ECON 4160, Spring term 2015 Lecture 7

Identification and estimation of SEMs (Part 1)

Ragnar Nymoen

Department of Economics

8 Oct 2015

- HN Ch 15
- References to Davidson and MacKinnon,
  - Ch 8.1-8.5
  - Ch 12.4.15
- (BN2011 kap 9; BN2014 Kap 7.9}

Structural form and reduced form I

- Consider the following two equations for simultaneous equilibrium in a single marked (partial equilibrium)

$$Q_t = \beta_{11} + \beta_{12} P_t + \varepsilon_{dt} \text{ (demand)} \tag{1}$$
$$Q_t = \beta_{21} + \beta_{22} P_t + \varepsilon_{st} \text{ (supply)} \tag{2}$$

- ε_{dt} and ε_{st} are white-noise processes, e.g., $(\varepsilon_{dt}, \varepsilon_{st})' \sim IN(\mathbf{0}, \boldsymbol{\Sigma})$.
- In this lecture we will not require from the outset that $\boldsymbol{\Sigma}$ is diagonal: We want to study identification both without imposing any restrictions of the covariance matrix $\boldsymbol{\Sigma}$, and with such restrictions (typically that $\boldsymbol{\Sigma}$ is diagonal).

Structural form and reduced form II

- In matrix notation, the model is
$$\begin{pmatrix} 1 & -\beta_{12} \\ 1 & -\beta_{22} \end{pmatrix} \begin{pmatrix} Q_t \\ P_t \end{pmatrix} = \begin{matrix} \beta_{11} + \varepsilon_{dt} \\ \beta_{21} + \varepsilon_{st} \end{matrix}$$
- Since the market is always in equilibrium, we observe not (1) and (2), but a sequence of equilibrium variables $\{\dots, (Q_{t-1}, P_{t-1}), (Q_t, P_t), (Q_{t+1}, P_{t+1}), \dots\}$ from the reduced-form of the simultaneous equations model

$$\begin{pmatrix} Q_t \\ P_t \end{pmatrix} = \begin{pmatrix} 1 & -\beta_{12} \\ 1 & -\beta_{22} \end{pmatrix}^{-1} \begin{pmatrix} \beta_{11} + \varepsilon_{dt} \\ \beta_{21} + \varepsilon_{st} \end{pmatrix} \tag{3}$$

Since

$$\begin{pmatrix} 1 & -\beta_{12} \\ 1 & -\beta_{22} \end{pmatrix}^{-1} = \begin{pmatrix} -\frac{\beta_{22}}{\beta_{12}-\beta_{22}} & \frac{\beta_{12}}{\beta_{12}-\beta_{22}} \\ -\frac{1}{\beta_{12}-\beta_{22}} & \frac{1}{\beta_{12}-\beta_{22}} \end{pmatrix}$$

Structural form and reduced form III

the Reduced Form (RF) becomes:

$$Q_t = \frac{\beta_{12}\beta_{21} - \beta_{11}\beta_{22}}{\beta_{12} - \beta_{22}} + \frac{\beta_{12}\varepsilon_{st} - \beta_{22}\varepsilon_{dt}}{\beta_{12} - \beta_{22}} \tag{4}$$

$$P_t = \frac{\beta_{21} - \beta_{11}}{\beta_{12} - \beta_{22}} + \frac{\varepsilon_{st} - \varepsilon_{dt}}{\beta_{12} - \beta_{22}} \tag{5}$$

- **The identification issue:** From the RF (4)-(5), can we obtain consistent estimators of the parameters of the simultaneous equation model (1)-(2)?
- REMARK: We discuss identification as a logical property of the theoretical model, as in HN and BN. This corresponds to the term "Asymptotic Identification" (page 529) used by DM.

