This might be considered a misnomer as we analyse a "completed" model instead of a truly structural one. However, we retain this denomination for lack of a clear alternative.
with

\[ b_1 = (1 : \beta_1') \Sigma \begin{pmatrix} 1 \\ \beta_1 \end{pmatrix} - \sigma_{00} \]  

(15)

and

\[ b_2 = (1 : \beta_1' \Gamma') \Sigma \begin{pmatrix} 1 \\ \Gamma \beta_1 \end{pmatrix} - \sigma_{00}. \]  

(16)

where we defined \( P_Q = Q(Q'Q)^{-1}Q' \) and \( M_Q = I - P_Q \) for any matrix \( Q \) of full column rank. Using equation (14) the large sample variance of \( \sqrt{T} \beta_{OLS} \) can be expressed as

\[ \text{Avar}(\beta_{OLS}) = A^{-1} \left\{ \sigma_{00} A + b_2 (A-B) \right\} A^{-1}. \]  

(17)

where \( A = \text{plim} T^{-1} \Pi Z Z \Pi \) and \( B = \text{plim} T^{-1} \Pi Z Z P Z Z \Pi \). Note that as \( A-B \) is positive semi-definite the sign of the coefficient \( b_2 \) in case of IIV determines the relative efficiency of CIV and IIV, i.e. of ignoring the restrictions versus incorporating them exactly in the first stage. If e.g. \( k = 1 \) and \( \beta_1 \) and \( \sigma_{01} \) have similar signs CIV will be more efficient than IIV which implies that we lose efficiency by taking the true exclusion restrictions into account in the first stage in this manner (see Turkington (1985)). From (16) and (17) it is evident that the weights of the optimal MIV estimator should satisfy

\[ \Gamma \beta_1 = - \Sigma^{-1} \sigma_{11} \sigma_{10}. \]  

(18)

in which case the large sample variance is given by (17) with \( b_2 \) replaced by \( b_2^{\text{opt}} \) defined as

\[ b_2^{\text{opt}} = - \sigma_{01} \Sigma^{-1} \sigma_{10}. \]  

(19)

As \( \Gamma \) contains \( k \times k \) free elements and (18) only defines \( k \) equations we have (unless \( k = 1 \)) a continuum of solutions for the weighting matrix \( \Gamma \) that will yield an efficient MIV estimator. The restric-
restrictions in instrumental variables equations

...imposed on $\bar{H}$ if $\Gamma = I$. Note that the optimal weights depend on unknown parameters and therefore have to be replaced by preliminary consistent estimators in applications. In the appendix we show that this does not affect the large sample variance of the resulting estimator.

Let us now consider the GLS estimator defined in (11). The structure of $\Omega$ in (14) gives rise to a simple expression for its inverse using the properties of projection matrices,

$$\Omega^{-1} = (\sigma_{00} + b_1)^{-1} M_{Z^*} + (\sigma_{00} + b_2)^{-1} \{ P_{Z^*} + b_2 \sigma_{00}^{-1} P_Z \}. \tag{20}$$

which avoids the need to invert a $T \times T$ matrix numerically. Using this result the large sample variance of $\sqrt{T} \hat{\beta}_{\text{GLS}}$ can be shown to be

$$\text{Avar}(\hat{\beta}_{\text{GLS}}) = (\sigma_{00} + b_2) \{ A + b_2 \sigma_{00}^{-1} B \}^{-1}, \tag{21}$$

where $A$ and $B$ are defined below (17). The most efficient estimator is again obtained if $b_2$ is minimal, that is, if $\Gamma$ satisfies equation (18). Finally, note that the assumption of independence between $e$ and $U$ on the one hand and $Z^*$ on the other is only required to interpret $\Omega$ as the variance covariance matrix of $w$ conditional on $Z^*$. The results in (17) and (21) remain valid if this assumption is not met as long as $\tilde{Z}$, $W$ and $Z$ are weakly exogenous for the parameters in (1) and (2) as is assumed throughout.

Let us now consider whether the proposed two-stage estimators will be as efficient as an efficient full system estimator such as 3SLS. In the appendix the large sample variance of $\sqrt{T} \hat{\beta}_{\text{3SLS}}$ is shown to be

$$\text{Avar}(\hat{\beta}_{\text{3SLS}}) = \{ \sigma_{00} A - \sigma_{01} (\Sigma^{11})^{-1} \sigma_{10} B \}^{-1} \tag{22}$$

where superscripts refer to the corresponding blocks in the inverse. If the optimal weighting matrix $\Gamma$ which satisfies (18) is used...
the variance covariance matrix of $\sqrt{T}\hat{\beta}_{GLS}$ is given in (21) with $b_2$ replaced by $b_2^{opt}$ in (19). It can be verified that this expression coincides with (22) so that in case of optimal weighting $\sqrt{T}\hat{\beta}_{GLS}$ is as efficient as $\hat{\beta}_{3SLS}$. The estimator $\beta_{OLS}$ will be less efficient unless $\sigma_{10}^{-1} = 0$.

