SIMULTANEOUS EQUATIONS SYSTEMS, I

Econometric Analysis of Cross Section and Panel Data, 2e
MIT Press
Jeffrey M. Wooldridge

1. Introduction
2. Identification
3. Estimation After Identification

1. INTRODUCTION

- We now study a linear set of equations that determines jointly a set of G outcomes, where endogenous variables may appear on the right hand side with exogenous variables.
- Actually, only the statistical structure is important for identification and estimation; the system could have omitted variables, for example. But we are mostly interested in simultaneous equations models.
- The leading cases are demand and supply systems, which jointly determine quantities and prices.

EXAMPLE: Labor Supply-Wage Offer system for married women. In equilibrium, write the system as

$$
\begin{aligned}
h &= \gamma_1 w + \delta_{11} exper + \delta_{12} exper^2 + \delta_{13} othinc + \delta_{14} kids + u_1 \\
w &= \gamma_2 h + \delta_{21} exper + \delta_{22} exper^2 + \delta_{23} educ + u_2,
\end{aligned}
$$

so other sources of income and number of children affect labor supply but not the wage offer, and education affects the wage offer but not labor supply. (Omit intercepts for simplicity.)

• In this system, *h* and *w* are endogenous. Traditional SEM analysis would take everything else as exogenous. Nonlinearity in *exper* requires no special treatment.

• The idea of an underlying counterfactual is critical to sensible applications of SEMs. It makes sense to think of a demand function in isolation, and similarly with a supply function. They are brought together as a way of determining the observed data.

EXAMPLE: City Crime Rates and Size of Police Force:

$$\begin{aligned} crime &= \gamma_1 police + \delta_{11} age + \delta_{12} unem + \delta_{13} wage + u_1 \\ police &= \gamma_2 crime + \delta_{21} age + \delta_{22} unem + \delta_{23} wage + \delta_{24} election + u_2 \end{aligned}$$

• These two equations form a legitimate SEM: each equation stands on its own. In effect, they describe two different sides of a “market.” They come together as a system under assumptions about how the observed outcomes, $(crime_i, police_i)$.

• The next example is a poor application of SEMs. Suppose the population is all families in a particular country.

EXAMPLE: Joint Determination of Family Retirement Saving and Housing Expenditure:

$$\begin{aligned} retirement &= \gamma_1 housing + \delta_{11} inc + \delta_{12} educ + \delta_{13} age + u_1 \\ housing &= \gamma_2 retirement + \delta_{21} inc + \delta_{22} educ + \delta_{32} age + u_2 \end{aligned}$$

• Neither of these equations stands on its own. What would it mean, in the first equation, to study the effect of changing income on retirement holding housing expenditure fixed? Even if one wants to model the joint determination of y_1 and y_2 in this way, the parameters are not interesting. There is no interesting counterfactual.

- Suppose we now have a binary variable, *class*, which indicates whether an adult has taken a course in financial planning. Does not make sense to add this just to the retirement equation.
- Instead, should just analyze

$$retirement = \beta_{11}inc + \beta_{12}educ + \beta_{13}age + \beta_{14}class + u_1.$$

Maybe we need an instrument for *class* (perhaps via randomization), but this has nothing to do with a simultaneous equations model.

- What do you think of the following? For a firm operating in a foreign country, y_1 = R&D spending, y_2 = foreign technology purchases.

2. IDENTIFICATION

- Identification is a feature of a population. Sampling is a separate issue.
- It is true that certain sampling schemes can cause a lack of identification, but the population is always the best starting point.
- With random sampling (as we assume here – remember, in the cross section), it always makes sense to study the population model.

Two-Equation System

- General two-equation *structural system* (in the population):

$$
\begin{aligned}
y_1 &= \gamma_1 y_2 + \mathbf{z}_1 \boldsymbol{\delta}_1 + u_1 \\
y_2 &= \gamma_2 y_1 + \mathbf{z}_2 \boldsymbol{\delta}_2 + u_2
\end{aligned}
$$

where $\mathbf{z}_1$ is $1 \times M_1$ and $\mathbf{z}_2$ is $1 \times M_2$. Let $\mathbf{z}$ be $1 \times M$ contain all (nonredundant) exogenous variables

$$
E(\mathbf{z}' u_1) = E(\mathbf{z}' u_2) = \mathbf{0}
$$

where in almost all applications $\mathbf{z}_1$ and $\mathbf{z}_2$ (and therefore $\mathbf{z}$) include unity. We act as if that is true here, so that the *structural errors* u_1 and u_2 have zero means.

