SYSTEMS OF EQUATIONS: INSTRUMENTAL VARIABLES

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

1. Examples of Systems with IVs
2. The System IV Estimator
3. Generalized Method of Moments
4. The Generalized IV Estimator
5. Testing
6. More on Efficiency
7. Summary

1. EXAMPLES OF SYSTEMS WITH IVS

- As in the OLS and GLS case, carry along two examples: a system that looks like a SUR setup, but where some explanatory variables are endogenous in their own equation (at least), and the panel data case, where some explanatory variables are contemporaneously endogenous.
- The sampling environment is the same as before. We assume random sampling of units from a well-defined population.
- Generally, the analysis is similar to estimation by SOLS and GLS, but we must distinguish explanatory variables from instruments.

The SUR Case with Endogenous Explanatory Variables

- As before, write the population model as

$$\begin{aligned} y_1 &= \mathbf{x}_1\boldsymbol{\beta}_1 + u_1 \\ y_2 &= \mathbf{x}_2\boldsymbol{\beta}_2 + u_2 \\ &\vdots \\ y_G &= \mathbf{x}_G\boldsymbol{\beta}_G + u_G \end{aligned} \tag{1.1}$$

where y_g is a response variable, $g = 1,\ldots,G$. The explanatory variables, $\mathbf{x}_g$, can be different across equations.

- Now we want to allow $E(\mathbf{x}_g'u_g) \neq 0$ for at least some equations g.

• **EXAMPLE**: Individual Labor Supply. Consider a labor supply function and a wage offer (inverse labor demand) function:

$$h^s(w) = \gamma_1 w + \mathbf{z}_1\boldsymbol{\delta}_1 + u_1 \tag{1.2}$$

$$w^o(h) = \gamma_2 h + \mathbf{z}_2\boldsymbol{\delta}_2 + u_2 \tag{1.3}$$

Equation (1.2) is the labor supply function, which shows how much each unit in the population would work at any given wage, w. Once we hold fixed the observed characteristics $\mathbf{z}_1$ and unobserved characteristics u_1, we can trace out the (linear) supply curve as a function of w.

• The labor supply function stands on its own, representing the individual side of the market.

- If we could run the right experiment, we could use OLS to estimate the labor supply function: randomly assign wages, w_i, to individuals, and record the resulting labor supply, h_i.
- We bring in the wage offer function to recognize that, for retrospective data, we observe the pair (h_i, w_i), with w_i not being randomly assigned. What is a sensible assumption about how (h_i, w_i) are generated? A standard approach is to assume that we observe equalibrium hours and wages for each individual. That is, the data are generated as

$$h_i = \gamma_1 w_i + \mathbf{z}_{i1}\boldsymbol{\delta}_1 + u_{i1} \quad (1.4)$$

$$w_i = \gamma_2 h_i + \mathbf{z}_{i2}\boldsymbol{\delta}_2 + u_{i2} \quad (1.5)$$

- In the usual case, we assume that the elements of $\mathbf{z}_1$ and $\mathbf{z}_2$ are exogenous to both equations, but we will consider other cases.
- The labor supply-wage offer example is a simultaneous equations model (SEM). We will talk more about identification of SEMs later.

EXAMPLE: Effects of Head Start on Student Peformance. Suppose for a population of disadvantaged students we postulate the equation

$$score_i = \gamma_1 HeadStart_i + \mathbf{z}_{i1}\boldsymbol{\delta}_1 + u_{i1}, \tag{1.6}$$

where $HeadStart_i = 1$ means that the student participated in the Head Start program as a child.

- There is no "simultaneity" here because Head Start participation is determined in advance of the standardized test score. But there could be a self-selection problem: participation in Head Start may be related to unobserved factors – such as parental involvement – that affect test scores. In other words, $HeadStart_i$ might be correlated with u_{i1}.
- As we know from single equation analysis, we can write a linear projection for $HeadStart_i$ as

$$HeadStart_i = \mathbf{z}_i \boldsymbol{\delta}_2 + u_{i2}, \tag{1.7}$$

where $\mathbf{z}_i$ includes some other exogenous variables in addition to $\mathbf{z}_{i1}$.

- There is nothing “structural” about (1.7). In the terminology of Chapter 5, it is a “reduced form.” We could just estimate (1.6) by 2SLS. But for some purposes – mainly efficiency considerations – it is useful to consider (1.6) and (1.7) together as a system.

• The previous examples can be written as

$$y_{i1} = \mathbf{x}_{i1}\boldsymbol{\beta}_1 + u_{i1} \tag{1.8}$$

$$y_{i2} = \mathbf{x}_{i2}\boldsymbol{\beta}_2 + u_{i2} \tag{1.9}$$

where we now need IVs for one or both equations.

• In many cases, the same instruments can be used for each equation, but in other cases the instruments will differ across equations.

Panel Data Models with Endogenous Explanatory Variables

- Write a panel data model as

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + u_{it} \tag{1.10}$$

$$E(\mathbf{z}_{it}'u_{it}) = 0,\ t = 1,\ldots,T, \tag{1.11}$$

where $\mathbf{x}_{it}$ is $1 \times K$. In the general case, the dimension of the instruments $\mathbf{z}_{it}$ can change with t, so $1 \times L_t$. (We will see specific examples with unobserved effects models.)

- In many cases, elements of $\mathbf{z}_{it}$ might be correlated with the errors in other time periods, so care is used in choosing $\mathbf{z}_{it}$ for each t.)

EXAMPLE: Estimating a Passenger Demand Function. Suppose passenger demand for flights is given by

$$\begin{aligned}\log(passen_{it}) = {} & \theta_t + \beta_1 \log(fare_{it}) \\ & + \beta_2 \log(dist_i) + \beta_3[\log(dist_i)]^2 + u_{it}\end{aligned} \tag{1.13}$$

where $lfare_{it} = \log(fare_{it})$ is generally correlated with u_{it}. If fares are generally higher with a higher concentration ratio, use $concen_{it}$ as an instrument for $lfare_{it}$. Then

$$\mathbf{z}_{it} = (1, d2_t, \ldots, dT_t, concen_{it}, ldist_i, ldist_i^2) \tag{1.14}$$

which recognizes that a fully set of period dummies has been included.

- Suppose that $lfare_{it}$ is also correlated with past $concen_{it}$, and

$$E(u_{it}|concen_{it}, concen_{i,t-1}, \ldots, concen_{i1}, ldist_i) = 0. \tag{1.15}$$

Then we could take

$$\mathbf{z}_{it} = (1, d2_t, \ldots, dT_t, concen_{it}, concen_{i,t-1}, \ldots, concen_{i1}, dist_i, ldist_i^2), \tag{1.16}$$

which has dimension increasing in t. Have to think about how to use all of this information.

- As before, the SUR and panel data cases can both be written as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{u}_i \tag{1.17}$$

where $\mathbf{y}_i$ is $G \times 1$ (or $T \times 1$) and $\mathbf{X}_i$ is $G \times K$ or $T \times K$, and $\boldsymbol{\beta}$ is the $K \times 1$ vector of parameters to be estimated.

- How do we choose the matrix of instruments, $\mathbf{Z}_i$?

SUR: Assume that for each equation g, the moment conditions are

$$E(\mathbf{z}_{ig}' u_{ig}) = \mathbf{0}, \tag{1.18}$$

for a $1 \times L_g$ vector $\mathbf{z}_{ig}$, written for a random draw.

Then the $G \times L$ matrix of instruments is

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{z}_{i1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_{i2} & \mathbf{0} & \vdots \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{z}_{iG} \end{pmatrix}, \tag{1.19}$$

where $L = L_1 + L_2 + \ldots + L_G$.

- Of course, we get the system OLS setup when $\mathbf{z}_{ig} = \mathbf{x}_{ig}$ for *all* g.

Panel Data: The instrument matrix that gives us the moment conditions $E(\mathbf{z}_{it}'u_{it}) = \mathbf{0}$ looks just like (1.19) with the small notational change of replacing G with T. Then

$$\mathbf{Z}_i'\mathbf{u}_i = \begin{pmatrix} \mathbf{z}_{i1}' & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_{i2}' & \mathbf{0} & \vdots \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{z}_{iT}' \end{pmatrix} \begin{pmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iT} \end{pmatrix} = \begin{pmatrix} \mathbf{z}_{i1}'u_{i1} \\ \mathbf{z}_{i2}'u_{i2} \\ \vdots \\ \mathbf{z}_{iT}'u_{iT} \end{pmatrix}$$

and so $E(\mathbf{Z}_i'\mathbf{u}_i) = \mathbf{0}$ is the same as

$$E(\mathbf{z}_{it}'u_{it}) = \mathbf{0},\ t = 1, 2, \ldots, T.$$

• When the dimension of the instruments is the same for all t, $L_t = L$, another choice is possible:

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{z}_{i1} \\ \mathbf{z}_{i2} \\ \vdots \\ \mathbf{z}_{iT} \end{pmatrix}. \tag{1.20}$$

• The IV matrix in (1.20) uses only a linear combination of the original moment conditions, namely,

$$E(\mathbf{Z}_i'\mathbf{u}_i) = \sum_{t=1}^{T} E(\mathbf{z}_{it}'u_{it}) = \mathbf{0}.$$

• If used properly, the instrument matrix in (1.19) will give an asymptotically more efficient estimator than that in (1.20). (But using many extra moment conditions can cause the finite-sample properties to be worse.)