Under-identification I

- Given our assumptions, OLS estimators of the RF parameters in (4) and (5) will be MLEs that are consistent estimators.
- Let us therefore assume a **perfect sample** that gives us knowledge of the *plim*-values of the OLS estimators of $E(Q_t)$ and $E(P_t)$. Denote these *plim*-values by γ_{Q0} and γ_{P0}

$$\gamma_{Q0} = \frac{\beta_{12}\beta_{21} - \beta_{11}\beta_{22}}{\beta_{12} - \beta_{22}} \tag{6}$$

$$\gamma_{P0} = \frac{\beta_{21} - \beta_{11}}{\beta_{12} - \beta_{22}} \tag{7}$$

  - Two equations in **two known RF parameters**, γ_{Q0} and γ_{P0}, and **four unknown structural parameters**.

Under-identification II

- Cannot determine β_{11}, β_{12}, β_{21} and β_{22} from (6) and (7) even if we have perfect knowledge of the reduced form parameters γ_{Q0} and γ_{P0}
- The parameters of (1) and/or (2) cannot be identified.
- Therefore, neither of the equations in the SEM are identified.

Partial identification by imposing restrictions I

- Let us assume completely inelastic supply, $\beta_{22} = 0$. The SEM is now

$$Q_t = \beta_{11} + \beta_{12} P_t + \varepsilon_{dt} \text{ (demand)} \tag{8}$$
$$Q_t = \beta_{21} + \varepsilon_{st} \text{ (supply)} \tag{9}$$

and keep $(\varepsilon_{dt}, \varepsilon_{st})' \sim IN(\mathbf{0}, \mathbf{\Sigma})$ as before.

- RF for this case:

$$\gamma_{Q0} = \frac{\beta_{12}\beta_{21}}{\beta_{12}} = \beta_{21} \tag{10}$$
$$\gamma_{P0} = \frac{\beta_{21} - \beta_{11}}{\beta_{12}} \tag{11}$$

Partial identification by imposing restrictions II

- Assume again that γ_{Q0} and γ_{P0} are known (the same perfect sample assumption as above).
  - The structural parameter β_{21} is found (determined) from (10), hence the supply equation (9) is identified
  - In (11) we still have "two unknowns" and only one equation: The demand equation (8) is not identified
- This is an example of partial identification (i.e. of one equation namely (9) in a SEM.

Full identification by imposing restrictions I

- In addition to $\beta_{22} = 0$ we assume that $\boldsymbol{\Sigma}$ is diagonal:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2_{d\varepsilon} & 0 \\ 0 & \sigma^2_{s\varepsilon} \end{pmatrix}$$

- Does this lead to (more) identification?

Full identification by imposing restrictions II

- To find the answer, write down all the first and second order parameters (moments) from the reduced form

$$E(Q_t) \equiv \gamma_{Q0} = \beta_{21} \tag{12}$$

$$E(P_t) \equiv \gamma_{P0} = \frac{\beta_{21} - \beta_{11}}{\beta_{12}} \tag{13}$$

$$Var(Q_t) = \frac{\beta_{12}^2 \sigma_{s\varepsilon}^2}{\beta_{12}^2} \tag{14}$$

$$Var(P_t) = \frac{\sigma_{s\varepsilon}^2 - \sigma_{d\varepsilon}^2}{\beta_{12}^2} \tag{15}$$

$$Cov(Q_t, P_t) = \frac{\beta_{12} \sigma_{s\varepsilon}^2}{\beta_{12}^2} \tag{16}$$

Full identification by imposing restrictions III

- We now have a determined equation system for the RF parameters and the structural form parameters: 5 independent equations in the five unknown structural parameters: β_{11}, β_{12}, β_{21}, $\sigma^2_{s\varepsilon}$ and $\sigma^2_{d\varepsilon}$.
- Both equations are therefore identified, we call this *exact-identification* also called *just-identification*.