- γ_1, $\boldsymbol{\delta}_1$, γ_2, $\boldsymbol{\delta}_2$ are the *structural parameters*.
- The moment conditions imply that if a variable is exogenous in any equation, it is exogenous in all equations; this is the traditional starting point.
- Although we do not need them to study identification, we can obtain *reduced forms* for y_1 and y_2 if

$$\gamma_1\gamma_2 \neq 1.$$

- Generally, a reduced form expresses an endogenous variable as a function of exogenous variables and unobserved errors.
- In this case, solve the two equations for y_1 and y_2:

$$\begin{aligned} y_1 &= \gamma_1(\gamma_2 y_1 + \mathbf{z}_2\boldsymbol{\delta}_2 + u_2) + \mathbf{z}_1\boldsymbol{\delta}_1 + u_1 \\ &= \gamma_1\gamma_2 y_1 + \mathbf{z}_1\boldsymbol{\delta}_1 + \mathbf{z}_2\gamma_1\boldsymbol{\delta}_2 + u_1 + \gamma_1 u_2. \end{aligned}$$

- Therefore, if $\gamma_1\gamma_2 \neq 1$,

$$\begin{aligned} y_1 &= (1 - \gamma_1\gamma_2)^{-1}(\mathbf{z}_1\boldsymbol{\delta}_1 + \mathbf{z}_2\gamma_1\boldsymbol{\delta}_2 + u_1 + \gamma_1 u_2) \\ &\equiv \mathbf{z}\boldsymbol{\pi}_1 + v_1 \end{aligned}$$

where $\boldsymbol{\pi}_1$ is the $M \times 1$ vector of *reduced form parameters* and $v_1 = (1 - \gamma_1\gamma_2)^{-1}(u_1 + \gamma_1 u_2)$ is a *reduced form error*.

- We can do the same for y_2, so we have

$$
\begin{aligned}
y_1 &= \mathbf{z}\boldsymbol{\pi}_1 + v_1 \\
y_2 &= \mathbf{z}\boldsymbol{\pi}_2 + v_2.
\end{aligned}
$$

Both reduced form errors satisfy

$$E(\mathbf{z}'v_1) = E(\mathbf{z}'v_2) = \mathbf{0},$$

which means $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$ can be consistently estimated by OLS on a random sample provided $E(\mathbf{z}'\mathbf{z})$ is nonsingular. (SUR estimation would not improve estimation of the RF parameters. Why?)

• We can always consistently estimate the RF parameters. When are the structural parameters identified

• Identification in the two-equation case is straightforward. Consider identification of the first structural equation. Write it with the RF of y_2:

$$
\begin{aligned}
y_1 &= \gamma_1 y_2 + \mathbf{z}_1 \boldsymbol{\delta}_1 + u_1 \\
y_2 &= \mathbf{z} \boldsymbol{\pi}_2 + v_2
\end{aligned}
$$

- Because y_2 is the only endogenous explanatory variable, we need at least one instrument for it. That means we must have something in $\mathbf{z}$ in the RF with a nonzero coefficient that is not also in $\mathbf{z}_1$.
- But $y_2 = \gamma_2 y_1 + \mathbf{z}_2\boldsymbol{\delta}_2 + u_2$ and so $\boldsymbol{\pi}_2$ has a nonzero coefficient on something not in $\mathbf{z}_1$ if and only if there is at least one element of $\mathbf{z}_2$ that is not also in $\mathbf{z}_1$ with nonzero coefficient in $\boldsymbol{\delta}_2$.

• So, we can read identification of each equation off of the structural system:

$$\begin{aligned} y_1 &= \gamma_1 y_2 + \mathbf{z}_1 \boldsymbol{\delta}_1 + u_1 \\ y_2 &= \gamma_2 y_1 + \mathbf{z}_2 \boldsymbol{\delta}_2 + u_2 \end{aligned}$$

The first equation is identified if and only if there is at least one element in $\mathbf{z}_2$ not in $\mathbf{z}_1$ with a nonzero coefficient (element of $\boldsymbol{\delta}_2$) in the second equation. Similarly, the second equation is identified if and only if there is something in $\mathbf{z}_1$ not in $\mathbf{z}_2$ with a corresponding nonzero element in $\boldsymbol{\delta}_1$.

EXAMPLE: In the system

$$
\begin{aligned}
h &= \gamma_1 w + \delta_{11} exper + \delta_{12} exper^2 + \delta_{13} othinc + \delta_{14} kids + u_1 \\
w &= \gamma_2 h + \delta_{21} exper + \delta_{22} exper^2 + \delta_{23} educ + u_2,
\end{aligned}
$$

the labor supply function is identified if and only if $\delta_{23} \neq 0$. The wage offer function is identified if and only if at least one of δ_{13} and δ_{14} is different from zero.

• Important: Our imposing of exclusion restrictions means that it must be the case that *educ* is legitimately excluded from the supply equation and *othinc* and *kids* are properly excluded from the wage offer equation.

- Estimation of each equation could proceed by 2SLS.
- Key Point: In general, identification of any particular equation of an SEM depends on the structure of *other* equations in the SEM.

General Linear System

• Up to us to make sure the system makes economic sense. From there, identification analysis is purely mechanical.