2. THE SYSTEM IV ESTIMATOR

- Write the system as in the system OLS/GLS case:

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

Assumption SIV.1 (Moment Conditions): For a $G \times L$ matrix $\mathbf{Z}_i$,

$$E(\mathbf{Z}_i'\mathbf{u}_i) = \mathbf{0}. \ \square \tag{2.1}$$

- The assumption is weaker, often in important ways, than the assumption that all elements of $\mathbf{Z}_i$ are uncorrelated with all elements in $\mathbf{u}_i$: $E(\mathbf{Z}_i \otimes \mathbf{u}_i) = \mathbf{0}$. We return to this distinction later.

Assumption SIV.2 (Rank Condition):

$$rank\ E(\mathbf{Z}_i'\mathbf{X}_i) = K. \quad \square \tag{2.2}$$

- Looks like the key rank condition for 2SLS except that $\mathbf{Z}_i$ and $\mathbf{X}_i$ are matrices.
- A necessary conditon for SIV.2 to hold is the order condition, $L \geq K$.
- Later it is useful to define $\mathbf{C} = E(\mathbf{Z}_i'\mathbf{X}_i)$.

- In the SUR case

$$E(\mathbf{Z}_i'\mathbf{X}_i) = \begin{pmatrix} E(\mathbf{z}_{i1}'\mathbf{x}_{i1}) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & E(\mathbf{z}_{i2}'\mathbf{x}_{i2}) & \mathbf{0} & \vdots \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & E(\mathbf{z}_{iG}'\mathbf{x}_{iG}) \end{pmatrix}$$

and so SIV.2 holds if and only if

$$rank\, E(\mathbf{z}_{ig}'\mathbf{x}_{ig}) = K_g,\ g = 1,\ldots,G. \tag{2.3}$$

This is the same as saying the rank condition holds for each equation g.

(The nonsingularity requirement on $E(\mathbf{z}_{ig}'\mathbf{z}_{ig})$ will come later.)

- In the panel data case, if

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{z}_{i1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_{i2} & \mathbf{0} & \vdots \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{z}_{iT} \end{pmatrix} \tag{2.4}$$

then

$$E(\mathbf{Z}_i'\mathbf{X}_i) = \begin{pmatrix} E(\mathbf{z}_{i1}'\mathbf{x}_{i1}) \\ E(\mathbf{z}_{i2}'\mathbf{x}_{i2}) \\ \vdots \\ E(\mathbf{z}_{iT}'\mathbf{x}_{iT}) \end{pmatrix}. \tag{2.5}$$

- If

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{z}_{i1} \\ \mathbf{z}_{i2} \\ \vdots \\ \mathbf{z}_{iT} \end{pmatrix} \tag{2.6}$$

then

$$E(\mathbf{Z}_i'\mathbf{X}_i) = \sum_{t=1}^{T} E(\mathbf{z}_{it}'\mathbf{x}_{it}). \tag{2.7}$$

• Estimation proceeds from the method of moments. The moment condition in Assumption SIV.1 can be written as

$$E[\mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})] = \mathbf{0} \tag{2.7}$$

or

$$E(\mathbf{Z}_i'\mathbf{X}_i)\boldsymbol{\beta} = E(\mathbf{Z}_i'\mathbf{y}_i). \tag{2.8}$$

• If the rank condition is violated (for example, if $L < K$), the system $E(\mathbf{Z}_i'\mathbf{X}_i)\mathbf{b} = E(\mathbf{Z}_i'\mathbf{y}_i)$ has more than one solution for $\mathbf{b}$, and $\boldsymbol{\beta}$ is not identified.

• Suppose $L = K$. Then

$$\boldsymbol{\beta} = [E(\mathbf{Z}_i'\mathbf{X}_i)]^{-1}E(\mathbf{Z}_i'\mathbf{y}_i) \tag{2.9}$$

• From here, there is only one thing to do: replace population averages with sample averages to get the **system instrumental variables (SIV)** estimator:

$$\hat{\boldsymbol{\beta}}_{SIV} = \left(N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{X}_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{y}_i\right). \tag{2.10}$$

Consistency is immediate by the usual WLLN argument.

- Alternative, we can write

$$\hat{\boldsymbol{\beta}}_{SIV} = \boldsymbol{\beta} + \left(N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{X}_i \right)^{-1} \left(N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{u}_i \right) \tag{2.11}$$

and use $N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{u}_i \xrightarrow{p} E(\mathbf{Z}_i' \mathbf{u}_i) = \mathbf{0}$.

- Equation (2.11) is more convenient for studying the asymptotic distribution.

- The SIV estimator for the SUR system where we have $L_g = K_g$ for all g is IV equation-by-equation. The algebra is straightforward.
- The SIV estimator for the panel data system with IVs stacked as in (2.6) with $L = K$ is a pooled IV estimator:

$$\hat{\boldsymbol{\beta}}_{PIV} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{z}_{it}'\mathbf{x}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{z}_{it}'y_{it}\right). \tag{2.12}$$

- Inference is covered generally in the next section.

3. GENERALIZED METHOD OF MOMENTS ESTIMATION

• We now turn to a general treatment with (potential) overidentification, that is, $L > K$.

3.1 A General Weighting Matrix

• Though we assume the population moment conditions $E(\mathbf{Z}_i'\mathbf{X}_i)\boldsymbol{\beta} = E(\mathbf{Z}_i'\mathbf{y}_i)$ uniquely determine $\boldsymbol{\beta}$, the sample analog,

$$\left(\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{X}_i\right)\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{y}_i\right), \tag{3.1}$$

generally has no solution when $L > K$ (L equations in K unknowns).

• We could choose $\hat{\boldsymbol{\beta}}$ to make the Euclidean length of the vector $\sum_{i=1}^{N} \mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}})$ as small as possible, that is, choose $\hat{\boldsymbol{\beta}}$ to solve

$$\min_{\mathbf{b}\in\mathbb{R}^K} \left[\sum_{i=1}^{N} \mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\mathbf{b}) \right]' \left[\sum_{i=1}^{N} \mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\mathbf{b}) \right]. \tag{3.2}$$

• This estimator is consistent and is sometimes used as an initial estimator, but it is essentially never efficient.

• Consider a general class of estimators that use a *weighted* Euclidean length.

- Let $\hat{\mathbf{W}}$ be an $L \times L$ symmetric, positive semi-definite matrix, which can be random (and usually depends on the same random sample of data). Consider now the problem

$$\min_{\mathbf{b} \in \mathbb{R}^K} \left[\sum_{i=1}^{N} \mathbf{Z}_i' (\mathbf{y}_i - \mathbf{X}_i \mathbf{b}) \right]' \hat{\mathbf{W}} \left[\sum_{i=1}^{N} \mathbf{Z}_i' (\mathbf{y}_i - \mathbf{X}_i \mathbf{b}) \right]. \tag{3.3}$$

- The solution to (3.3) is called a **generalized method of moments** (**GMM**) estimator.
- Can solve this problem using multivariable calculus.

• Let $\mathbf{Z}$ be the $NG \times L$ matrix of instruments stacked by observation, and similarly for $\mathbf{X}$ ($NG \times K$) and $\mathbf{Y}$ ($NG \times 1$). Can show that

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{Z}\hat{\mathbf{W}}\mathbf{Z}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Z}\hat{\mathbf{W}}\mathbf{Z}'\mathbf{Y}), \tag{3.4}$$

where

$$\mathbf{Z}'\mathbf{X} = \sum_{i=1}^{N} \mathbf{Z}_i'\mathbf{X}_i, \ \mathbf{Z}'\mathbf{Y} = \sum_{i=1}^{N} \mathbf{Z}_i'\mathbf{y}_i. \tag{3.5}$$

Assumption SIV.3 (Positive Definite Limit): For an $L \times L$ nonrandom positive definite matrix $\mathbf{W}$,

$$\hat{\mathbf{W}} \xrightarrow{p} \mathbf{W} \text{ as } N \to \infty. \quad \square \tag{3.6}$$

- Positive definiteness is stonger than needed. As will be clear, having $\mathbf{C}'\mathbf{W}\mathbf{C}$ full rank (nonsigular) is sufficient, where $\mathbf{C} = E(\mathbf{Z}_i'\mathbf{X}_i)$.
- Usually the law of large numbers, combined with consistency of a first-stage estimator, is used to establish SIV.3.

Theorem (Consistency): Under SIV.1 to SIV.3, $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$.

Proof: Write

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left[\left(N^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{Z}_i \right) \hat{\mathbf{W}} \left(N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{X}_i \right) \right]^{-1} \cdot \left(N^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{Z}_i \right) \hat{\mathbf{W}} \left(N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{u}_i \right) \tag{3.7}$$

$$\begin{aligned} \operatorname{plim}_{N\to\infty}(\hat{\boldsymbol{\beta}}) &= \boldsymbol{\beta} + (\mathbf{C}'\mathbf{W}\mathbf{C})^{-1}\mathbf{C}'\mathbf{W} \left(\operatorname{plim}_{N\to\infty} N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{u}_i \right) \\ &= \boldsymbol{\beta} + (\mathbf{C}'\mathbf{W}\mathbf{C})^{-1}\mathbf{C}'\mathbf{W} \cdot \mathbf{0} = \boldsymbol{\beta}. \end{aligned} \tag{3.8}$$

Theorem (Asymptotic Normality): Under SIV.1, SIV.2, and SIV.3, $\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically normal with mean zero and variance matrix

$$Avar\,\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = (\mathbf{C}'\mathbf{W}\mathbf{C})^{-1}\mathbf{C}'\mathbf{W}\boldsymbol{\Lambda}\mathbf{W}\mathbf{C}(\mathbf{C}'\mathbf{W}\mathbf{C})^{-1}. \tag{3.9}$$

where

$$\boldsymbol{\Lambda} = Var(\mathbf{Z}_i'\mathbf{u}_i) = E(\mathbf{Z}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i). \tag{3.10}$$

- Consistent estimation of $Avar\,\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ uses $\hat{\mathbf{C}} = N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{X}_i$, $\hat{\mathbf{W}}$ and a consistent estimator $\hat{\boldsymbol{\Lambda}}$ of $\boldsymbol{\Lambda}$ (more later).

