"Method-of-Moment View of Linear Simultaneous Equation Systems"

Myoung-jae Lee

Discussion Paper No. 07-19 (June 2007)
The Institute of Economic Research Korea University
Anam-dong, Sungbuk-ku,
Seoul, 136-701, Korea
Tel: (82-2) 3290-1632  Fax: (82-2) 928-4948
In this paper, we review the modern method-of-moment-based approaches to identification and estimation of linear simultaneous equation systems. First, we present the rank condition for the structural form (SF) parameter identification. The rank condition comes naturally and is much easier to understand than that in the conventional reduced-form-based indirect approach. Then, we show how to estimate all SF parameters jointly (in a single step) with method-of-moment estimators. As it turns out, using only unconditional moments, but not any conditional moments, greatly simplifies the identification and estimation issues, and makes light work of conveying the essential ideas involved.

Key Words: methods of moments, linear simultaneous equations, system estimation.

Running Head: Methods-of-Moments for Simultaneous Equations.

*The author is grateful to the Associate Editor and two anonymous referees for their detailed comments.
1 Introduction

Linear simultaneous (equation) systems (LSS) are widely used in econometrics, both micro and macro, ever since the pioneering contributions from the Cowles Commission. Discussions of LSS routinely appear in almost all econometrics textbooks; e.g., Amemiya (1985), Davidson and Mackinnon (1993), and Green (2003). Despite this popularity, however, the way the identification and estimation of LSS are discussed in the literature is mostly “archaic”: unnecessarily long-winded (for identification) and more parametric than the modern econometric trend (for estimation). The goal of this expository paper is to provide a modern review of LSS, both for identification and estimation, based on a methods-of-moment viewpoint which has appeared in a number of papers and books.

In doing so, we use only unconditional moments in the form \( E(\text{instruments} \times \text{error}) = 0 \). This greatly simplifies the presentation and makes light work of conveying the essential ideas involved without getting bogged down with various scenarios when conditional moments are considered. It is hoped that our (unconditional) method-of-moment view of LSS helps understand LSS with more ease and thus facilitates the applications.

For identification, we will come up with a rank condition, which is obtained “naturally” once a LSS and a moment condition are given. For estimation, under the given unconditional moment condition, generalized method of moments (GMM; Hansen, 1982) is the most efficient estimator as shown by Chamberlain (1987), which settles the issue of which estimation method to use. As GMM is a two-stage procedure using the usual linear-projection-based instrumental variable estimator (IVE) in the first stage, we will need to discuss only IVE and GMM for the estimation of LSS. This way, we also part from another old textbook discussion on the so-called ‘Gauss-Markov’ efficiency of the least-squares-type estimators (LSE) which require fixed regressors and unbiasedness.

The rest of this paper is organized as follows. Section 2 introduces notations and sets up LSS to be used in the rest of the paper. Section 3 examines the identification issue for LSS. Section 4 examines method-of-moment estimation of LSS. Finally, Section
2 Notations and Moment Conditions

Consider $H$-many linear equations:

$$y_{hi} = x_{hi}' \beta_h + u_{hi} \quad \iff \quad y_{hi} - x_{hi}' \beta_h = u_{hi}, \quad i = 1, \ldots, N \text{ and } h = 1, \ldots, H$$

where $y_{hi}$ is a response variable, $x_{hi}$ is a $k_h \times 1$ regressor vector, $\beta_h$ is a conformably defined vector of structural form (SF) parameters, and $u_{hi}$ is an error term. With some components of $x_{hi}$ possibly endogenous, suppose we adopt an unconditional moment condition

$$E(x_i u_{hi}) = 0 \quad \forall h$$

where $x_i$ is the $K \times 1$ system exogenous regressors consisting of elements of $x_{1i}, \ldots, x_{Hi}$; unity is included in $x_i$, and each exogenous element of $x_{1i}, \ldots, x_{Hi}$ appears no more than once in $x_i$. Assume that all random variables have finite second moments, $E(x_i x_i')$ is of full rank, and $\{(x_{hi}, y_{hi}), \ h = 1, \ldots, H\}$ is independent and identically distributed (iid) across $i = 1, \ldots, N$.