• Notation: let $\mathbf{y} = (y_1, y_2, \ldots, y_G)$ be a $1 \times G$ vector of endogenous variables, $\mathbf{z} = (z_1, z_2, \ldots, z_M)$ be a $1 \times M$ vector of exogenous variables, and $\mathbf{u} = (u_1, u_2, \ldots, u_G)$ a $1 \times G$ vector of structural errors. Write the system as

$$
\begin{aligned}
\mathbf{y}\boldsymbol{\gamma}_1 + \mathbf{z}\boldsymbol{\delta}_1 + u_1 &= 0 \\
\mathbf{y}\boldsymbol{\gamma}_2 + \mathbf{z}\boldsymbol{\delta}_2 + u_2 &= 0 \\
&\vdots \\
\mathbf{y}\boldsymbol{\gamma}_G + \mathbf{z}\boldsymbol{\delta}_G + u_G &= 0
\end{aligned}
$$

or

$$
\begin{aligned}
\gamma_{11}y_1 + \gamma_{12}y_2 + \ldots + \gamma_{1G}y_G + \delta_{11}z_1 + \delta_{12}z_2 + \ldots + \delta_{1M}z_M + u_1 &= 0 \\
\gamma_{21}y_1 + \gamma_{22}y_2 + \ldots + \gamma_{2G}y_G + \delta_{21}z_1 + \delta_{22}z_2 + \ldots + \delta_{2M}z_M + u_2 &= 0 \\
&\vdots \\
\gamma_{G1}y_1 + \gamma_{G2}y_2 + \ldots + \gamma_{GG}y_G + \delta_{G1}z_1 + \delta_{G2}z_2 + \ldots + \delta_{GM}z_M + u_G &= 0
\end{aligned}
$$

• Note that this setup allows for traditional specification of supply and demand, where y_1 (quantity, say) can be on the left hand side of both equations with y_2 (price) on the RHS in both equations ($G = 2$).

• In each equation g, $\boldsymbol{\gamma}_g$ is $1 \times G$ and $\boldsymbol{\delta}_g$ is $1 \times M$.

• The vector $\mathbf{z}$ is exogenous in all equations:

$$E(\mathbf{z}'u_g) = \mathbf{0},\, g = 1,, \ldots, G,$$

and we assume that $E(\mathbf{z}'\mathbf{z})$ is nonsingular.

• There are $G(G + M)$ structural parameters, but we have only GM moment conditions. We need a lot more information via restrictions on the parameters.

• Write the system of all G equations as

$$\mathbf{y}\boldsymbol{\Gamma} + \mathbf{z}\boldsymbol{\Delta} + \mathbf{u} = \mathbf{0}$$

$$\underset{G\times G}{\boldsymbol{\Gamma}} = \begin{pmatrix} \boldsymbol{\gamma}_1 & \boldsymbol{\gamma}_2 & \cdots & \boldsymbol{\gamma}_G \end{pmatrix}$$

$$\underset{M\times G}{\boldsymbol{\Delta}} = \begin{pmatrix} \boldsymbol{\delta}_1 & \boldsymbol{\delta}_2 & \cdots & \boldsymbol{\delta}_G \end{pmatrix}$$

• We assume that $\boldsymbol{\Gamma}$ is nonsingular, so that a reduced form exists.

Define the $G \times G$ variance-covariance matrix of the structural errors as

$$\boldsymbol{\Sigma} = E(\mathbf{u}'\mathbf{u}).$$

• The reduced form for all G endogenous variables is obtained as

$$\begin{aligned}
\mathbf{y\Gamma} &= -\mathbf{z\Delta} - \mathbf{u} \\
\mathbf{y} &= \mathbf{z}(-\mathbf{\Delta\Gamma}^{-1}) + \mathbf{u}(-\mathbf{\Gamma}^{-1}) \equiv \mathbf{z\Pi} + \mathbf{v} \\
\mathbf{\Pi} &\equiv -\mathbf{\Delta\Gamma}^{-1} \\
\mathbf{v} &\equiv \mathbf{u}(-\mathbf{\Gamma}^{-1}).
\end{aligned}$$

Define the $G \times G$ variance-covariance matrix of the reduced form errors:

$$\mathbf{\Lambda} \equiv E(\mathbf{v}'\mathbf{v}) = \mathbf{\Gamma}^{-1\prime}\mathbf{\Sigma\Gamma}^{-1}.$$

• Because $E(\mathbf{z}'\mathbf{v}) = \mathbf{0}$ and $E(\mathbf{z}'\mathbf{z})$ is nonsingular, $\mathbf{\Pi}$ and $\mathbf{\Lambda}$ are always identified. That is, the reduced form parameters are identified.

• When can we recover the structural parameters, $\mathbf{\Gamma}$, $\mathbf{\Delta}$, and $\mathbf{\Sigma}$ from the RF parameters? (Sometimes $\mathbf{\Sigma}$ is of interest, but not usually. If we can estimate all elements of $\mathbf{\Gamma}$ then we can estimate $\mathbf{\Sigma} = \mathbf{\Gamma}'\mathbf{\Lambda}\mathbf{\Gamma}$.)