3.2. System 2SLS

- The **System 2SLS** estimator uses weight matrix

$$\hat{\mathbf{W}} = \left(N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{Z}_i \right)^{-1}, \tag{3.11}$$

and the estimator can be written as

$$\hat{\boldsymbol{\beta}}_{S2SLS} = [\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y} \tag{3.12}$$

- Assumption SIV.3 is equivalent to

$$rank\, E(\mathbf{Z}_i'\mathbf{Z}_i) = L, \tag{3.13}$$

which extends the single equation assumption, S2SLS.2(a), in the obvious way.

• Inference with S2SLS is possible without further assumptions.

• In the SUR case, the S2SLS estimator is **2SLS equation-by-equation**.

• In the panel data case with $\mathbf{Z}_i = (\mathbf{z}_{i1}', \ldots, \mathbf{z}_{iT}')'$, can show that S2SLS is the **pooled 2SLS estimator**:

$$\mathbf{Z}'\mathbf{X} = \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{z}_{it}'\mathbf{x}_{it}, \ \mathbf{Z}'\mathbf{Z} = \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{z}_{it}'\mathbf{z}_{it}, \ \mathbf{Z}'\mathbf{Y} = \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{z}_{it}'y_{it}. \tag{3.14}$$

• When $\mathbf{Z}_i$ is block diagonal as in (2.4), the system 2SLS estimator is different. It can be obtained as follows. (1) For each t, estimate a reduced from from a cross section regression

$$\mathbf{x}_{it} \text{ on } \mathbf{z}_{it},\ i = 1, \ldots, N, \tag{3.15}$$

and obtain the fitted values,

$$\hat{\mathbf{x}}_{it} = \mathbf{z}_{it}\hat{\mathbf{\Pi}}_t, \tag{3.16}$$

where $\hat{\mathbf{\Pi}}_t$ is $L_t \times K$ in general. Note that $\hat{\mathbf{x}}_{it}$ is always $1 \times K$. (2) Estimate $\boldsymbol{\beta}$ via pooled IV in the equation

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + u_{it},\ t = 1,\ldots,T;\ i = 1,\ldots,N \tag{3.17}$$

using $\hat{\mathbf{x}}_{it}$ as instruments (not regressors!).

$$\hat{\boldsymbol{\beta}}_{S2SLS} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\mathbf{x}}_{it}'\mathbf{x}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\mathbf{x}}_{it}'y_{it}\right) \tag{3.18}$$

- Can use "cluster robust" inference in the pooled IV, ignoring estimation of the $\hat{\boldsymbol{\Pi}}_t$ in the first stage.
- The pooled 2SLS estimator estimates a common $\hat{\boldsymbol{\Pi}}$ by pooling the regression in (3.15) across t and i.
- There is a set of no serial correlation and homoskedasticity assumptions that justifies use of the usual statistics.

3.3. Optimal Weighting Matrix

• Given an infinite number of choices for $\mathbf{W}$, can we choose the best?

Yes. (And then it is obvious how to estimate the optimal $\mathbf{W}$.)

• Recall that $\boldsymbol{\beta}$ is defined by

$$E[\mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})] = E(\mathbf{Z}_i'\mathbf{u}_i) = \mathbf{0}. \tag{3.19}$$

• The optimal weighting matrix is the inverse of the variance matrix of $\mathbf{Z}_i'\mathbf{u}_i$, $\boldsymbol{\Lambda} = Var(\mathbf{Z}_i'\mathbf{u}_i)$.

Assumption SIV.4 (Optimal Weighting Matrix):

$$\mathbf{W} = \mathbf{\Lambda}^{-1}. \tag{3.20}$$

• With this choice of $\mathbf{W}$, the asymptotic variance collapses to

$$(\mathbf{C}'\mathbf{\Lambda}^{-1}\mathbf{C})^{-1}\mathbf{C}'\mathbf{\Lambda}^{-1}\mathbf{\Lambda}\mathbf{\Lambda}^{-\mathbf{1}}\mathbf{C}(\mathbf{C}'\mathbf{\Lambda}^{-1}\mathbf{C})^{-1} = (\mathbf{C}'\mathbf{\Lambda}^{-1}\mathbf{C})^{-1}. \tag{3.21}$$

• It can be shown that $(\mathbf{C}'\mathbf{\Lambda}^{-1}\mathbf{C})$ is the "smallest" possible by showing

$$(\mathbf{C}'\mathbf{\Lambda}^{-1}\mathbf{C}) - (\mathbf{C}'\mathbf{W}\mathbf{C})(\mathbf{C}'\mathbf{W}\mathbf{\Lambda}\mathbf{W}\mathbf{C})^{-1}(\mathbf{C}'\mathbf{W}\mathbf{C}) \tag{3.22}$$

is positive semi-definite for any $L \times L$ positive definite matrix $\mathbf{W}$.

- Let $\mathbf{D} = \mathbf{\Lambda}^{1/2}\mathbf{W}\mathbf{C}$ and show the difference can be expressed as

$$\mathbf{C}'\mathbf{\Lambda}^{-1/2}[\mathbf{I}_L - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}']\mathbf{\Lambda}^{-1/2}\mathbf{C} \tag{3.23}$$

and note that $\mathbf{I}_L - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$ is symmetric and idempotent (and hence positive semi-definite).

• To obtain an actual GMM estimator using an efficient weighting matrix, use a two-step procedure.

(1) Let $\check{\boldsymbol{\beta}}$ be an initial consistent estimator of $\boldsymbol{\beta}$, usually the system 2SLS estimator or the estimator using $\hat{\mathbf{W}} = \mathbf{I}_L$. Obtain the $G \times 1$ residual vectors, $\check{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i\check{\boldsymbol{\beta}}$, $i = 1, \ldots, N$, and compute

$$\hat{\boldsymbol{\Lambda}} = N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i'\check{\mathbf{u}}_i\check{\mathbf{u}}_i'\mathbf{Z}_i \xrightarrow{p} \boldsymbol{\Lambda}. \tag{3.24}$$

(2) Choose

$$\hat{\mathbf{W}} = \left(N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i'\check{\mathbf{u}}_i\check{\mathbf{u}}_i'\mathbf{Z}_i \right)^{-1}. \tag{3.25}$$

• We call such an estimator an **optimal GMM** estimator. It is sometimes called a **minimum chi-square** estimator (reason will be clear later).

• Important: the optimal weighting matrix provides an asymptotically efficient estimator in the class of estimators based on

$$E[\mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})] = \mathbf{0}. \tag{3.26}$$

In other words, the estimator is asymptotically efficient for the given set of moment conditions (instruments).

• It is possible we can find additional moment conditions that can enhance efficiency.

- When $L = K$ the weighting matrix is irrelevant. There is only one estimator consistent under (3.26), and that is the system IV estimator.
- If $\hat{\boldsymbol{\beta}}$ is now an optimal GMM estimator and $\hat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}$ are the optimal GMM residuals,

$$\widehat{Avar}(\hat{\boldsymbol{\beta}}) = N^{-1}\left[(\mathbf{X}'\mathbf{Z}/N)\left(N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\mathbf{Z}_i \right)^{-1} (\mathbf{Z}'\mathbf{X}/N) \right]^{-1}$$

$$= \left[(\mathbf{X}'\mathbf{Z})\left(\sum_{i=1}^{N}\mathbf{Z}_i'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\mathbf{Z}_i \right)^{-1} (\mathbf{Z}'\mathbf{X}) \right]^{-1}. \qquad (3.27)$$

- There is usually more than one way to estimate $\mathbf{\Lambda}$, and so the optimal GMM estimator is not unique. But all are $\sqrt{N}$-asymptotically equivalent.
- Recent research has focused on the small-sample properties of optimal GMM estimators. While replacing $\mathbf{\Lambda}$ with $\hat{\mathbf{\Lambda}}$ does not affect the $\sqrt{N}$-asymptotic distribution of $\hat{\boldsymbol{\beta}}$, it can have deleterious effects on the actual (finite sample) distribution. (More sophisticated asymptotic analysis picks this up.)

• It is using a first-stage estimate of $\boldsymbol{\beta}$ that causes finite-sample problems for two-step optimal GMM. System S2SLS does not have the same problems, but it is asymptotically inefficient.

• *Empirical likelihood* has been proposed as an alternative to GMM. See Imbens and Wooldridge (2007, NBER lecture notes).

3.4. The GMM Three Stage Least Squares Estimator

• Mainly for historical reasons, but also because it can have better small sample properties, we can consider a restricted version of the optimal weighting matrix.

Assumption SIV.5 (System Homoskedasticity): Let $\mathbf{\Omega} = E(\mathbf{u}_i\mathbf{u}_i')$. Then

$$E(\mathbf{Z}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i) = E(\mathbf{Z}_i'\mathbf{\Omega}\mathbf{Z}_i). \tag{3.28}$$

• SIV.5 means that all the squares and cross products u_{ig}^2, $u_{ig}u_{ih}$ are uncorrelated with the squares and cross products in $\mathbf{Z}_i \otimes \mathbf{Z}_i$.

- Can use the vec and Kronecker product operators to show this:

$$vec\, E(\mathbf{Z}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i) = E[(\mathbf{Z}_i \otimes \mathbf{Z}_i)'vec(\mathbf{u}_i\mathbf{u}_i')] = E[(\mathbf{Z}_i \otimes \mathbf{Z}_i)']E[vec(\mathbf{u}_i\mathbf{u}_i')]$$

or

$$E[(\mathbf{Z}_i \otimes \mathbf{Z}_i)'\mathbf{r}_i] = E[(\mathbf{Z}_i \otimes \mathbf{Z}_i)']E(\mathbf{r}_i)$$

where $\mathbf{r}_i = vec(\mathbf{u}_i\mathbf{u}_i')$ contains the u_{ig}^2, $u_{ig}u_{ih}$.