Define the stacked versions of the random variables and parameters: with $k \equiv k_1 + \ldots + k_H$,

$$y_i \equiv \begin{bmatrix} y_{1i} \\ \vdots \\ y_{Hi} \end{bmatrix}_{H \times 1}, \quad u_i \equiv \begin{bmatrix} u_{1i} \\ \vdots \\ u_{Hi} \end{bmatrix}_{H \times 1}, \quad w_i \equiv \begin{bmatrix} x_{1i} \\ \vdots \\ 0 \\ 0 \end{bmatrix}_{k \times H}, \quad \gamma \equiv \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_H \end{bmatrix}_{k \times 1},$$

where the numbers below a matrix denote its dimension. With these notations, the LSS can be written simply as

$$y_i = w_i' \gamma + u_i, \quad i = 1, \ldots, N$$

as if we had only one equation to deal with.

Rewrite the moment conditions $E(u_{hi} x_i) = 0 \forall h$ compactly as

$$E(u_i \otimes x_i) = \begin{bmatrix} E(u_{1i} x_i) \\ \vdots \\ E(u_{Hi} x_i) \end{bmatrix}_{(HK) \times 1} = 0.$$
Define

$$z_i \equiv I_H \otimes x_i = \begin{bmatrix} x_i & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & x_i \end{bmatrix} \quad \Rightarrow \quad z_iu_i = \begin{bmatrix} x_iu_{1i} \\ \vdots \\ x_iu_{Hi} \end{bmatrix}.$$ 

Hence, the (system) moment condition $E(u_i \otimes x_i) = 0$ is equivalent to

$$E(z_iu_i) \{ = E(\text{instruments} \times \text{error}) \} = 0.$$

### 3 Identification Conditions for LSS

To motivate our approach for identification, first we quickly review the classical textbook introduction of the rank condition of identification as can be seen in, e.g., Amemiya (1985) and Green (2003). In the LSS, denote the endogenous regressors as $y_j’s$ to rewrite the LSS as

$$\Gamma_H y_{H \times 1} - B_{H \times K} x_{K \times 1} = u_{H \times 1}$$

(3.1)

where $\Gamma$ and $B$ are the matrix of SF parameters, and $\Gamma$ is invertible. Sometimes we call $\Gamma$ ‘the endogenous SF parameters’, and $B$ ‘the exogenous SF parameters’. Solving the SF equations for $y_i$, we get the reduced forms (RF)

$$y_i = \Gamma^{-1}Bx_i + \Gamma^{-1}u_i = \Pi x_i + v_i,$$

where $\Pi \equiv \Gamma^{-1}B$, $v_i \equiv \Gamma^{-1}u_i$

where $\Pi$ is the matrix of RF parameters, and $v_i$ is the RF error vector.

Traditionally, the identification issue in LSS has been centered around the relation between the SF and RF parameters embodied in $\Pi = \Gamma^{-1}B$. Specifically, given that $\Pi$ is identified, the issue has been how to identify $\Gamma$ and $B$; i.e., whether the equation determines uniquely $\Gamma$ and $B$ or not. But this way of looking at identification is indirect through $\Pi$, and this seems to be a legacy of the old times where only LSE is feasible (for the RF’s). Although this kind of indirect identification seems unavoidable for some nonlinear models (e.g., Lee and Kimhi, 2005), the indirect identification of the SF parameters through the RF parameters is unwarranted for LSS because the identification can be done directly in a single step as shown in the following. Because
of the iid condition, we will often omit the subscript $i$ indexing individuals in the remainder of this paper.

Observe

$$E(zu) = E\{z \cdot (y - w'\gamma)\} = 0 \iff E(zw')\gamma = E(zy) \tag{3.2}$$

$$\implies \gamma = \{E(wz')E^{-1}(zz')E(zw')\}^{-1} \cdot E(wz')E^{-1}(zz')E(zy) \tag{3.3}$$

if the $k \times k$ matrix $E(wz')E^{-1}(zz')E(zw')$ is invertible where $E^{-1}(\cdot)$ denotes $\{E(\cdot)\}^{-1}$.

To see when this is invertible, note

$$wz'_{k \times HK} = \begin{bmatrix} x_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & x_H \end{bmatrix} \cdot \begin{bmatrix} x' & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & x' \end{bmatrix} = \begin{bmatrix} x_1x' & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & x_Hx' \end{bmatrix};$$

$wz'$ is of dimension $(k \times H) \cdot (H \times HK) = k \cdot HK$, which is block-diagonal. $E(zz')$ is also block diagonal. Thus $E(wz')E^{-1}(zz')E(zw')$ is block-diagonal as well:

$$\text{diag}\{E(x_1x')E^{-1}(xx')E(xx'), ..., E(x_Hx')E^{-1}(xx')E(xx_H')\}.$$

For this to be invertible, each diagonal matrix should be invertible, for which it is necessary and sufficient to have

$$\text{rank}\{E(x_hx')\} = k_h \text{ (requiring } k_h \leq K) \quad \forall h.$$

This is the ‘rank condition of identification’, whereas ‘$k_h \leq K$’ is the ‘order condition of identification’; the order condition is only a necessary condition.