• Let $\mathbf{F}$ be any $G \times G$ nonsingular matrix. Postmultiply the structural system by $\mathbf{F}$:

$$\mathbf{y}\mathbf{\Gamma}\mathbf{F} + \mathbf{z}\mathbf{\Delta}\mathbf{F} + \mathbf{u}\mathbf{F} = \mathbf{0}$$

or

$$\mathbf{y}\mathbf{\Gamma}^{*} + \mathbf{z}\mathbf{\Delta}^{*} + \mathbf{u}^{*} = \mathbf{0}$$

• The RF of the new system is easily seen to be the same as the RF of the original system:

$$\begin{aligned}
\mathbf{y}\boldsymbol{\Gamma}^* &= -\mathbf{z}\boldsymbol{\Delta}^* - \mathbf{u}^* \\
\mathbf{y} &= \mathbf{z}(-\boldsymbol{\Delta}^*\boldsymbol{\Gamma}^{*-1}) + \mathbf{u}^*(-\boldsymbol{\Gamma}^{*-1}) \\
&= \mathbf{z}(-\boldsymbol{\Delta}\mathbf{F}\mathbf{F}^{-1}\boldsymbol{\Gamma}^{-1}) + \mathbf{u}(-\mathbf{F}\mathbf{F}^{-1}\boldsymbol{\Gamma}^{-1}) \\
&= \mathbf{z}(-\boldsymbol{\Delta}\boldsymbol{\Gamma}^{-1}) + \mathbf{u}(-\boldsymbol{\Gamma}^{-1})
\end{aligned}$$

• The transformed system is called an *equivalent structure* because it is indistinguishable from the original system.

• Our only hope is to place restrictions on $(\mathbf{\Gamma}, \mathbf{\Delta}, \mathbf{\Sigma})$ so that there are no equivalent structures.

• Define the $(G + M) \times G$ matrix of structural parameters on endogenous and exogenous variables:

$$\underset{(G+M)\times G}{\mathbf{B}} = \begin{pmatrix} \mathbf{\Gamma} \\ \mathbf{\Delta} \end{pmatrix}.$$

• A $G \times G$ nonsingular matrix $\mathbf{F}$ is an *admissible linear structure if*

1. $\mathbf{BF}$ satisfies all of the restrictions $\mathbf{B}$ does.
2. $\mathbf{F}'\mathbf{\Sigma}\mathbf{F}$ satisfies all the restrictions $\mathbf{\Sigma}$ does.

- Need enough restrictions on $\mathbf{B}$ and $\mathbf{\Sigma}$ so that the only admissible linear structure is $\mathbf{F} = \mathbf{I}_G$.
- In cross section settings, $\mathbf{\Sigma}$ is rarely restricted, and so for now we ignore the second requirement. Proceed as if we only place restrictions on $\mathbf{B}$.
- We consider identification of the first equation, assuming no cross-equation restrictions (which are rare in true SEMs with autonomous equations):

$$\mathbf{y}\boldsymbol{\gamma}_1 + \mathbf{z}\boldsymbol{\delta}_1 + u_1 = 0$$

- First impose a *normalization restriction* that one of the elements in $\boldsymbol{\gamma}_1$ equals –1. This defines the LHS variable. For example, if $\gamma_{11} = -1$,

$$y_1 = \gamma_{12}y_2 + \ldots + \gamma_{1G}y_G + \delta_{11}z_1 + \ldots + \delta_{1M}z_M + u_1$$

- The normalization restriction means that we cannot get an admissible structure by multiplying through by any nonzero constant.

- Let

$$\underset{(G+M)\times 1}{\boldsymbol{\beta}_1} = \begin{pmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\delta}_1 \end{pmatrix}$$

and assume that we have imposed the normalization restriction. In addition, we have restrictions on $\boldsymbol{\beta}_1$ of the form

$$\mathbf{R}_1\boldsymbol{\beta}_1 = \mathbf{0},$$

where $\mathbf{R}_1$ is $J_1 \times (G + M)$.

• Often called *homogeneous linear restrictions*, but, because of the normalization restriction, it actually allows nonhomogeneous restrictions by choosing $\mathbf{R}_1$ appropriately.

EXAMPLE: $G = 3$, $M = 4$

$$y_1 = \gamma_{12} y_2 + \gamma_{13} y_3 + \delta_{11} z_1 + \delta_{12} z_2 + \delta_{13} z_3 + \delta_{14} z_4 + u_1$$

$$\gamma_{12} = 0, \delta_{13} + \delta_{14} = 3$$

There are $J_1 = 2$ restrictions (in addition to $\gamma_{11} = -1$).

$$\mathbf{R}_1 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

$$\boldsymbol{\beta}_1' = (-1, \gamma_{12}, \gamma_{13}, \delta_{11}, \delta_{12}, \delta_{13}, \delta_{14})$$

• When are the restrictions defined by $\mathbf{R}_1$, plus normalization, enough to identify $\boldsymbol{\beta}_1$?