- In the SUR case, can show SIV.5 is the same as

$$E(u_{ig}^2 \mathbf{z}_{ig}' \mathbf{z}_{ig}) = E(u_{ig}^2) E(\mathbf{z}_{ig}' \mathbf{z}_{ig}) \tag{3.29}$$

$$E(u_{ig} u_{ih} \mathbf{z}_{ig}' \mathbf{z}_{ih}) = E(u_{ig} u_{ih}) E(\mathbf{z}_{ig}' \mathbf{z}_{ih}), \; g \neq h. \tag{3.30}$$

- By the usual iterated expectations argument, a sufficient condition is

$$E(\mathbf{u}_i \mathbf{u}_i' | \mathbf{Z}_i) = E(\mathbf{u}_i \mathbf{u}_i') \tag{3.31}$$

and sufficient for (3.31) (and Assumption SIV.1) are

$$E(\mathbf{u}_i | \mathbf{Z}_i) = \mathbf{0} \tag{3.32}$$

$$Var(\mathbf{u}_i | \mathbf{Z}_i) = Var(\mathbf{u}_i). \tag{3.33}$$

• Under SIV.5, we can estimate $\boldsymbol{\Lambda} = E(\mathbf{Z}_i'\boldsymbol{\Omega}\mathbf{Z}_i)$ differently. First,

$$\hat{\boldsymbol{\Omega}} = N^{-1}\sum_{i=1}^{N}\check{\mathbf{u}}_i\check{\mathbf{u}}_i' \tag{3.34}$$

and then

$$\hat{\boldsymbol{\Lambda}} = N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i'\hat{\boldsymbol{\Omega}}\mathbf{Z}_i = \mathbf{Z}'(\mathbf{I}_N \otimes \hat{\boldsymbol{\Omega}})\mathbf{Z}/N. \tag{3.35}$$

• Under SIV.1 to SIV.5, an optimal GMM estimator that uses the inverse of (3.35) as the weighting matrix can be written

$$\hat{\boldsymbol{\beta}}_{GMM3SLS} = \{\mathbf{X}'\mathbf{Z}[\mathbf{Z}'(\mathbf{I}_N \otimes \hat{\boldsymbol{\Omega}})\mathbf{Z}]^{-1}\mathbf{Z}'\mathbf{X}\}^{-1}\mathbf{X}'\mathbf{Z}[\mathbf{Z}'(\mathbf{I}_N \otimes \hat{\boldsymbol{\Omega}})\mathbf{Z}]^{-1}\mathbf{Z}'\mathbf{Y}. \tag{3.36}$$

• We call (3.36) the **GMM three stage least squares (3SLS) estimator**. (One does not need three steps to obtain this estimator; more later on the name.)

• For first-order asymptotics, there is no gain in using SIV.5. In other words, the general weighting matrix gives the same $\sqrt{N}$-asymptotic distribution as the GMM-3SLS weighting matrix. And, of course, the latter is inefficient in the presence of system heteroskedasticity.

• The argument for $\hat{\boldsymbol{\beta}}_{GMM3SLS}$ must come from finite-sample considerations or, as we see in the next section, from a different kind of efficiency argument.

4. THE GENERALIZED IV ESTIMATOR

4.1 Derivation of the GIV Estimator and Its Asymptotic Properties

- Rather than estimating $\boldsymbol{\beta}$ using the moment conditions $E[\mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})] = \mathbf{0}$, an alternative is to transform the moment conditions in a way analogous to generalized least squares. Let $\boldsymbol{\Omega} = E(\mathbf{u}_i\mathbf{u}_i')$ as before, and assume $\boldsymbol{\Omega}$ is positive definite. Consider the transformed equation

$$\boldsymbol{\Omega}^{-1/2}\mathbf{y}_i = \boldsymbol{\Omega}^{-1/2}\mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\Omega}^{-1/2}\mathbf{u}_i. \tag{4.1}$$

Apply the system 2SLS estimator to (4.1) with instruments $\boldsymbol{\Omega}^{-1/2}\mathbf{Z}_i$.

- The resulting estimator can be written as sums of matrices across i as

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{GIV} &= \left[\left(\sum_{i=1}^{N} \mathbf{X}_i' \boldsymbol{\Omega}^{-1} \mathbf{Z}_i \right) \left(\sum_{i=1}^{N} \mathbf{Z}_i' \boldsymbol{\Omega}^{-1} \mathbf{Z}_i \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{Z}_i' \boldsymbol{\Omega}^{-1} \mathbf{X}_i \right) \right]^{-1} \\
&\quad \cdot \left(\sum_{i=1}^{N} \mathbf{X}_i' \boldsymbol{\Omega}^{-1} \mathbf{Z}_i \right) \left(\sum_{i=1}^{N} \mathbf{Z}_i' \boldsymbol{\Omega}^{-1} \mathbf{Z}_i \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{Z}_i' \boldsymbol{\Omega}^{-1} \mathbf{y}_i \right) \qquad (4.2 \\
&= \boldsymbol{\beta} + \left[\left(\sum_{i=1}^{N} \mathbf{X}_i' \boldsymbol{\Omega}^{-1} \mathbf{Z}_i \right) \left(\sum_{i=1}^{N} \mathbf{Z}_i' \boldsymbol{\Omega}^{-1} \mathbf{Z}_i \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{Z}_i' \boldsymbol{\Omega}^{-1} \mathbf{X}_i \right) \right]^{-1} \\
&\quad \cdot \left(\sum_{i=1}^{N} \mathbf{X}_i' \boldsymbol{\Omega}^{-1} \mathbf{Z}_i \right) \left(\sum_{i=1}^{N} \mathbf{Z}_i' \boldsymbol{\Omega}^{-1} \mathbf{Z}_i \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{Z}_i' \boldsymbol{\Omega}^{-1} \mathbf{u}_i \right)
\end{aligned}
$$

Assumption GIV.1 (Exogeneity): $E(\mathbf{Z}_i \otimes \mathbf{u}_i) = \mathbf{0}$. $\square$

- GIV.1 implies

$$E(\mathbf{Z}_i' \boldsymbol{\Omega}^{-1} \mathbf{u}_i) = \mathbf{0} \tag{4.3}$$

- Assumption GIV.1 is identical to the consistency condition for GLS when $\mathbf{Z}_i = \mathbf{X}_i$.

Naturally, we also need a rank condition:

Assumption GIV.2 (Rank Condition): (a) *rank* $E(\mathbf{Z}_i'\mathbf{\Omega}^{-1}\mathbf{Z}_i) = L$; (b) *rank* $E(\mathbf{Z}_i'\mathbf{\Omega}^{-1}\mathbf{X}_i) = K$. $\square$

- When $G = 1$, Assumption GIV.2 reduces to Assumption 2SLS.2.

• Of course, to operationalize equation (4.2), we replace $\mathbf{\Omega}$ with an estimator $\hat{\mathbf{\Omega}}$, probably based on the system 2SLS residuals – just as with the GMM-3SLS estimator.

• When $\mathbf{\Omega}$ is replaced with a consistent estimator, $\hat{\mathbf{\Omega}}$, we obtain the **generalized instrumental variables (GIV) estimator**.

• Implicit in GIV estimation is the first-stage regression $\hat{\boldsymbol{\Omega}}^{-1/2}\mathbf{X}_i$ on $\hat{\boldsymbol{\Omega}}^{-1/2}\mathbf{Z}_i$ which yields fitted values $\hat{\boldsymbol{\Omega}}^{-1/2}\mathbf{Z}_i\hat{\boldsymbol{\Pi}}^*$, where, in terms of full data matrices, $\hat{\boldsymbol{\Pi}}^* = [\mathbf{Z}'(\mathbf{I}_N \otimes \hat{\boldsymbol{\Omega}}^{-1})\mathbf{Z}]^{-1}[\mathbf{Z}'(\mathbf{I}_N \otimes \hat{\boldsymbol{\Omega}}^{-1})\mathbf{X}]$.

• In comparing GIV to other estimators, it can be useful to think of GIV as system IV estmation of $\hat{\boldsymbol{\Omega}}^{-1/2}\mathbf{y}_i = \hat{\boldsymbol{\Omega}}^{-1/2}\mathbf{X}_i\boldsymbol{\beta} + \hat{\boldsymbol{\Omega}}^{-1/2}\mathbf{u}_i$ using IVs $\hat{\boldsymbol{\Omega}}^{-1/2}\mathbf{Z}_i\hat{\boldsymbol{\Pi}}^*$.

• As with SGLS.1, that is, $E(\mathbf{X}_i \otimes \mathbf{u}_i) = \mathbf{0}$, GIV.1 can impose unintended restrictions on the relationships between instruments and errors across equations or time or both. GMM estimators based on $E(\mathbf{Z}_i'\mathbf{u}_i) = \mathbf{0}$ are more robust in terms of consistency.

• Fully robust inference is possible, but a version of the system homoskedasticity assumption simplifies the variance matrix:

Assumption GIV.3 (System Homoskedasticity):

$E(\mathbf{Z}_i'\mathbf{\Omega}^{-1}\mathbf{u}_i\mathbf{u}_i'\mathbf{\Omega}^{-1}\mathbf{Z}_i) = E(\mathbf{Z}_i'\mathbf{\Omega}^{-1}\mathbf{Z}_i)$. $\square$

• This condition can be shown to hold if every square and cross product of $\mathbf{u}_i$, that is, u_{ig}^2 and $u_{ig}u_{ih}$, is uncorrelated with all elements of $\mathbf{Z}_i \otimes \mathbf{Z}_i$ (which consists of levels, squares, and cross products of all IVs for all equations.