The following cases occur:

- $k_h < K$: equation $h$ is ‘over-identified’ (more instruments than necessary)
- $k_h = K$: equation $h$ is ‘just-identified’ (just enough instruments)
- $k_h > K$: equation $h$ is ‘under-identified’ (not enough instruments).

This exposition of the rank and order conditions for LSS identification is much simpler than the typical econometric textbook exposition as in the discussion following (3.1).
The order condition is easy to check in practice. Consider equation $h$ in (3.1). Suppose it has $k_h^{EXO}$ exogenous regressors and $k_h - k_h^{EXO}$ endogenous regressors. Then the order condition is, as well known,

$$k_h \leq K \iff k_h - k_h^{EXO} \leq K - k_h^{EXO} \iff$$

$$\# \text{ endo. regressors in eq. } h \leq \# \text{ exo. regressors excluded from eq. } h. \tag{3.4}$$

The difficulty in applying the order condition in practice is not verifying (3.4), but finding plausible exclusion restrictions. Lee and Chang (2007) provide an answer to this problem using ratios of RF parameters.

In contrast to the order condition, the rank condition is “tricky” to check, because, for an estimator $A_N$ consistent for a constant matrix $A$, it can happen that $\text{rank}(A_N) \neq A$. For instance, suppose (Andrews, 1987)

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad A_N = \begin{bmatrix} 1 & N^{-1} \\ N^{-1} & \tau N^{-2} \end{bmatrix} \implies |A_N| = \frac{\tau - 1}{N^2}. \tag{3.5}$$

Clearly $A_N \to^p A$, but unless $\tau = 1$, $\text{rank}(A_N) = 2 > 1 = \text{rank}(A)$.

In practice, the opposite case $\text{rank}(A_N) < \text{rank}(A)$ would be more prevalent. For instance, consider a dummy variable $d$ with an extremely small $P(d = 1)$. In a given sample, all observations may have $d = 0$. In this case, if there is a single endogenous regressor in equation $h$ and $d$ is the single exogenous regressor excluded from equation $h$, then although (3.4) holds, the rank condition fails in the sample whereas it does not in the population. Equally worrisome is only a couple of individuals having $d = 1$ in the sample. Although the rank condition holds in the sample, this case is close to (3.5) in the sense that the rank holds up only barely in the sample.

In going from equation (3.2) to (3.3), we multiplied (3.2) by $E(wz')E^{-1}(zz')$. In fact, we could have used another matrix of the same dimension, say $L$, so long as $L \cdot E(zw')$ is invertible. It might look as if the above identification result holds only for the particular choice $L = E(wz')E^{-1}(zz')$. That this is not the case can be shown using the following fact for a system of equations.
Consider a system of \( HK \) equations with \( k \) unknowns:

\[
A^{(HK) \times k} \gamma = c^{HK \times 1}.
\]

For these equations to have any solution, \( c \) has to be in the column space of \( A \), as \( A \gamma \) is a linear combination of the columns of \( A \). Then the equation system is said to be ‘consistent’. Given that the equation system is consistent, a necessary and sufficient condition for the solution to be unique is

\[
\text{rank}(A) = k
\]

\[\iff\]

\[
\text{rank}\{E(x_h x')\} = k_h \forall h \text{ for (3.2) with } A = E(zw') \text{ and } c = E(zy)
\]

see, e.g., Searle (1982, p.233) or Schott (2005, p.227). We state this identification result as a theorem:

**Theorem 1.** For a linear simultaneous equation system \( y_i = w_i' \gamma + u_i \) consisting of \( H \)-many structural form equations where \( w_i = \text{diag}(x_{1i}, \ldots, x_{Hi}) \), suppose that all second moments of the random variables are finite, that \( E(zu) = 0 \) holds where \( z \equiv I_H \otimes x \) and \( x \) consists of all exogenous elements of \( x_1, \ldots, x_H \), and that \( E(xx') \) is of full rank. Then, given that there is any solution to \( E(zw') \gamma = E(zy) \) which is equivalent to \( E(zu) = 0 \), the solution is unique (that is, \( \gamma \) is uniquely identified) iff

\[
\text{rank}\{E(x_h x')\} = k_h \text{ (requiring } k_h \leq K) \ \forall h.
\]