4. EXTENSION TO MORE GENERAL RECURSIVE RESTRICTIONS ON THE IV EQUATIONS

The results obtained in the previous sections can be extended to more general models with block-recursive restrictions. The only extension we consider in this paper is to the case where some instruments are excluded from a subset of the IV equations:

\[ y = X \beta + e \quad (23) \]
\[ X = \Pi \Lambda + U \quad (24) \]

where we now distinguish between the first $k_1$ variables in $Y$, say $Y_1$, and the other $k-k_1$ ($=k_2$) variables, say $Y_2$, so that

\[ X = (Y_1: Y_2: Z: W), \]
\[ Z = (Z_1: Z_2: W), \]
\[ U = (U_1: U_2: 0: 0), \]
\[ \beta' = (\beta_{11}': \beta_{12}': \beta_{21}'; \beta_{31}'), \]

\[ \Pi = \begin{pmatrix}
\Pi_{11} & \Pi_{12} & S_1 & 0 \\
\Pi_{21} & \Pi_{22} & S_2 & 0 \\
\Pi_{31} & \Pi_{32} & 0 & 1 \\
\end{pmatrix}, \text{ with } \Pi_{22} = 0, \Pi_{31} = 0 \\
\]

and $\Pi_{32} = 0$.

and

\[ V(e : U_1 : U_2) = \Xi \Lambda \Lambda' \]

\[ T \]

\[ \]
with

$$\Sigma = \begin{bmatrix}
\sigma_{00} & \sigma_{01} & \sigma_{02} \\
\sigma_{10} & \Sigma_{11} & \Sigma_{12} \\
\sigma_{20} & \Sigma_{21} & \Sigma_{22}
\end{bmatrix}.$$  

The above model coincides with the one in (5) and (6) if the fact that $\pi_{22} = 0$ is ignored. Analogously to (7), define $\tilde{\pi}$ to be some consistent estimator of the matrix $\pi$ used in constructing the matrix $\pi^*$ and reconsider the $\beta_{OLS}$ in (10) which is consistent for the parameter $\beta$ in (23) under the assumptions made. Many choices of $\tilde{\pi}$ can be thought of, e.g. $\tilde{\pi}$ in (8) where no restrictions are imposed at all, $\tilde{\pi}$ in (9) which implies that only those restrictions common to all IV equations are imposed, or

$$\pi^O = \begin{bmatrix}
(Z'Z)^{-1}Z'Y_1 & (Z_1'Z_1)^{-1}Z_1'Y_2 \\
0 & 0
\end{bmatrix},$$  

(25)

which imposes all restrictions in the system, including the exclusion of certain instruments for the equations describing $Y_2$. We consider MIV estimators for which

$$\tilde{\pi} = \pi^O \Delta + \tilde{\pi} (\Lambda^\ast \Delta) + \tilde{\pi} (I - \Lambda).$$  

(26)

Again, the weighting matrices $\Delta$ and $\Lambda$ will be of dimension $k \times k$, where $\Lambda$ performs the same role as $\Gamma$ in Section 3 and mixes the unrestricted $\tilde{\pi}$ with the other two estimators where restrictions are imposed on both sets of IV equations. Since the first $k_1$ columns of $\pi^O$ and $\tilde{\pi}$ coincide, only the last $k_2$ rows of $\Delta$, denoted by $\Delta^2$, will enter the analysis.

The OLS estimator in (10) is the OLS estimator of $\beta$ in

$$y = Z_2 \tilde{\pi} \beta + w$$  

(27)
with
\[
w = e + M_{\varepsilon} U \beta_1 + (P_{\varepsilon} - P_{Z_0}) U \lambda \beta_1 + (P_{Z_0} - P_{Z_1}) U_2 \Delta_2 \lambda_2 \beta_1.
\]  
(28)

so that (14) can be generalized to
\[
E[ww' | Z] = \Omega = \sigma_{00} I + c_1 M_{\varepsilon} U \beta_1 + c_2 (P_{\varepsilon} - P_{Z_0}) + c_3 (P_{Z_0} - P_{Z_1})
\]  
(29)

with
\[
c_1 = (1 \beta_1') \Sigma \left( \begin{array}{c} 1 \\ \beta_1 \end{array} \right) - \sigma_{00},
\]
\[
c_2 = (1 \beta_1') \Sigma \left( \begin{array}{c} 1 \\ \lambda \beta_1 \end{array} \right) - \sigma_{00},
\]
\[
c_3 = (1 \beta_1' \Delta_2) \Sigma \left( \begin{array}{c} 1 \\ \Delta_2 \lambda_2 \end{array} \right) - \sigma_{00},
\]

where
\[
\Sigma = \left( \begin{array}{cc} \sigma_{00} & \sigma_{02} \\ \sigma_{20} & \sigma_{22} \end{array} \right).
\]