- $\mathbf{\Omega}^{-1}\mathbf{u}_i$ is a linear combinationof the original errors, so $\mathbf{\Omega}^{-1}\mathbf{u}_i$ is also uncorrelated with all elements of $\mathbf{Z}_i \otimes \mathbf{Z}_i$. Then

$$\begin{aligned} vec\ E(\mathbf{Z}_i'\mathbf{\Omega}^{-1}\mathbf{u}_i\mathbf{u}_i'\mathbf{\Omega}^{-1}\mathbf{Z}_i) &= E[(\mathbf{Z}_i \otimes \mathbf{Z}_i)' vec(\mathbf{\Omega}^{-1}\mathbf{u}_i\mathbf{u}_i'\mathbf{\Omega}^{-1})] \\ &= E[(\mathbf{Z}_i \otimes \mathbf{Z}_i)']E[vec(\mathbf{\Omega}^{-1}\mathbf{u}_i\mathbf{u}_i'\mathbf{\Omega}^{-1})] \\ &= E[(\mathbf{Z}_i \otimes \mathbf{Z}_i)']vec[E(\mathbf{\Omega}^{-1}\mathbf{u}_i\mathbf{u}_i'\mathbf{\Omega}^{-1})] \\ &= E[(\mathbf{Z}_i \otimes \mathbf{Z}_i)']vec[\mathbf{\Omega}^{-1}E(\mathbf{u}_i\mathbf{u}_i')\mathbf{\Omega}^{-1}] \\ &= E[(\mathbf{Z}_i \otimes \mathbf{Z}_i)']vec(\mathbf{\Omega}^{-1}) = vec[E(\mathbf{Z}_i'\mathbf{\Omega}^{-1}\mathbf{Z}_i)] \end{aligned}$$

- A sufficient conditon for GIV.3 is $E(\mathbf{u}_i\mathbf{u}_i'|\mathbf{Z}_i) = E(\mathbf{u}_i\mathbf{u}_i')$.

• Under Assumptions GIV.1, GIV.2, and GIV.3,

$$Avar[\sqrt{N}(\hat{\boldsymbol{\beta}}_{GIV} - \boldsymbol{\beta})] = \{E(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{Z}_i)[E(\mathbf{Z}_i'\boldsymbol{\Omega}^{-1}\mathbf{Z}_i)]^{-1}E(\mathbf{Z}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i)]\}^{-1}, \quad (4.4)$$

and this matrix is easily estimated by the usual process of replacing expectations with sample averages and $\boldsymbol{\Omega}$ with $\hat{\boldsymbol{\Omega}}$.

- As with FGLS, improperly imposing restrictions in estimating $\mathbf{\Omega}$ does not lead to inconsistency in the GIV estimator. More precisely, suppose $\hat{\mathbf{\Lambda}}$ is obtained by restricting the variances and covariances in some way, and $\hat{\mathbf{\Lambda}} \overset{p}{\to} \mathbf{\Lambda} \neq \mathbf{\Omega}$. Then, under GIV.1, $E(\mathbf{Z}_i'\mathbf{\Lambda}^{-1}\mathbf{u}_i) = \mathbf{0}$, and this is the moment condition that implies consistency of the GIV estimator that uses $\hat{\mathbf{\Lambda}}$ as the variance matrix estimator.
- Of course, we will not have the simple expression for the asymptotic variance in equation (4.4), but we can easily compute a fully robust variance matrix estimator.
- We will use this for unobserved effects panel data models.

4.2 Comparison of GMM, GIV, and the Traditional 3SLS Estimator

• The **traditional 3SLS estimator** is typically motivated as follows. The first-stage uses untransformed $\mathbf{X}_i$ and $\mathbf{Z}_i$, giving fitted values $\hat{\mathbf{X}}_i = \mathbf{Z}_i\hat{\mathbf{\Pi}}$, where $\hat{\mathbf{\Pi}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$ is the matrix of first-stage regression coefficients. Then, $\mathbf{\Omega}^{-1/2}\mathbf{y}_i = \mathbf{\Omega}^{-1/2}\mathbf{X}_i\boldsymbol{\beta} + \mathbf{\Omega}^{-1/2}\mathbf{u}_i$ is estimated by system IV, but with instruments $\mathbf{\Omega}^{-1/2}\hat{\mathbf{X}}_i$. When we replace $\mathbf{\Omega}$ with its estimate, we arrive at the estimator

$$\hat{\boldsymbol{\beta}}_{T3SLS} = \left(\sum_{i=1}^{N}\hat{\mathbf{X}}_i'\hat{\mathbf{\Omega}}^{-1}\mathbf{X}_i\right)^{-1}\left(\sum_{i=1}^{N}\hat{\mathbf{X}}_i'\hat{\mathbf{\Omega}}^{-1}\mathbf{y}_i\right). \tag{4.5}$$

• Substituting $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{u}_i$ and rearranging shows that the orthogonality condition needed for consistency is

$$E[(\mathbf{Z}_i\boldsymbol{\Pi})'\boldsymbol{\Omega}^{-1}\mathbf{u}_i] = \boldsymbol{\Pi}'E(\mathbf{Z}_i'\boldsymbol{\Omega}^{-1}\mathbf{u}_i] = \mathbf{0} \tag{4.6}$$

where $\boldsymbol{\Pi} = plim(\hat{\boldsymbol{\Pi}})$. It is easily seen that (4.6) is implied by Assumption GIV.1 [although not by condition (4.3)]. Like the GIV estimator, the consistency of the traditional 3SLS estimator does not follow from $E(\mathbf{Z}_i'\mathbf{u}_i) = \mathbf{0}$.

- We have three different estimators of systems of equations based on first estimating $\mathbf{\Omega} = E(\mathbf{u}_i \mathbf{u}_i')$. Question: Why have seemingly different estimators all been given the label “three stage least squares”? Answer: In the setting that the traditional 3SLS estimator was proposed – the SUR system with the *same* instruments used in every equation – all estimates are identical.
- In fact, the equivalence of all estimates holds if we just impose the common instrument assumption in the general system $\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{u}_i$. Let $\mathbf{w}_i$ denote a vector assumed to be exogenous in every equation in the sense that $E(\mathbf{w}_i' u_{ig}) = \mathbf{0}$ for $g = 1, \ldots, G$ and choose

$$\mathbf{Z}_i = \mathbf{I}_G \otimes \mathbf{w}_i. \tag{4.7}$$

- Im, Ahn, Schmidt, and Wooldridge (1999) show that the GMM 3SLS estimator and the GIV estimator (using the same $\hat{\mathbf{\Omega}}$) are identical.
- The GIV estimator and the traditional 3SLS estimator are also identical: The first stage regressions involve the same set of explanatory variables, $\mathbf{w}_i$, in each equation, and so it does not matter whether the matrix of first stage regression coefficients is obtained via FGLS (used by GIV) or system OLS (used by T3SLS).

- In many modern applications of system IV methods, to both simultaneous equations and panel data, instruments that are exogenous in one equation are not exogenous in all other equations. In such cases it is important to use the GMM 3SLS estimator once $\mathbf{Z}_i$ has been properly chosen.
- Warning: Packages such as Stata, which have built-in 3SLS commands, maintain that a variable assumed to be exogenous in one equation is exogenous in every equation. Plus, inference cannot be made robust to system heteroskedasticity.

- Of course, the minimum chi square estimator that does not impose SIV.5 is always available, too, and often desirable. The GIV estimator and the traditional 3SLS estimator generally induce correlation between the transformed instruments and the structural errors.
- Most applications of method of moments tend to focus on GMM methods based on the original orthogonality conditions. Nevertheless, for unobserved effects models we will see that the GIV approach can provide insights into the workings of certain panel data estimators.

5. TESTING

- Consider linear hypotheses of the form

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}, \tag{5.1}$$

where $\mathbf{R}$ is $Q \times K$, $\mathbf{r}$ is $Q \times 1$, $Q \leq K$. We can always use the Wald statistic,

$$\mathcal{W} = (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})'(\mathbf{R}\hat{\mathbf{V}}\mathbf{R}')^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' \overset{a}{\sim} \chi^2_Q \tag{5.2}$$

where $\hat{\mathbf{V}}$ is an appropriate estimate of $\widehat{Avar}(\hat{\boldsymbol{\beta}})$.

- If we are using a minimum chi-square estimator, we can use a test based on the change in objective function when the restrictions are imposed.

• Let $\tilde{\boldsymbol{\beta}}$ be the restricted estimator using the same weighting matrix, $\hat{\mathbf{W}}$.

Then, under H_0, it can be shown that

$$N^{-1}\left[\left(\sum_{i=1}^{N}\mathbf{Z}_i'\tilde{\mathbf{u}}_i\right)'\hat{\mathbf{W}}\left(\sum_{i=1}^{N}\mathbf{Z}_i'\tilde{\mathbf{u}}_i\right)-\left(\sum_{i=1}^{N}\mathbf{Z}_i'\hat{\mathbf{u}}_i\right)'\hat{\mathbf{W}}\left(\sum_{i=1}^{N}\mathbf{Z}_i'\hat{\mathbf{u}}_i\right)\right] \quad (5.3)$$

$$\stackrel{d}{\to} \chi^2_Q$$

- For GMM 3SLS, the statistic is

$$\left(\sum_{i=1}^{N}\mathbf{Z}_i'\tilde{\mathbf{u}}_i\right)'\left(\sum_{i=1}^{N}\mathbf{Z}_i'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Z}_i\right)\left(\sum_{i=1}^{N}\mathbf{Z}_i'\tilde{\mathbf{u}}_i\right) - \left(\sum_{i=1}^{N}\mathbf{Z}_i'\hat{\mathbf{u}}_i\right)'\left(\sum_{i=1}^{N}\mathbf{Z}_i'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{Z}_i\right)\left(\sum_{i=1}^{N}\mathbf{Z}_i'\hat{\mathbf{u}}_i\right). \tag{5.4}$$

If system homoskedasticity fails, (5.4) does not have a limiting chi-square distribution.