• Let $\mathbf{F} = \begin{pmatrix} \mathbf{f}_1 & \mathbf{f}_2 & \cdots & \mathbf{f}_G \end{pmatrix}$ be a potential admissible structure. Then the first column of $\mathbf{B}^* = \mathbf{BF}$ is $\boldsymbol{\beta}_1^* = \mathbf{B}\mathbf{f}_1$. Now

$$\mathbf{R}_1\boldsymbol{\beta}_1^* = \mathbf{R}_1(\mathbf{B}\mathbf{f}_1) = (\mathbf{R}_1\mathbf{B})\mathbf{f}_1 = \mathbf{0}$$

holds if $\mathbf{f}_1' = \mathbf{e}_1' = \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}$ because $\mathbf{B}\mathbf{e}_1 = \boldsymbol{\beta}_1$. And, the restrictions hold for any scalar multiple of $\mathbf{e}_1$. But we will be able to rule out scalar multiples of $\mathbf{e}_1$, other than $\mathbf{e}_1$ itself, by normalization.

- Need to conclude that the dimension of the null space of $\mathbf{R}_1\mathbf{B}$, that is, the set of G vectors satisfying $(\mathbf{R}_1\mathbf{B})\mathbf{f}_1 = \mathbf{0}$, is one. Because $\mathbf{R}_1\mathbf{B}$ has G columns, its null space has rank one if and only if

$$rank(\mathbf{R}_1\mathbf{B}) = G - 1.$$

- This is the important *rank condition* for identification. With the normalization it is necessary and sufficient for identification of $\boldsymbol{\beta}_1$. We must study how the restrictions for the first equation act on the rest of the system.

- Necessary but not sufficient is the *order condition*: the number of rows of $\mathbf{R}_1$ is large enough so that $\mathbf{R}_1\mathbf{B}$ *could* have rank $G - 1$:

$$J_1 \geq G - 1.$$

If $J_1 < G - 1$, there is no need to check the rank condition. If $J_1 \geq G - 1$, the rank condition might hold, but it needs to be checked.

- Go back to the equation without the normalization restrictions, $\mathbf{y}\boldsymbol{\gamma}_1 + \mathbf{z}\boldsymbol{\delta}_1 + u_1 = 0$. Makes sense to express the order condition as

$$J_1 + 1 \geq G,$$

so that the total number of restrictions in the equation is as large as the number of endogenous variables.

• After we specify the entire system, we can find

$\mathbf{R}_1\mathbf{B} = \begin{pmatrix} \mathbf{R}_1\boldsymbol{\beta}_1 & \mathbf{R}_1\boldsymbol{\beta}_2 & \cdots & \mathbf{R}_1\boldsymbol{\beta}_G \end{pmatrix}$, which has G columns. The first is, by definition, the zero vector. So, we must check to make sure the last $G - 1$ columns are linearly independent.

EXAMPLE: Consider the three-equation system

$$\begin{aligned}
y_1 &= \gamma_{12}y_2 + \gamma_{13}y_3 + \delta_{11}z_1 + \delta_{13}z_3 + u_1 \\
y_2 &= \gamma_{21}y_1 + \delta_{21}z_1 + u_2 \\
y_3 &= \delta_{31}z_1 + \delta_{32}z_2 + \delta_{33}z_3 + \delta_{34}z_4 + u_3
\end{aligned}$$

where all restrictions have been imposed.

• We have $G = 3$ and $M = 4$, and the 2×7 matrix that imposes the restrictions on the first equation is

$$\mathbf{R}_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

If $\mathbf{B}$ is the 7×3 matrix without any restrictions yet imposed, then

$$\mathbf{R}_1\mathbf{B} = \begin{pmatrix} \delta_{12} & \delta_{22} & \delta_{32} \\ \delta_{14} & \delta_{24} & \delta_{34} \end{pmatrix}.$$

- Now we impose all of the restrictions in the system:

$$\mathbf{R}_1\mathbf{B} = \begin{pmatrix} 0 & 0 & \delta_{32} \\ 0 & 0 & \delta_{34} \end{pmatrix}.$$

- Regardless of the values of δ_{32} and δ_{34}, $rank(\mathbf{R}_1\mathbf{B})$ is at most one, yet $G - 1 = 2$. Therefore, the first equation fails the order condition (even though the order condition just holds: $J_1 = 2$).