Interpreting the example of identification by restrictions: recursive structure I

- The restrictions $\beta_{21} = 0$ (inelastic supply) together with diagonal $\mathbf{\Sigma}$ makes the structure become **recursive:**
  - In expectation, supply is fixed from period to period, and there is no indirect correlation between ε_{st} and ε_{dt} via $\mathbf{\Sigma}$
- Using our new concepts, Q_t is *strictly exogenous* in a regression between P_t and Q_t, and the *plim* of the OLS estimated regression coefficient is

$$\frac{Cov(Q_t, P_t)}{Var(Q_t)} = \frac{\frac{\beta_{12}\sigma^2_{s\varepsilon}}{\beta^2_{12}}}{\frac{\beta^2_{12}\sigma^2_{s\varepsilon}}{\beta^2_{12}}} = \frac{\frac{\sigma^2_{s\varepsilon}}{\beta_{12}}}{\sigma^2_{s\varepsilon}} = \frac{1}{\beta_{12}}$$

Interpreting the example of identification by restrictions: recursive structure II

- Hence, the slope parameter $\beta_{11}^{'}$ of the *inverted demand function*

$$P_t = \beta_{11}' + \beta_{12}^{'} Q_t + \varepsilon_{st}'$$

  is consistently estimated by OLS, and a consistent estimator of β_{12} is the reciprocal of that OLS estimator $\widehat{\beta_{12}^{'}}$.
- It is also MLE, given the normal/gaussian properties of the disturbances.

Summing up the first example I

- Identification is a logical property of the structural (theory) model. It is **not** a property of the sample.
- Consistent estimation of the RF parameters is necessary for identification of structural parameters.
- Restriction on one or more of the structural parameters (β_{22} above), may lead to at least partial identification
- Restriction on the covariance matrix of the structural disturbances ($\sigma_{sd} = 0$, above) can also increase the degree of identification.
- Next: Motivate an easy-to-use method of checking identification *without* considering any restrictions on the Σ matrix for the disturbances

Summing up the first example II

- This is the method of the Order- and Rank-conditions
- Before presenting the general result, we will develop our understanding in steps.

Another partially identified structure I

- Before we restricted (1) and (2), call it *Structure 1*, none of the structural parameters were identified. We can make it more concrete by writing it as

$$\textit{Structure 1}$$

$$Q_t = 15 - P_t + \varepsilon_{dt}$$
$$Q_t = -0.2 + 0.5P_t + \varepsilon_{st}$$

  Look at a picture of what a scatter plot from this structure could look like.

- Intuitively, the lack of identification is due to the fact that there is no *independent variation* in the supply schedule that can help "trace out" the demand curve, and vice versa for the supply curve.

Another partially identified structure II

- When we discuss identification, it can be use an "identification table"

	Q	1	P
Demand equation	1	β_{11}	β_{12}
Supply equation	1	β_{21}	β_{22}

- We regard the constant term as a variable, denoted by the "1" in the table.
- In the detailed *Structure 1* $\beta_{11} = 15$, $\beta_{22} = 0.5$ and so on.

Another partially identified structure III

- Now consider a second structure:

$$\begin{aligned} &\textit{Structure 2} \\ Q_t &= 15 - P_t + 1.5X_{dt} + \varepsilon_{dt} \\ Q_t &= -0.2 + 0.5P_t + \varepsilon_{st} \end{aligned}$$

where X_{dt} is an observable variable, so the identification table is

	Q	1	P	X_d
Demand equation	1	β_{11}	β_{12}	β_{13}
Supply equation	1	β_{21}	β_{22}	0

where the exclusion X_d from the supply equation is market by the "0".

Another partially identified structure IV

- Intuition: the supply equation is now identified, because the demand equation shifts due to changes in X_d, and the supply is unaffected by the X_d shifts
- More formally: We cannot obtain an observational equivalent supply-equation by taking a linear combination of the two relationships in *Structure 2* (expect by giving zero weight to the demand equation).
- However, such a linear relationship is observational equivalent to the *Demand equation* which therefore is not identified in *Structure 2.*

What about?