• A test of the overidentifying restrictions, if $L > K$, is also based on the value of the objective function. Under H_0: $E(\mathbf{Z}_i'\mathbf{u}_i) = \mathbf{0}$,

$$\left(N^{-1/2}\sum_{i=1}^{N}\mathbf{Z}_i'\hat{\mathbf{u}}_i\right)'\hat{\mathbf{W}}\left(N^{-1/2}\sum_{i=1}^{N}\mathbf{Z}_i'\hat{\mathbf{u}}_i\right) \xrightarrow{d} \chi^2_{L-K} \tag{5.5}$$

where $\hat{\mathbf{W}}$ is an optimal weighting matrix.

6. MORE ON EFFICIENCY

6.1. Adding Instruments to Enhance Efficiency

- We first show that adding more instruments – that are exogenous, of course – can never hurt asymptotic efficiency provided an optimal weighting matrix is used.

- Let $\mathbf{Z}_{i1}$ be a $G \times L_1$, $L_1 \geq K$, submatrix of $\mathbf{Z}_i$, which is $G \times L$. It can be shown that

$$(\mathbf{C}_1'\mathbf{\Lambda}_1^{-1}\mathbf{C}_1)^{-1} - (\mathbf{C}'\mathbf{\Lambda}^{-1}\mathbf{C})^{-1} \tag{6.1}$$

is p.s.d., where $\mathbf{C}_1 = E(\mathbf{Z}_{i1}'\mathbf{X}_i)$, $\mathbf{\Lambda}_1 = E(\mathbf{Z}_{i1}'\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_{i1})$ and $\mathbf{C}$ and $\mathbf{\Lambda}$ are defined for the entire matrix of instruments, $\mathbf{Z}_i$.

- In other words, we can never do worse – using first-order asymptotics – by adding more IVs and using optimal GMM. (Caution: Recent research has shown that adding lots of poor IVs can seriously deteriorate the asymptotic approximations of GMM estimators.)

• Are there cases when asymptotic efficiency does not improve by adding more moment conditions? Generally, difficult to characterize. When $G = 1$ and the system homoskedasticity assumption $E(u_i^2\mathbf{z}_i'\mathbf{z}_i) = E(u_i^2)E(\mathbf{z}_i'\mathbf{z}_i)$ holds, the condition for no gain reduces to

$$E[(\mathbf{z}_{i2} - \mathbf{z}_{i1}\mathbf{D}_1)'\mathbf{x}_i] = \mathbf{0} \tag{6.2}$$

where

$$\mathbf{D}_1 = [E(\mathbf{z}_{i1}'\mathbf{z}_{i1})]^{-1}E(\mathbf{z}_{i1}'\mathbf{z}_{i2}) \quad (L_1 \times L_2) \tag{6.3}$$

- For a single endogenous explanatory variable x_{iK}, the condition is

$$L(x_{iK}|\mathbf{z}_{i1}, \mathbf{z}_{i2}) = L(x_{iK}|\mathbf{z}_{i1}), \tag{6.4}$$

that is, zero coefficients on $\mathbf{z}_{i2}$ in the reduced form for x_{iK}:

$$x_{iK} = \mathbf{z}_{i1}\boldsymbol{\pi}_1 + \mathbf{z}_{i2}\boldsymbol{\pi}_2 + r_{iK}, \ \boldsymbol{\pi}_2 = \mathbf{0}. \tag{6.5}$$

- This means that, given $\mathbf{z}_{i1}$, $\mathbf{z}_{i2}$ does not helpt to predict x_{iK} (in a linear sense).

• In the $G = 1$ case, if 2SLS.3 does not hold, so there is heteroskedasticity, 2SLS is generally inefficient if $L > K$. Should use GMM with a weighting matrix that allows $E(u_i^2 \mathbf{z}_i' \mathbf{z}_i)$ to be unrestricted.

• Generally, the optimal GMM estimator is more efficient than 2SLS. The optimal weighting matrix has the form

$$\left(N^{-1} \sum_{i=1}^{N} \check{u}_i^2 \mathbf{z}_i' \mathbf{z}_i \right)^{-1}, \tag{6.6}$$

where $\check{u}_i$ are the 2SLS residuals.

• If heteroskedasticity is present, we can keep adding instruments that would otherwise be redundant in order to improve efficiency.

• This finding for GMM has surprising efficiency implications for OLS. Suppose

$$y = \mathbf{x}\boldsymbol{\beta} + u,\ E(u|\mathbf{x}) = 0. \tag{6.7}$$

Under the zero conditional mean assumption, we can choose as instruments

$$\mathbf{z} = [\mathbf{x}, \mathbf{h}(\mathbf{x})] \tag{6.8}$$

for any (row) vector of functions $\mathbf{h}(\mathbf{x})$.

• Of course, with this choice of $\mathbf{z}$,

$$L(\mathbf{x}|\mathbf{z}) = L[\mathbf{x}|\mathbf{x}, \mathbf{h}(\mathbf{x})] = L(\mathbf{x}|\mathbf{x}) = \mathbf{x}, \tag{6.9}$$

and so, if

$$E(u^2|\mathbf{x}) = Var(u|\mathbf{x}) = \sigma^2, \tag{6.10}$$

we cannot improve efficiency over OLS by adding nonlinear functions of $\mathbf{x}$ to the instrument list.

- But if heteroskedasticity is present, the GMM estimator that uses the extended instrument list and a heteroskedasticity-robust weighting matrix is generally more efficient, asymptotically, than OLS!
- This result is due to Cragg (1983).
- In practice, the Cragg result is hardly used. For one, how should we choose $\mathbf{h}(\mathbf{x})$? And we might just be interested in the linear projection

6.2. Finding the Optimal Instruments

- So far, we know (i) how to obtain an optimal weighting matrix given $\mathbf{Z}_i$; (ii) adding valid instruments can improve efficiency, provided the optimal weighting matrix is used; (iii) in some cases, adding more instruments does not help.
- Now: Is there a way to choose a small number of instruments in an optimal way?

• The easiest case to describe is when we have a common set of exogenous variables for each equation, call these $\mathbf{w}_i$. Then

$$E(u_{ig}|\mathbf{w}_i) = 0, g = 1, \ldots, G. \tag{6.11}$$

or $E(\mathbf{u}_i|\mathbf{w}_i) = \mathbf{0}$.

• Under (6.11), any function of $\mathbf{w}_i$ is uncorrelated with u_{ig} for all g. Therefore, we can define a matrix of instruments $\mathbf{Z}_i$ and $G \times L$ matrix, $L \geq K$, that has any function of $\mathbf{w}_i$ as elements. But where would we stop? L could be very large (leading to poor small-sample properties).

• Question: Is there a way to reduce the functions of $\mathbf{w}_i$ we should consider? Yes.

• It turns out the optimal instruments, call them $\mathbf{Z}_i^*$, are

$$\mathbf{Z}_i^* = [Var(\mathbf{u}_i|\mathbf{w}_i)]^{-1}E(\mathbf{X}_i|\mathbf{w}_i), \tag{6.12}$$

which is a function of $\mathbf{w}_i$ only. Note that $\mathbf{Z}_i^*$ is $G \times K$, just like $\mathbf{X}_i$.

• If $\mathbf{Z}_i^*$ were available, we would have no reason to try other functions of $\mathbf{w}_i$ as instruments: once we have $\mathbf{Z}_i^*$, all other functions of $\mathbf{w}_i$ are redundant.

• The optimal instruments depend on a matrix of conditional means and a conditional variance-covariance matrix, and is generally unknown. Some have suggested "nonparametric" estimation of them.

• Suppose that $E(\mathbf{X}_i|\mathbf{w}_i) = (\mathbf{I}_G \otimes \mathbf{w}_i)\mathbf{\Pi} \equiv \mathbf{Z}_i\mathbf{\Pi}$ and $E(\mathbf{u}_i\mathbf{u}_i'|\mathbf{w}_i) = E(\mathbf{u}_i\mathbf{u}_i') = \mathbf{\Omega}$. Then the optimal instruments are simply $\mathbf{Z}_i^* = \mathbf{\Omega}^{-1}(\mathbf{Z}_i\mathbf{\Pi})$.

• The choice $\mathbf{Z}_i^* = \mathbf{\Omega}^{-1}(\mathbf{Z}_i\mathbf{\Pi})$ leads directly to the traditional 3SLS estimator when $\mathbf{\Omega}$ and $\mathbf{\Pi}$ are replaced by estimates. Because the same IVs show up in each equation, this estimator is the same as GIV and GMM-3SLS.