The optimal values of \( \Delta \) and \( \lambda \) should satisfy \(^1\)

\[
\Delta_2 \beta_1 = - \Sigma^{-1}_{22} \sigma_{20}
\]  
(30)

---

\(^1\) It can be checked that both Zellner's (1962) and Richard's (1984) restricted maximum likelihood estimator of \( \hat{\Sigma} \) coincide with an MIV estimator. However, their implicit choice of the weighting matrices does not satisfy the optimality criterion in (30) and (31) and, therefore, the use of such an efficient estimator of the parameters in the IV equations in order to obtain the OLS estimator in (6) leads to an efficiency loss for the parameters of interest.
and

$$\mathbf{A} = -\left[ \begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array} \right]^{-1} \left[ \begin{array}{c} \sigma_{10} \\ \sigma_{20} \end{array} \right],$$

(31)

in which case we obtain (29) with $c_2$ and $c_3$ replaced by

$$c_{2, opt} = -\left( \begin{array}{cc} \sigma_{01} & \sigma_{02} \end{array} \right) \left[ \begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array} \right]^{-1} \left[ \begin{array}{c} \sigma_{10} \\ \sigma_{20} \end{array} \right] = (\sigma_{00})^{-1} - \sigma_{00},$$

(32)

and

$$c_{3, opt} = -\sigma_{02} \Sigma_{22}^{-1} \sigma_{20} = (\sigma_{00})^{-1} - \sigma_{00},$$

(33)

respectively, where $\sigma_{00}$ is the upper-left element of the inverse of the "shrunken" matrix $\Sigma$.

In general, the OLS estimator will not be as efficient as a system estimator like 3SLS. However, one can use the above results for the GLS estimator defined in (11). The properties of projection matrices can be used to show that if (30) and (31) hold the inverse of $Q$ is given by

$$Q^{-1} = (\sigma_{00} + c_1)^{-1} M_{Z^*} + \sigma_{00} P_{Z^*} + (\sigma_{00} - \sigma_{00}) P_{Z^*} + (\sigma_{00} - \sigma_{00}) P_{Z^*},$$

(34)

so that the large sample variance of the optimal GLS estimator will be

$$\text{Avar}(\mathbf{b}_{GLS}) = (\sigma_{00} + \sigma_{00} - \sigma_{00} + \sigma_{00} - \sigma_{00}) (\sigma_{00} - \sigma_{00})^{-1},$$

(35)

where $A$ and $B$ were defined in section 3 and

$$C = \text{plim} \left( T^{-1} \Pi_{e} Z_{e} P_{Z_{e}} Z_{e} \Pi_{e} \right).$$

(36)
Finally, the theory of partitioned matrices can be used to show that the large sample variance of an efficient estimator of $\beta$ such as 3SLS coincides with (35) (see appendix). As in the previous section, a correctly weighted MIV estimator will be fully efficient.

5. CONCLUDING REMARKS

In this paper we considered two-stage estimators of parameters of a structural equation in a model with recursive exclusion restrictions on the instrumental variables equations. Neither imposing these restrictions in the first stage regression (IIV) nor ignoring them (CIV) will in general be more efficient than the alternative even if e.g. the efficient SUR estimator of the reduced form coefficients is used. We introduced a class of mixed instrumental variables estimators (MIV) with IIV and CIV as special cases which yields a two-stage estimator which is more efficient than IIV and CIV and as efficient as an efficient system estimator such as 3SLS. The estimator of the reduced form coefficients to be substituted in the structural equation is a weighted average of the standard restricted and unrestricted estimators. The resulting estimator of the structural parameters is computationally much more attractive than other efficient estimators such as 3SLS. The results can be extended to more general patterns of recursive restrictions.