Structure 3

$$Q_t = 15 - P_t + \varepsilon_{dt}$$
$$Q_t = -0.2 + 0.5P_t + 2X_{st} + \varepsilon_{st}$$

	Q	1	P	X_s
Demand equation	1	β_{11}	β_{12}	0
Supply equation	1	β_{21}	β_{22}	β_{23}

Finally the case of "exact identification" I

Structure 4

$$Q_t = 15 - P_t + 1.5X_{dt} + \varepsilon_{dt}$$
$$Q_t = -0.2 + 0.5P_t + 2X_{st} + \varepsilon_{st}$$

	Q	1	P	X_d	X_s
Demand equation	1	β_{11}	β_{12}	β_{13}	0
Supply equation	1	β_{21}	β_{22}	0	β_{24}

- No linear combination can "give back" the original two structural equations, hence both are identified

The order and rank conditions for exact identification

- The following identification rule suggests itself: In a SEM with N equations and therefore N endogenous variables, any structural equation is identified if that equation excludes $N-1$ variables. The excluded variables can be endogenous or exogenous.
- This is called the **order condition**.
- An equivalent formulation of the order condition is that equation number i is identified if

$$\underbrace{(K - K_i)}_{\text{excluded exogenous}} = \underbrace{N_i - 1}_{\text{included endogenous minus one}}$$

- In fact the order condition is only necessary. The necessary and sufficient condition says that the excluded variables in the equation under inspection must have coefficients that are different from zero in the other equations in the SEM.
- This is the **rank condition** for identification

Overidentification

- We typically distinguish between just (or exact) identification and overidentification.
- **Overidentification** means that we can derive, from the RF parameters, more than one solution for the structural parameters.
- Hence the RF form in this case has *more* information than we need in order to estimate the structural parameters consistently.
- Over-identification is not an obstacle to estimation of the SEM, as we shall see later, the only problem is how to use the information to give efficient estimation (lowest possible standard errors).

Order and rank condition, general formulation I

Rank condition

In a SEM with N linear equations, an equation is identified if and only if at least one non-zero $(N-1) \times (N-1)$ determinant is contained in the array of the coefficients which those variables excluded from the equation in question appear in the other equations of the SEM.

Remarks:

- Recall that the rank of a matrix is the order of the largest non-zero determinant that it contains.
- If the rank condition is satisfied, the order condition is automatically satisfied, but not vice versa
- If the order of the non-zero determinant is larger than $(N-1)$ the equation is *over-identified.*

Order and rank condition, general formulation II

- Identities are often part of SEMs. They are counted among the N equation of the model. Identities are identified equations, but the identification of the *other* structural equations should be investigated with the identities taken into account.
- Exclusion restrictions are a special case of linear restrictions on the parameters and an even more general formulation of the order condition is:
  In a SEM with N linear equations, a necessary condition for identification of an equation is that there are $(N-1) \times (N-1)$ linearly independent restrictions on the parameters of the equation.

Order and rank condition, general formulation III

- As noted, in PcGive, identification is always checked prior to estimation, hence it is asymptotic identification that is checked by the program.

Identification of dynamic SEMs

- Predetermined variables count as exogenous variables when we investigate identification
- This is because identification is about obtaining consistent estimators of structural parameters, and as we know, predeterminedness is enough for consistent estimation of RF parameters, cf. the estimation theory of for VARs!
- Therefore, the order and rank conditions apply to dynamic simultaneous equations models

Identification in recursive systems I

- In our first example, we saw that the supply-demand system could be identified if the structure was recursive
- Recursive generally requires two things:
  1. The matrix of contemporaneous coefficients must be (upper or lower) triangular)
  2. The $\mathbf{\Sigma}$ matrix must be diagonal
- Note from previous lectures, that this is exactly what we achieve if we model the system in terms of conditional and marginal models.