4 **Method-of-Moment Estimation of LSS**

Turning to estimation of LSS, as already noted, GMM is the efficient estimator under a given unconditional moment condition. But GMM is a two-stage procedure with a \( \sqrt{N} \)-consistent initial estimator needed to estimate the variance matrix of \( N^{-1/2} \sum_i z_i u_i \) in the first stage. Equation (3.3) motivates such an estimator, which is nothing but the usual IVE. In the following, we call the IVE applied to LSS ‘system IVE’ and the IVE applied to each equation separately ‘separate IVE’. Analogous names for GMM are ‘System GMM’ and ‘separate GMM’.
Assuming that we have only over- or just-identified equations in the system, the system IVE $g_{ive}$ for $\gamma$ is

$$g_{ive} = \left\{ \sum_i w_i z_i' \left( \sum_i z_i z_i' \right)^{-1} \sum_i z_i w_i' \right\} \cdot \sum_i w_i z_i' \left( \sum_i z_i z_i' \right)^{-1} \sum_i z_i y_i$$

$$= \gamma + \left\{ \sum_i w_i z_i' \left( \sum_i z_i z_i' \right)^{-1} \sum_i z_i w_i' \right\} \cdot \sum_i w_i z_i' \left( \sum_i z_i z_i' \right)^{-1} \sum_i z_i y_i \quad \text{using} \quad y_i = w_i' \gamma + u_i.$$

From this, the consistency of $g_{ive}$ for $\gamma$ follows immediately, and the asymptotic distribution is ($'\approx'$ denotes convergence in law)

$$\sqrt{N} (g_{ive} - \gamma) \approx N(0, A \cdot E(z w u') \cdot A')$$

where

$$A \equiv \left\{ E(w z' \cdot E^{-1}(z z') \cdot E(z w')) \right\} \cdot E(w z') E^{-1}(z z').$$

Observe

$$E(z y) = \begin{bmatrix} E(x y_1) \\
\vdots \\
E(x y_H) \end{bmatrix} \quad \text{and} \quad E(w z') E^{-1}(z z') E(z y) = \begin{bmatrix} E(x_1 x') E^{-1}(x x') E(x y_1) \\
\vdots \\
E(x_H x') E^{-1}(x x') E(x y_H) \end{bmatrix}.$$}

The population moments show that

$$g_{ive} = \begin{bmatrix}
\left\{ \sum_i x_{1i} x_{i}' \left( \sum_i x_i x_i' \right)^{-1} \sum_i x_i x_{1i}' \right\}^{-1} \cdot \sum_i x_{1i} x_i' \left( \sum_i x_i x_i' \right)^{-1} \sum_i x_i y_{1i} \\
\vdots \\
\left\{ \sum_i x_{Hi} x_{i}' \left( \sum_i x_i x_i' \right)^{-1} \sum_i x_i x_{Hi}' \right\}^{-1} \cdot \sum_i x_{Hi} x_i' \left( \sum_i x_i x_i' \right)^{-1} \sum_i x_i y_{Hi} \end{bmatrix}.$$}

Thus, the system IVE is nothing but the stacked version of the separate IVE’s $b_{h,ive}$, $h = 1, \ldots, H$, which are on the right-hand side of this display. Hence, there is no efficiency gain in the system IVE over the separate IVE’s. Wooldridge (2002) mentions this point without showing the derivation. This equivalence holds despite no restriction whatsoever put on the relationship among $u_{hi}$, $h = 1, \ldots, H$; i.e., dependence among these errors is neither disallowed nor specified. The advantage of the system IVE over the separate IVE is that the system IVE yields asymptotic covariances between $b_{h,ive}$ and $b_{j,ive}$ for all $h \neq j$; also it can be used to impose cross-equation restrictions.

The system GMM corresponding to the system IVE is obtained by taking one step
from \( g_{ive} \). Define the residuals \( \hat{u}_i \equiv y_i - w'_i g_{ive} \) to get

\[
g_{gmm} = \left\{ \sum_i w_i z_i' \left( \sum_i z_i \hat{u}_i \hat{u}'_i \right)^{-1} \sum_i z_i w_i \right\}^{-1} \left( \sum_i w_i z_i' \left( \sum_i z_i \hat{u}_i \hat{u}'_i \right)^{-1} \sum_i z_i y_i, \right.
\]

\[
\sqrt{N}(g_{gmm} - \gamma) \sim N(0, \{E(w'z')E^{-1}(zu'z')E(zw')\}^{-1}),
\]

\[
\frac{1}{\sqrt{N}} \sum_i \hat{u}_i z_i' \left( \frac{1}{N} \sum_i z_i \hat{u}_i \hat{u}'_i z_i' \right)^{-1} \left( \frac{1}{\sqrt{N}} \sum_i z_i \hat{u}_i \right) \sim \chi^2_{HK-k}, \quad \hat{u}_i \equiv y_i - w'_i g_{gmm}
\]

which is the GMM over-identification test statistic. The system GMM shows the form of the best matrix to multiply to (3.2) to solve (3.2) for \( \gamma \), where ‘best’ means ‘efficient’.