- The matrix imposing the restrictions of the second equation, $\gamma_{23} = 0, \delta_{22} = 0, \delta_{23} = 0$, and $\delta_{24} = 0$, is

$$\mathbf{R}_2 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

- With all restrictions on the entire system imposed,

$$\mathbf{R}_2\mathbf{B} = \begin{pmatrix} \gamma_{13} & 0 & -1 \\ 0 & 0 & \delta_{32} \\ \delta_{13} & 0 & \delta_{33} \\ 0 & 0 & \delta_{34} \end{pmatrix}.$$

- A sufficient condition for this matrix to have rank two is $\delta_{13} \neq 0$ and at least one of δ_{32} and δ_{34} different from zero. The rank condition fails if $\gamma_{13} = \delta_{13} = 0$, in which case y_1 and y_2 form a two equation system with only one exogenous variable, z_1, appearing in both equations
- The third equation is identified because it contains no endogenous explanatory variables.

• The most common situation is like in the previous example, where all restrictions are *exclusion restrictions*, in which case the restriction matrix consistst of zeros and ones. For the first equation, the number of rows in $\mathbf{R}_1$ is the number of endogenous variables excluded from the right hand side, $G - G_1 - 1$, where G_1 is the number of included endogenous variables on the RHS, plus the number of excluded exogenous variables from the first equation, $M - M_1$. So

$$J_1 = (G - G_1 - 1) + (M - M_1).$$

- Consequently, the order condition is

$$(G - G_1 - 1) + (M - M_1) \geq G - 1$$

or

$$M - M_1 \geq G_1.$$

- In other words, the number of exogenous variables not appearing in the first equation must be at least as large as the number of included RHS endogenous variables. We know this just from counting to see if we have enough potential instrumental variables. But even if we do, they partial correlations with the RHS endogenous variables may not be sufficient (rank condition).

Unidentified, Just Identified, and Overidentified

• If $rank(\mathbf{R}_1\mathbf{B}) < G - 1$ the first equation is *unidentified* or *underidentified*. This always happens if the order condition fails.

• If $J_1 = G - 1$ and the rank condition holds, we say equation one is *just identified*: we have just enough of the right restrictions to estimate $\boldsymbol{\beta}_1$.

• If $J_1 > G - 1$ and the rank condition holds, equation one is (potentially) *overidentified*, and $J_1 - (G - 1)$ is the number of *overidentifying restrictions*. (We say "potentially" because we might not really have overidentifying restrictions.)

EXAMPLE: Consider the $G = 2$, $M = 4$ system

$$y_1 = \gamma_{12}y_2 + \delta_{11}z_1 + \delta_{12}z_2 + \delta_{13}z_3 + \delta_{14}z_4 + u_1$$
$$y_2 = \gamma_{21}y_1 + \delta_{21}z_1 + \delta_{22}z_2 + u_2$$

The first equation fails the order condition, and is unidentified. The second equation satisfies the rank condition if at least one of $\delta_{13} \neq 0$ or $\delta_{14} \neq 0$. If δ_{13} and δ_{14} are both different from zero, there is one overidentifying restriction in equation two. But if, say, $\delta_{13} = 0$, the second equation is only just identified because there is only one instrument for y_1, z_4.

3. ESTIMATION AFTER IDENTIFICATION

- We can write the system as before

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} = \mathbf{u}_i$$

with IV matrix

$$\mathbf{Z}_i = \mathbf{I}_G \otimes \mathbf{z}_i.$$

- Therefore, there is no distinction between GMM 3SLS, GIV, and traditional 3SLS.
- We can, if we think there is system heteroskedasticity, use efficient GMM instead.

- As with estimating any system, there is a tradeoff between robustness and efficiency. Any identified equation can be estimated by 2SLS or single-equation efficient GMM. In general, a system procedure such as 3SLS or GMM on the system requires that all equations are correctly specified.
- What if we want to use a system procedure but some equations are unidentified? Can replace them with reduced forms.
- Easy algebraic equivalances for identified systems: 2SLS on each equation is the same as 3SLS if (i) Each equation is just identified or (ii) $\hat{\boldsymbol{\Sigma}}$ (the estimated variance-covariance matrix) used in 3SLS, is diagonal.

- More difficult to show : If only one equation in a system is overidentified, and the other equations are just identified, 2SLS on the overidentified equation is the same as the 3SLS estimates when 3SLS is applied to the system.
- In particular, since RFs are just identified, if only one equation is an overidentified structural equation and the remaining equations are RFs, 2SLS = 3SLS on the overidentified equation.

• Suppose for an identified system we put each equation into two groups, just identified and overidentified. For the overidentified set of equations, the 3SLS estimates based on the entire system are identical to the 3SLS estimates using only the overidentified subset. But 3SLS estimation on the whole system for the just identified set of equations is generally more efficient than just applying 3SLS on the just identified set (which is equivalent to 2SLS on each equation). See Schmidt (1976, Theorem 5.2.13)

• Equivalances do not hold – except when all equations are just identified – with efficient GMM. Accounting for system heteroskedasticity could be more efficient.