APPENDIX: DETAILS ON THE TECHNICALITIES

In this appendix we will derive the result in (14), we show that the substitution of consistent estimates for unknown coefficients in the weighting matrices in equations like (18), (30) and (31) does not affect the large sample variance of the two-stage estimators and, finally, we show how the expression for the asymptotic variance covariance matrix of the 3SLS estimator can be obtained.
In order to prove the result in (14) we first rewrite \( w \) in (13) as

\[
w = e + (I - P_Z) U \beta_1 + (P_Z - P_{Z'}) U r \beta_1. \tag{A1}
\]

Using the well-known fact that for matrices and vectors of appropriate dimensions it holds true that \( \text{vec} (NR^n) = (n' \otimes N) \text{vec} R \) one can establish that if \( V(R_1 : R_2) = G \otimes I_T \) \tag{A2}

where \( R_1 \) and \( R_2 \) are \( T \times r_1 \) and \( T \times r_2 \) matrices of random variables it will also be true that

\[
E \text{vec}(N_1 R_1 n_1) \text{vec}'(N_2 R_2 n_2) = n'_1 G_{12} n_2 N_1 N_2'. \tag{A3}
\]

if \( N_1 \) and \( N_2 \) are \( T \times T \) matrices and \( n_1 \) and \( n_2 \) are \( r_1 \times 1 \) and \( r_2 \times 1 \) vectors respectively, and \( G_{12} \) is the upper-right block of the matrix \( G \) defined in (A2). The evaluation of \( \Omega = E[ww' | Z] \) involves a large number of expressions of the type (A3), yielding

\[
\Omega = \sigma_{00} I_T + b_1 M_{Z'_*} + b_2 (P_{Z'_*} - P_Z) \tag{A4}
\]

with \( b_1 \) and \( b_2 \) defined in the main text.

As far as the estimation of the weighting matrices is concerned, reconsider e.g. equations (15) and (16). If the weighting matrix \( \Gamma \) is estimated (15) and (16) have to be replaced by

\[
y = Z_* \tilde{\Pi}_* \beta + w \tag{18'}
\]

with

\[
w = e + U \beta_1 + Z_* (\Pi - \tilde{\Pi}) \beta_1 + Z_* (\Pi - \tilde{\Pi}) \Gamma \beta_1. \tag{19'}
\]
Note, however, that

\[ (1/T) \pi \hat{z}_t \hat{w} = (1/T) \pi \hat{z}_t \hat{w} + \pi \hat{z}_t \hat{w}^T \pi (\hat{z}_t \hat{w})^{-1} (\pi - \pi) (\pi - \pi) \beta \]  \hspace{1cm} (A5) \]

and that the second term in the right hand side of (A5) converges to zero, in probability so that the large sample variance of \( \hat{\beta}_{OLS} \) is not affected by the estimation of \( \Gamma \). Similar arguments hold true for the case in section 4 and for GLS estimators.

Finally, we consider the expressions for the large sample variance covariance matrix of the 3SLS estimator. First of all write the model in (23)-(24) in vec form as

\[ y^* = H \delta + u^* \]  \hspace{1cm} (A6) \]

with

\[ y^* = \text{vec} \left( \begin{array}{c} y : Y_1 : Y_2 \end{array} \right) = \begin{bmatrix} y \\ \text{vec} Y_1 \\ \text{vec} Y_2 \end{bmatrix} \]  \hspace{1cm} (A7) \]

\[ H = \begin{bmatrix} (Y_1 : Y_2 : \tilde{Z} : \tilde{W}) & 0 & 0 \\ 0 & (I_k \bigotimes Z) & 0 \\ 0 & 0 & (I_k \bigotimes Z_1) \end{bmatrix} \]  \hspace{1cm} (A8) \]

\[ \delta^* = (\beta_1^* : \beta_2^* : \beta_3^* : \text{vec} \Pi_1 : \text{vec} \Pi_{12}) \]  \hspace{1cm} (A9) \]

and

\[ u^* = (e' : \text{vec} U_1 : \text{vec} U_2)' \]  \hspace{1cm} (A10) \]

In section 4 we have already assumed that \( V(u^*) = \Sigma \bigotimes \Sigma \). It is well known (see e.g. Schmidt (1976), p.207) that the asymptotic variance covariance matrix of the 3SLS estimator \( \sqrt{T} \delta_{3SLS} \) of \( \delta \) can be expressed as

\[ \text{Avar} (\delta_{3SLS}) = \text{plim} \left\{ H' \left( \Sigma^{-1} \bigotimes P_{Z_{\hat{w}}} \right) H \right\}^{-1} \]  \hspace{1cm} (A11) \]
with \( \mathbf{P}_{Z} \) defined in the main text. From (A11) we then deduce the asymptotic covariance matrix of the first \( k+h_1+h_2 \) elements of \( \hat{\beta}_{3SLS} \) using the theory of partitioned matrices and projection matrices.