- The proof of efficiency for the general case is not difficult. Let $\mathbf{X}_i^* \equiv E(\mathbf{X}_i|\mathbf{w}_i)$ and $\mathbf{\Omega}_i \equiv E(\mathbf{u}_i\mathbf{u}_i'|\mathbf{w}_i)$, where it is important to remember that these are both functions of $\mathbf{w}_i$. Then $\mathbf{Z}_i^* = \mathbf{\Omega}_i^{-1}\mathbf{X}_i^*$. Let $\boldsymbol{\beta}^*$ be the (generally infeasible) SIV estimator using instruments $\mathbf{Z}_i^*$ Then

$$\mathrm{V}_1 \equiv Avar[\sqrt{N}(\boldsymbol{\beta}^* - \boldsymbol{\beta}) = [E(\mathbf{Z}_i^{*\prime}\mathbf{X}_i)]^{-1}[E(\mathbf{Z}_i^{*\prime}\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i^*)][E(\mathbf{X}_i'\mathbf{Z}_i^*)]^{-1}. \quad (6.13)$$

Now, by iterated expectations and using the fact that $\mathbf{Z}_i^*$ is a function of $\mathbf{w}_i$, $E(\mathbf{Z}_i^{*\prime}\mathbf{X}_i) = E[E(\mathbf{Z}_i^{*\prime}\mathbf{X}_i|\mathbf{w}_i)]$
$= E[\mathbf{Z}_i^{*\prime}E(\mathbf{X}_i|\mathbf{w}_i)] \;= E(\mathbf{Z}_i^{*\prime}\mathbf{X}_i^*) = E(\mathbf{X}_i^{*\prime}\mathbf{\Omega}_i^{-1}\mathbf{X}_i^*)$. Similarly,
$E(\mathbf{Z}_i^{*\prime}\mathbf{u}_i\mathbf{u}_i^{\prime}\mathbf{Z}_i^*) = E[E(\mathbf{Z}_i^{*\prime}\mathbf{u}_i\mathbf{u}_i^{\prime}\mathbf{Z}_i^*|\mathbf{w}_i)]$
$= E[\mathbf{Z}_i^{*\prime}E(\mathbf{u}_i\mathbf{u}_i^{\prime}|\mathbf{w}_i)\mathbf{Z}_i^*] = E(\mathbf{Z}_i^{*\prime}\mathbf{\Omega}_i\mathbf{Z}_i^*) = E(\mathbf{X}_i^{*\prime}\mathbf{\Omega}_i^{-1}\mathbf{X}_i^*)$. Therefore, we have shown that $\mathbf{V}_1$ simplifies to $[E(\mathbf{X}_i^{*\prime}\mathbf{\Omega}_i^{-1}\mathbf{X}_i^*)]^{-1}$.

• Next, let $\mathbf{Z}_i$ be any other $G \times K$ function of $\mathbf{w}_i$ and let $\hat{\boldsymbol{\beta}}$ be the SIV estimator that uses these instruments. Then

$$\begin{aligned}\mathbf{V}_2 \equiv Avar[\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= [E(\mathbf{Z}_i'\mathbf{X}_i)]^{-1}[E(\mathbf{Z}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i)][E(\mathbf{X}_i'\mathbf{Z}_i)]^{-1} \\ &= [E(\mathbf{Z}_i'\mathbf{X}_i^*)]^{-1}[E(\mathbf{Z}_i'\boldsymbol{\Omega}_i\mathbf{Z}_i)][E(\mathbf{X}_i^{*\prime}\mathbf{Z}_i)]^{-1} \qquad (6.14)\end{aligned}$$

where we again use iterated expectations and the fact that $\mathbf{Z}_i$ is a function of $\mathbf{w}_i$.

To show $\mathbf{V}_2 - \mathbf{V}_1$ is positive semi-definite, we assume that $\mathbf{V}_2$ is nonsingular and show that $\mathbf{V}_1^{-1} - \mathbf{V}_2^{-1}$ is p.s.d. But

$$\mathbf{V}_1^{-1} - \mathbf{V}_2^{-1} = E(\mathbf{X}_i^{*\prime}\mathbf{\Omega}_i^{-1}\mathbf{X}_i^{*}) - E(\mathbf{X}_i^{*\prime}\mathbf{Z}_i)[E(\mathbf{Z}_i^{\prime}\mathbf{\Omega}_i\mathbf{Z}_i)]^{-1}\mathbf{E}(\mathbf{Z}_i^{\prime}\mathbf{X}_i^{*}) \tag{6.15}$$

which we can show is equal to $E(\mathbf{R}_i^{\prime}\mathbf{R}_i)$, where $\mathbf{R}_i$ is the matrix of population residuals from the regression $\mathbf{\Omega}_i^{-1/2}\mathbf{X}_i^{*}$ on $\mathbf{\Omega}_i^{1/2}\mathbf{Z}_i$. Therefore, $\mathbf{V}_1^{-1} - \mathbf{V}_2^{-1}$ is p.s.d. and so is $\mathbf{V}_2 - \mathbf{V}_1$.

• **EXAMPLE**: Consider the linear model $y = \mathbf{x}\boldsymbol{\beta} + u$ under $E(u|\mathbf{x}) = 0$. The optimal instruments are

$$\mathbf{z}^* = \sigma^{-2}(\mathbf{x})\mathbf{x}, \tag{6.16}$$

where $\sigma^2(\mathbf{x}) = Var(u|\mathbf{x}) = Var(y|\mathbf{x})$. The optimal IV estimator is the weighted least squares estimator,

$$\boldsymbol{\beta}^* = \left(\sum_{i=1}^{N} \sigma^{-2}(\mathbf{x}_i)\mathbf{x}_i'\mathbf{x}_i\right)^{-1}\left(\sum_{i=1}^{N} \sigma^{-2}(\mathbf{x}_i)\mathbf{x}_i'y_i\right). \tag{6.17}$$

• Of course, we would need to estimate $Var(y|\mathbf{x})$ to operationalize this estimator.

7. SUMMARY

- Generally, if we start with moment conditions of the form $E(\mathbf{Z}_i'\mathbf{u}_i) = \mathbf{0}$, GMM estimation based on this set of moment conditions will be more robust than estimators based on a transformed set of moment conditions, such as GIV. If we decide to use GMM, we can use the unrestricted weighting matrix.
- Under Assumption SIV.5, which is a system homoskedasticity assumption, the GMM 3SLS estimator is an asymptotically efficient GMM estimator.
- When the same instruments can be used in every equation, GMM 3SLS, GIV, and traditional 3SLS are identical.

- When GMM and GIV are both consistent but are not $\sqrt{N}$ –asymptotically equivalent, they cannot generally be ranked in terms of asymptotic efficiency.
- One can never do worse by adding instruments and using the efficient weighting matrix in GMM. This has implications for panel data applications. For example, if one has the option of choosing the instruments as a block diagonal matrix or as a stacked matrix, it is better in large samples to use the block diagonal form.

• What about System 2SLS and GMM 3SLS? Under system homoskedasticity, 3SLS is generally more efficient. But there are situations where they coincide. (1) If the system is just identified, that is, $L = K$, all estimators reduce to the SIV estimator. In the case of the SUR system, the system is just identified if and only if each equation is just idenfied: $L_g = K_g, g = 1, \ldots, G$ and the rank condition holds for each equation.

• When estimating the SUR system there is another case where S2SLS – which is 2SLS estimation of each equation – coincides with 3SLS, regardless of the degree of overidentification. The 3SLS estimator is equivalent to 2SLS equation-by-equation when $\hat{\mathbf{\Omega}}$ is a diagonal matrix, that is, $\hat{\mathbf{\Omega}} = diag(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_G^2)$.

• Therefore, if $\mathbf{\Omega}$ is diagonal, 2SLS and 3SLS are asymptotically equivalent.

- In cases where 2SLS on each equation is not algebraically or asymptotically equivalent to 3SLS, it is not necessarily true that we should prefer the 3SLS estimator (or the minimum chi-square estimator more generally). Why? For estimation of any particular equation, the system procedures generally require $E(\mathbf{z}_g' u_g) = 0$ for all g, not just the equation of interest.
- As with system OLS and FGLS, there is a trade-off between robustness and efficiency.
- When IVs are available that are exogenous in each equation in the zero conditional mean sense, can characterize the optimal IVs (and then no weighting matrix is needed.)

8. PANEL DATA IV APPLICATION

- Apply pooled IV to estimate the passenger demand function,

$$lpassen_{it}) = \theta_t + \beta_1 lfare_{it} + \beta_2 ldist_i + \beta_3 ldist_i^2 + u_{it},$$