- An important point that is somewhat subtle is that if an equation in a system has only exogenous variables on the right hand side, but some exogenous variables are excluded, then it is those exclusion restrictions that deliver increased efficiency of 3SLS (say) relative to 2SLS. The latter uses unrestricted reduced forms.
- To evaluate specific policies, but not to get estimates of "deep" (structural) parameters, we can just directly estimate the RFs. Simulations of the effects of policy changes on equilibrium values use the reduced form, whether they are estimated directly or obtained from the structural parameter estimates: $\hat{\mathbf{\Pi}} = -\hat{\mathbf{\Delta}}\hat{\mathbf{\Gamma}}^{-1}$

EXAMPLE: Labor Supply/Wage Offer System

• Only consider women who worked positive hours during the year (because we only observe an hourly wage for working women).

• We will revisit this example later and discuss how to deal with the sample selection problem caused by dropping the $hours = 0$ observations.

```
. use mroz

. reg hours lwage educ nwifeinc age kidslt6 kidsge6

      Source |       SS       df       MS              Number of obs =     428
-------------+------------------------------           F(  6,   421) =    5.04
       Model |  17228385.3     6  2871397.55           Prob > F      =  0.0001
    Residual |   240082635   421  570267.541           R-squared     =  0.0670
-------------+------------------------------           Adj R-squared =  0.0537
       Total |   257311020   427   602601.92           Root MSE      =  755.16

------------------------------------------------------------------------------
       hours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lwage |  -17.40781   54.21544    -0.32   0.748    -123.9745    89.15887
        educ |  -14.44486   17.96793    -0.80   0.422    -49.76289    20.87317
    nwifeinc |  -4.245807   3.655815    -1.16   0.246    -11.43173    2.940117
         age |  -7.729976    5.52945    -1.40   0.163    -18.59874    3.138792
     kidslt6 |  -342.5048   100.0059    -3.42   0.001     -539.078   -145.9317
     kidsge6 |  -115.0205   30.82925    -3.73   0.000    -175.6189   -54.42208
       _cons |   2114.697   340.1307     6.22   0.000     1446.131    2783.263
------------------------------------------------------------------------------

. * OLS gives essentially zero slope for the labor supply function.
```

```
. * Use 2SLS, with instruments exper, expersq for lwage:

. ivreg hours educ nwifeinc age kidslt6 kidsge6 (lwage = exper expersq)

Instrumental variables (2SLS) regression

      Source |       SS       df       MS              Number of obs =     428
-------------+------------------------------           F(  6,   421) =    3.41
       Model |  -456272250     6  -76045375            Prob > F      =  0.0027
    Residual |   713583270   421  1694972.14           R-squared     =       .
-------------+------------------------------           Adj R-squared =       .
       Total |   257311020   427   602601.92           Root MSE      =  1301.9

------------------------------------------------------------------------------
       hours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lwage |   1544.819   480.7387     3.21   0.001     599.8713    2489.766
        educ |   -177.449    58.1426    -3.05   0.002    -291.7349   -63.16302
    nwifeinc |  -9.249121   6.481116    -1.43   0.154     -21.9885    3.490256
         age |  -10.78409   9.577347    -1.13   0.261    -29.60946    8.041289
     kidslt6 |  -210.8339    176.934    -1.19   0.234    -558.6179    136.9501
     kidsge6 |  -47.55708   56.91786    -0.84   0.404    -159.4357     64.3215
       _cons |   2432.198   594.1719     4.09   0.000     1264.285    3600.111
------------------------------------------------------------------------------
Instrumented:  lwage
Instruments:   educ nwifeinc age kidslt6 kidsge6 exper expersq
------------------------------------------------------------------------------

. * Now very strong upward slope. Do we believe the exclusion restrictions
. * -- excluding exper, expersq -- in the labor supply function?
```

```
. * The heteroskedasticity-robust standard error is substantially larger
. * on lwage, and so we should use robust inference:

. ivreg hours educ nwifeinc age kidslt6 kidsge6 (lwage = exper expersq), robust

Instrumental variables (2SLS) regression               Number of obs =     428
                                                       F(  6,   421) =    2.53
                                                       Prob > F      =  0.0205
                                                       R-squared     =       .
                                                       Root MSE      =  1301.9

------------------------------------------------------------------------------
             |               Robust
       hours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lwage |   1544.819    603.758     2.56   0.011     358.0628    2731.574
        educ |   -177.449   67.39857    -2.63   0.009    -309.9286   -44.96935
    nwifeinc |  -9.249121   5.274702    -1.75   0.080    -19.61715    1.118911
         age |  -10.78409   10.66514    -1.01   0.313    -31.74764    10.17947
     kidslt6 |  -210.8339      205.6    -1.03   0.306    -614.9643    193.2965
     kidsge6 |  -47.55708   56.94704    -0.84   0.404     -159.493    64.37887
       _cons |   2432.198   616.2835     3.95   0.000     1220.822    3643.574