Rewrite (A11) in the notation of (4) and (5), denoting blocks of \( \mathbf{I}^{-1} \) by superscripts, to obtain \( \text{Avar}(\hat{\beta}_{3SLS}) = \)

\[
\text{plim} T \begin{pmatrix}
\sigma^{00} \mathbf{P}_{Z}^* \mathbf{Z}^* \mathbf{P}_{Z}^* & \sigma^{01} \mathbf{P}_{Z}^* \mathbf{Z}^* \mathbf{Z}^* & \sigma^{02} \mathbf{P}_{Z}^* \mathbf{Z}^* \mathbf{Z}^* \\
\sigma^{10} \mathbf{Z}^* \mathbf{P}_{Z}^* \mathbf{Z}^* & \sigma^{11} \mathbf{Z}^* \mathbf{Z}^* & \sigma^{12} \mathbf{Z}^* \mathbf{Z}^* \\
\sigma^{20} \mathbf{Z}^* \mathbf{P}_{Z}^* \mathbf{Z}^* & \sigma^{21} \mathbf{Z}^* \mathbf{Z}^* & \sigma^{22} \mathbf{Z}^* \mathbf{Z}^*
\end{pmatrix}^{-1}.
\]

(A12)

The upper-left block of this inverse corresponds to \( \beta \) and can be written as

\[
\text{Avar}(\hat{\beta}_{3SLS}) = \left( \sigma^{00} \mathbf{A} - \sigma^{01} (\mathbf{I}^{11})^{-1} \sigma^{10} \mathbf{B} - c_\ast \mathbf{C} \right)^{-1}
\]

(A13)

where

\[
c_\ast = -\left( \sigma^{02} - \sigma^{01} (\mathbf{I}^{11})^{-1} \sigma^{10} \right) (\mathbf{I}^{22} - \mathbf{I}^{21} (\mathbf{I}^{11})^{-1} \mathbf{I}^{12})^{-1} (\sigma^{20} - \mathbf{E}^{21} (\mathbf{I}^{11})^{-1} \sigma^{10}).
\]

and \( \mathbf{A}, \mathbf{B} \) and \( \mathbf{C} \) are defined in the text. Using the fact that the inverse of the "shrunken" \( \mathbf{I} \) matrix, \( \mathbf{\hat{I}} \), can be written as

\[
\mathbf{\hat{I}}^{-1} = \begin{pmatrix}
\sigma^{00} - \sigma^{01} (\mathbf{I}^{11})^{-1} \sigma^{10} & \sigma^{02} - \sigma^{01} (\mathbf{I}^{11})^{-1} \sigma^{12} \\
\sigma^{20} - \mathbf{E}^{21} (\mathbf{I}^{11})^{-1} \sigma^{10} & \sigma^{22} - \mathbf{E}^{21} (\mathbf{I}^{11})^{-1} \mathbf{I}^{12}
\end{pmatrix}
\]

(A14)

and realizing that

\[
\sigma^{-1} = \sigma^{00} - \sigma^{02} (\mathbf{I}^{22})^{-1} \sigma^{20}.
\]

(A15)

we obtain that \( \text{Avar}(\hat{\beta}_{3SLS}) \) coincides with \( \text{Avar}(\hat{\beta}_{GLS}) \) in (35).
Clearly the efficiency of GLS in the model in equations (1)-(3)
is a special case of this result.

ACKNOWLEDGEMENTS

We wish to thank Marno Verbeek and Sybrand Schim van der Loeff and
two anonymous referees for helpful comments that have greatly
improved the presentation of the paper. Of course, the usual
disclaimer applies. Both authors conducted their research under a
fellowship of the Royal Netherlands Academy of Arts and Sciences
(K.N.A.W.).

REFERENCES

Amemiya, T. (1975). The nonlinear limited-information maximum-
likelihood estimator and the modified nonlinear two-stage

Econometrica, 51, 277-304.

tors in multi-equation generated regressor models, Review of

UK money demand function: a Bayesian approach to testing
exogeneity, Review of Economic Studies, 53, 603-634.

Nijman, Th.E. (1985). Missing Observations in Dynamic Macro-
Economic Modeling, doctoral dissertation, Free University,
Amsterdam.

sions with generated regressors, International Economic
Review, 25, 221-247.


Blackwell, Oxford.

Richard, J.-F. (1979). Exogeneity, inference and prediction in so-
called incomplete dynamic simultaneous equations models, CORE
discussion paper 7922, Université Catholique de Louvain.


No. 8 Th. van de Klundert and F. van der Ploeg, Wage rigidity and capital mobility in an optimizing model of a small open economy, De Economist 137, nr. 1, 1989, pp. 47 - 75.