In contrast to the system GMM \( g_{gmm} \), the separate GMM is

\[
b_{n,gmm} = \left\{ \sum_i x_{hi} x_{hi}' \left( \sum_i x_i x_i' \hat{u}_h \hat{u}'_h \right)^{-1} \sum_i x_i x_{hi}' \right\}^{-1} \left( \sum_i x_{hi} x_{hi}' \left( \sum_i x_i x_i' \hat{u}_h \hat{u}'_h \right)^{-1} \sum_i x_i y_{hi}, \right. \forall h
\]

where \( \hat{u}_h = y_{hi} - x_{hi}' b_{n,ive} \). One may wonder whether the system GMM is more efficient than the separate GMM. Differently from the system IVE versus the separate IVE, the answer is positive in general as shown in the following. The system GMM appeared in Wooldridge (2002) and Green (2003).

The inverse of the GMM asymptotic variance matrix is

\[
\begin{bmatrix}
E(x_1 x') & 0 & 0 \\
0 & \ddots & 0 \\
0 & 0 & E(x_H x')
\end{bmatrix}
\begin{bmatrix}
E(x x' u_1^2) & \cdots & E(x x' u_1 u_H) \\
\vdots & \ddots & \vdots \\
E(x x' u_H u_1) & \cdots & E(x x' u_H^2)
\end{bmatrix}^{-1}
\begin{bmatrix}
E(x x') & 0 & 0 \\
0 & \ddots & 0 \\
0 & 0 & E(x x'_H)
\end{bmatrix}.
\]

If

\[
E(x x' u_h u_j) = 0 \forall h \neq j,
\]

then, all off-diagonal terms in the middle matrix are zero, and the system GMM asymptotic variance becomes diagonal:

\[
\text{diag}\{E(x_h x')E^{-1}(x x' u_h^2)E(x x'_h)\}^{-1}, \quad h = 1, \ldots, H
\]

This is the same as the asymptotic variance of the separate GMM’s. If the off-diagonal terms are not zero, then differently from the system IVE, the system GMM is not in general equal to the separate GMM’s. Having \( \sum_i z_i \hat{u}_i \hat{u}'_i z_i' \) in the system GMM that is not block-diagonal, instead of the block-diagonal \( \sum_i z_i z_i' \) in the system IVE, makes the difference.
The negation of $E(xx'uhuj) = 0 \forall h \neq j$, however, is not sufficient for the system GMM efficiency gain. Lee (2004) shows that, for the efficiency gain of equation $h$ in the system GMM over the separate GMM, in addition to $E(xx'uhuj) \neq 0$ for some $j$, it is necessary to have at least one over-identified SF, say $SF$ $m$ with $m \neq h$. An analogous condition in fact appeared in the literature under the traditional homoskedasticity framework: Hausman (1983, p.413) states that there is an efficiency gain in system estimation over the separate estimation of SF $h$ if $E(uhuj) \neq 0$ for some $j \neq h$ with SF $j$ over-identified.

Finally, we note that the system GMM is applicable to panel data. If we have panel data with $H$ ‘waves’ (i.e., each subject is observed for $H$-many periods), then there will be $H$ equations—one for each period. In this case, typically, different instruments are used for different equations, say $z_h$ for SF $h$, to result in the moment conditions

$$E(uhz_h) = 0, \quad h = 1, \ldots, H.$$  

This case can be dealt with simply by setting $z_i = \text{diag}(z_{1i}, \ldots, z_{Hi})$ in the above IVE and GMM. See Lee (2002) for this type of panel data IVE and GMM.

5 Conclusions

This paper provided a review on the modern method-of-moment-based approach to identification and estimation of linear simultaneous equation systems. First, we derived the rank and order conditions for the structural form (SF) parameter identification. Then we showed how to estimate the SF parameters jointly with the system IVE and then system GMM. Throughout our discussion, only unconditional moment conditions of the form $E(\text{instruments} \times \text{error}) = 0$ have been used—no homoskedasticity, no independence, neither any parametric distributions was specified or used. This narrowed scope simplified much our exposition, and made it possible to convey the essential ideas of linear simultaneous system identification and estimation without “distractions”.

10
REFERENCES

Amemiya, T., 1985, Advanced econometrics, Harvard University Press