------------------------------------------------------------------------------
Instrumented:  lwage
Instruments:   educ nwifeinc age kidslt6 kidsge6 exper expersq
------------------------------------------------------------------------------
```

```
. * Use 3SLS on system (but reg3 does not allow inference robust to
. * system heteroskedasticity):

. reg3 (hours lwage educ nwifeinc age kidslt6 kidsge6) (lwage hours
                 educ exper expersq )

Three-stage least-squares regression

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hours        |
       lwage |   1676.933    431.169     3.89   0.000     831.8577    2522.009
        educ |  -205.0267   51.84729    -3.95   0.000    -306.6455   -103.4078
    nwifeinc |   .3678943   3.451518     0.11   0.915    -6.396957    7.132745
         age |  -12.28121   8.261529    -1.49   0.137    -28.47351    3.911094
     kidslt6 |  -200.5673   134.2685    -1.49   0.135    -463.7287    62.59414
     kidsge6 |  -48.63986   35.95137    -1.35   0.176    -119.1032    21.82352
       _cons |   2504.799   535.8919     4.67   0.000      1454.47    3555.128
-------------+----------------------------------------------------------------
lwage        |
       hours |    .000201   .0002109     0.95   0.340    -.0002123    .0006143
        educ |   .1129699   .0151452     7.46   0.000     .0832858    .1426539
       exper |   .0208906   .0142782     1.46   0.143    -.0070942    .0488753
     expersq |  -.0002943   .0002614    -1.13   0.260    -.0008066     .000218
       _cons |  -.7051103   .3045904    -2.31   0.021    -1.302097   -.1081241
------------------------------------------------------------------------------
Endogenous variables:  hours lwage
Exogenous variables:   educ nwifeinc age kidslt6 kidsge6 exper expersq
------------------------------------------------------------------------------
```

```
. * In the system, hourly wage offer does not appear to depend on hours.
. * Key difference between 3SLS and 2SLS estimation of labor supply equation:
. * the former maintains the exclusion restrictions in the wage offer
. * equation. 2SLS uses an unrestricted reduced form for lwage.

. * What if we try to estimate labor demand instead of wage offer?

. ivreg hours educ exper expersq (lwage = nwifeinc age kidslt6 kidsge6)

Instrumental variables (2SLS) regression

      Source |       SS       df       MS              Number of obs =     428
-------------+------------------------------           F(  4,   423) =    6.55
       Model | -188692851     4 -47173212.7           Prob > F      =  0.0000
    Residual |  446003871   423  1054382.67           R-squared     =       .
-------------+------------------------------           Adj R-squared =       .
       Total |  257311020   427   602601.92           Root MSE      =  1026.8

------------------------------------------------------------------------------
       hours |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lwage |   1000.535   805.4179     1.24   0.215    -582.5845    2583.655
        educ |  -130.1076    89.2759    -1.46   0.146    -305.5873    45.37201
       exper |   13.88497   39.15248     0.35   0.723    -63.07268    90.84262
     expersq |  -.0257315   .8910653    -0.03   0.977    -1.777199    1.725736
       _cons |   1584.152   520.0551     3.05   0.002      561.938    2606.366
------------------------------------------------------------------------------
Instrumented:  lwage
Instruments:   educ exper expersq nwifeinc age kidslt6 kidsge6
------------------------------------------------------------------------------
```

```
. * The previous analysis shows that it matters for estimation whether we
. * specify a wage offer or labor demand function. The estimated labor
. * demand function gives nonsense. We can see the problem by looking at
. * the reduced form for lwage. The RF does not depend on the excluded
. * exogenous variables in the labor demand function:

. reg lwage educ exper expersq nwifeinc age kidslt6 kidsge6

      Source |       SS       df       MS              Number of obs =     428
-------------+------------------------------           F(  7,   420) =   11.78
       Model |  36.6476796     7  5.2353828            Prob > F      =  0.0000
    Residual |  186.679761   420  .444475622           R-squared     =  0.1641
-------------+------------------------------           Adj R-squared =  0.1502
       Total |  223.327441   427  .523015084           Root MSE      =  .66669

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0998844   .0150975     6.62   0.000     .0702084    .1295604
       exper |   .0407097   .0133723     3.04   0.002     .0144249    .0669946
     expersq |  -.0007473   .0004018    -1.86   0.064    -.0015371    .0000424
    nwifeinc |   .0056942   .0033195     1.72   0.087    -.0008307    .0122192
         age |  -.0035204   .0054145    -0.65   0.516    -.0141633    .0071225
     kidslt6 |  -.0558725   .0886034    -0.63   0.529    -.2300339    .1182889
     kidsge6 |  -.0176484    .027891    -0.63   0.527    -.0724718    .0371749
       _cons |  -.3579972   .3182963    -1.12   0.261    -.9836494    .2676551
------------------------------------------------------------------------------
```

```
. test nwifeinc age kidslt6 kidsge6

 ( 1)  nwifeinc = 0
 ( 2)  age = 0
 ( 3)  kidslt6 = 0
 ( 4)  kidsge6 = 0

       F(  4,   420) =    0.91
            Prob > F =    0.4555
```