first using $concen_{it}$ as an IV for $lfare_{it}$ and then using as many lags as is available for each t.

```
. tab year

1997, 1998, |
 1999, 2000 |      Freq.     Percent        Cum.
------------+-----------------------------------
       1997 |      1,149       25.00       25.00
       1998 |      1,149       25.00       50.00
       1999 |      1,149       25.00       75.00
       2000 |      1,149       25.00      100.00
------------+-----------------------------------
      Total |      4,596      100.00
```

```
. * Reproduce the first-stage regression with fully robust inference:

. reg lfare concen ldist ldistsq y98 y99 y00, cluster(id)

Linear regression                                      Number of obs =    4596
                                                       F(  6,  1148) =  205.63
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4062
                                                       Root MSE      =  .33651

                                  (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .3601203    .058556     6.15   0.000     .2452315    .4750092
       ldist |  -.9016004   .2719464    -3.32   0.001    -1.435168   -.3680328
     ldistsq |   .1030196   .0201602     5.11   0.000     .0634647    .1425745
         y98 |   .0211244   .0041474     5.09   0.000     .0129871    .0292617
         y99 |   .0378496   .0051795     7.31   0.000     .0276872     .048012
         y00 |     .09987   .0056469    17.69   0.000     .0887906    .1109493
       _cons |   6.209258   .9117551     6.81   0.000     4.420364    7.998151
------------------------------------------------------------------------------

.* Plenty of partial correlation between lfare and concen.
```

```
. * Now pooled IV, usual, heteroskedasticity-robust, and fully robust
. * (heteroskedasticity and serial correlation) inference:

. ivreg lpassen ldist ldistsq y98 y99 y00 (lfare = concen)

Instrumental variables (2SLS) regression

      Source |       SS       df       MS              Number of obs =    4596
-------------+------------------------------           F(  6,  4589) =   20.45
       Model | -556.334915     6 -92.7224858           Prob > F      =  0.0000
    Residual |  4147.02233  4589  .903687586           R-squared     =       .
-------------+------------------------------           Adj R-squared =       .
       Total |  3590.68741  4595  .781433605           Root MSE      =  .95062

------------------------------------------------------------------------------
     lpassen |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lfare |  -1.776549   .2358788    -7.53   0.000    -2.238985   -1.314113
       ldist |  -2.498972   .4058371    -6.16   0.000    -3.294607   -1.703336
     ldistsq |   .2314932   .0345468     6.70   0.000     .1637648    .2992216
         y98 |   .0616171   .0400745     1.54   0.124    -.0169481    .1401824
         y99 |   .1241675   .0405153     3.06   0.002      .044738    .2035971
         y00 |   .2542695   .0456607     5.57   0.000     .1647525    .3437865
       _cons |   21.21249   1.891586    11.21   0.000     17.50407     24.9209
------------------------------------------------------------------------------
Instrumented:  lfare
Instruments:   ldist ldistsq y98 y99 y00 concen
------------------------------------------------------------------------------
```

```
. ivreg lpassen ldist ldistsq y98 y99 y00 (lfare = concen), robust

Instrumental variables (2SLS) regression               Number of obs =    4596
                                                       F(  6,  4589) =   18.23
                                                       Prob > F      =  0.0000
                                                       R-squared     =       .
                                                       Root MSE      =  .95062

------------------------------------------------------------------------------
             |               Robust
     lpassen |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lfare |  -1.776549   .2500745    -7.10   0.000    -2.266815   -1.286283
       ldist |  -2.498972   .4233497    -5.90   0.000    -3.328941   -1.669002
     ldistsq |   .2314932   .0361533     6.40   0.000     .1606153    .3023711
         y98 |   .0616171   .0400086     1.54   0.124    -.016819     .1400532
         y99 |   .1241675   .0408092     3.04   0.002     .0441618    .2041733
         y00 |   .2542695   .0469737     5.41   0.000     .1621784    .3463606
       _cons |   21.21249   1.997197    10.62   0.000     17.29702    25.12795
------------------------------------------------------------------------------
Instrumented:  lfare
Instruments:   ldist ldistsq y98 y99 y00 concen
------------------------------------------------------------------------------
```

```
. ivreg lpassen ldist ldistsq y98 y99 y00 (lfare = concen), cluster(id)

Instrumental variables (2SLS) regression               Number of obs =    4596
                                                       F(  6,  1148) =   28.02
                                                       Prob > F      =  0.0000
                                                       R-squared     =       .
                                                       Root MSE      =  .95062

                                  (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
     lpassen |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lfare |  -1.776549   .4753368    -3.74   0.000    -2.709175   -.8439226
       ldist |  -2.498972    .831401    -3.01   0.003    -4.130207   -.8677356
     ldistsq |   .2314932   .0705247     3.28   0.001     .0931215    .3698649
         y98 |   .0616171   .0131531     4.68   0.000     .0358103    .0874239
         y99 |   .1241675   .0183335     6.77   0.000     .0881967    .1601384
         y00 |   .2542695   .0458027     5.55   0.000      .164403    .3441359
       _cons |   21.21249   3.860659     5.49   0.000     13.63775    28.78722
------------------------------------------------------------------------------
Instrumented:  lfare
Instruments:   ldist ldistsq y98 y99 y00 concen
------------------------------------------------------------------------------
```

```
. * Now estimate a reduced form separately for each year. First generate
. * the lags of concen.

. xtset id year
       panel variable:  id (strongly balanced)
        time variable:  year, 1997 to 2000
                delta:  1 unit

. gen concen_1 = l.concen
(1149 missing values generated)

. gen concen_2 = l2.concen
(2298 missing values generated)

. gen concen_3 = l3.concen
(3447 missing values generated)
```

```
. reg lfare concen ldist ldistsq if year == 1997

      Source |       SS       df       MS              Number of obs =    1149
-------------+------------------------------           F(  3,  1145) =  262.60
       Model |  99.8444059     3  33.2814686           Prob > F      =  0.0000
    Residual |  145.115611  1145  .126738525           R-squared     =  0.4076
-------------+------------------------------           Adj R-squared =  0.4060
       Total |  244.960016  1148  .213379805           Root MSE      =    .356

------------------------------------------------------------------------------
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .3950364   .0627179     6.30   0.000     .2719815    .5180914
       ldist |  -.9360734   .2718439    -3.44   0.001    -1.469441   -.4027054
     ldistsq |     .10807   .0206224     5.24   0.000     .0676079     .148532
       _cons |   6.190051   .8898786     6.96   0.000     4.444075    7.936026
------------------------------------------------------------------------------

. predict lfareh97
(option xb assumed; fitted values)
```

```
. reg lfare concen concen_1 ldist ldistsq if year == 1998

      Source |       SS       df       MS              Number of obs =    1149
-------------+------------------------------           F(  4,  1144) =  218.70
       Model |  90.6172553     4  22.6543138           Prob > F      =  0.0000
    Residual |  118.502855  1144  .103586411           R-squared     =  0.4333
-------------+------------------------------           Adj R-squared =  0.4313
       Total |   209.12011  1148  .182160374           Root MSE      =  .32185

------------------------------------------------------------------------------
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .1864114    .123059     1.51   0.130    -.0550353    .4278581
    concen_1 |   .2103054   .1205194     1.74   0.081    -.0261586    .4467693
       ldist |  -.9217479   .2457973    -3.75   0.000    -1.404012   -.4394837
     ldistsq |   .1053951   .0186503     5.65   0.000     .0688024    .1419878
       _cons |   6.236879   .8045257     7.75   0.000     4.658368    7.815391
------------------------------------------------------------------------------

. predict lfareh98
(option xb assumed; fitted values)
(1149 missing values generated)
```

```
. reg lfare concen concen_1 concen_2 ldist ldistsq if year == 1999

      Source |       SS       df       MS              Number of obs =    1149
-------------+------------------------------           F(  5,  1143) =  162.41
       Model |  89.5585078     5  17.9117016           Prob > F      =  0.0000
    Residual |  126.059565  1143  .110288333           R-squared     =  0.4154
-------------+------------------------------           Adj R-squared =  0.4128
       Total |  215.618073  1148  .187820621           Root MSE      =   .3321

------------------------------------------------------------------------------
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |  -.1435733   .1477566    -0.97   0.331    -.4334778    .1463312
    concen_1 |   .3274273   .1685066     1.94   0.052    -.0031897    .6580444
    concen_2 |   .2672412   .1267665     2.11   0.035     .0185201    .5159624
       ldist |  -.8896912   .2538571    -3.50   0.000    -1.387769    -.391613
     ldistsq |   .1030365   .0192529     5.35   0.000     .0652615    .1408115
       _cons |   6.104077   .8317747     7.34   0.000       4.4721    7.736054
------------------------------------------------------------------------------

. predict lfareh99
(option xb assumed; fitted values)
(2298 missing values generated)
```

```
. reg lfare concen concen_1 concen_2 concen_3 ldist ldistsq if year == 2000

      Source |       SS       df       MS              Number of obs =    1149
-------------+------------------------------           F(  6,  1142) =  122.92
       Model |  78.3465194     6  13.0577532           Prob > F      =  0.0000
    Residual |   121.31513  1142  .106230412           R-squared     =  0.3924
-------------+------------------------------           Adj R-squared =  0.3892
       Total |   199.66165  1148  .173921298           Root MSE      =  .32593

------------------------------------------------------------------------------
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |  -.4329714    .134417    -3.22   0.001    -.6967035   -.1692394
    concen_1 |   .1853204   .1781258     1.04   0.298    -.1641701    .5348108
    concen_2 |   .3333175   .1655192     2.01   0.044     .0085616    .6580735
    concen_3 |   .3429416   .1254792     2.73   0.006      .096746    .5891373
       ldist |  -1.135407   .2493245    -4.55   0.000    -1.624593   -.6462217
     ldistsq |   .1189046   .0189071     6.29   0.000     .0818079    .1560012
       _cons |   7.103373   .8171934     8.69   0.000     5.500005    8.706742
------------------------------------------------------------------------------

. predict lfareh00
(option xb assumed; fitted values)
(3447 missing values generated)
```

```
. gen lfareh = lfareh97

. replace lfareh = lfareh98 if year == 1998
(1149 real changes made)

. replace lfareh = lfareh99 if year == 1999
(1149 real changes made)

. replace lfareh = lfareh00 if year == 2000
(1149 real changes made)
```

```
. ivreg lpassen ldist ldistsq y98 y99 y00 (lfare = lfareh), cluster(id)

Instrumental variables (2SLS) regression              Number of obs =    4596
                                                       F(  6,  1148) =   31.11
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0242
                                                       Root MSE      =   .8738

                                  (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
     lpassen |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lfare |  -1.082542    .358565    -3.02   0.003    -1.786058   -.3790258
       ldist |  -1.955129   .7552349    -2.59   0.010    -3.436925   -.4733338
     ldistsq |   .1691905   .0618604     2.74   0.006     .0478184    .2905625
         y98 |   .0447243   .0097519     4.59   0.000     .0255908    .0638577
         y99 |   .0998176   .0135394     7.37   0.000     .0732527    .1263824
         y00 |    .187701   .0337105     5.57   0.000     .1215598    .2538422
       _cons |   16.88214   3.229308     5.23   0.000     10.54613    23.21815
------------------------------------------------------------------------------
Instrumented:  lfare
Instruments:   ldist ldistsq y98 y99 y00 lfareh
------------------------------------------------------------------------------
```