Identification in recursive systems II

- Suppose that the RF of a dynamic SEM for Y_t and X_t is the stationary VAR with white-noise Gaussian:

$$\underbrace{\begin{pmatrix} Y_t \\ X_t \end{pmatrix}}_{\mathbf{y}_t} = \underbrace{\begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix}}_{\boldsymbol{\Pi}} \underbrace{\begin{pmatrix} Y_{t-1} \\ X_{t-1} \end{pmatrix}}_{\mathbf{y}_{t-1}} + \underbrace{\begin{pmatrix} \varepsilon_{yt} \\ \varepsilon_{xt} \end{pmatrix}}_{\boldsymbol{\varepsilon}_t}, \tag{17}$$

- $Cov(\varepsilon_{yt}, \varepsilon_{xt}) \neq 0$ usually. but if we represent the VAR in terms of a conditional ADL and the second line in the VAR

$$Y_t = \phi_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \varepsilon_t \tag{18}$$

$$X_t = \pi_{21} Y_{t-1} + \pi_{22} X_{t-1} + \varepsilon_{xt} \tag{19}$$

Identification in recursive systems III

we have a model with a recursive structure

$$Y_t - \beta_0 X_t = \phi_1 Y_{t-1} + \beta_1 X_{t-1} + \varepsilon_t \tag{20}$$

$$X_t = \pi_{21} Y_{t-1} + \pi_{22} X_{t-1} + \varepsilon_{xt} \tag{21}$$

since $Cov(\varepsilon_t, \varepsilon_{xt}) = 0$ from the construction of the conditional model

- Because of $Cov(\varepsilon_t, \varepsilon_{xt}) = 0$ OLS on the marginal model and ADL separately give consistent estimators of all the parameters.
- One way to put this, is that, if our parameters of interest are the parameters of the conditional-marginal equation, then that model is the structure, and it is identified.

Notation for a structural equation I

- Without loss of generality, we consider equation # 1 in a SEM
- Apart from the disturbance, adopt the notation in DM page 522:

$$\mathbf{y}_1 = \mathbf{Z}_1 \boldsymbol{\beta}_{11} + \mathbf{Y}_1 \boldsymbol{\beta}_{21} + \boldsymbol{\epsilon}_1 \tag{22}$$

- $\mathbf{y}_1$ is $n \times 1$, with observations of the variable that # 1 in the SEM is normalized on.
- $\mathbf{Z}_1$ is $n \times k_{11}$ with observations of the k_{11} included predetermined or exogenous variables.
- $\mathbf{Y}_1$, $n \times k_{12}$ holds the included endogenous explanatory variables.

Notation for a structural equation II

The total number of explanatory variable in the first equation is

$$k_{11} + k_{12} = k_1 \tag{23}$$

For simplicity assume that the structural disturbance is Gaussian white-noise.

$$\boldsymbol{\epsilon}_1 = IN(\mathbf{0}, \sigma_1^2 \mathbf{I}_{nxn})$$

Simultaneity bias of OLS estimators I

By defining the two partitioned matrices:

$$\mathbf{X}_1 = \begin{pmatrix} \mathbf{Z}_1 & \mathbf{Y}_1 \end{pmatrix} \tag{24}$$

$$\boldsymbol{\beta}_1 = \begin{pmatrix} \boldsymbol{\beta}_{11} & \boldsymbol{\beta}_{21} \end{pmatrix}' \tag{25}$$

(22) can be written compactly as

$$\mathbf{y}_1 = \mathbf{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{\epsilon}_\mathbf{1} \tag{26}$$

Simultaneity bias of OLS estimators II

- (26) looks like an ordinary regression, **but of course we know better**: Since it is the first structural equation in a SEM (a system!) we have

$$plim(\frac{1}{n}\mathbf{X}_1'\epsilon_1) \neq \mathbf{0} \quad (27)$$

since $\mathbf{X}_1$ includes k_{12} endogenous explanatory variables.

- Hence the OLS estimator $\hat{\boldsymbol{\beta}}_1$ will be inconsistent by simultaneous equations bias.
- All the estimators will be affected, not only $\hat{\beta}_{21}$.

Medium-term macro model example

- Medium term macro model of the Keynesian type.
  - C_t : private consumption in year t (in constant prices)
  - GDP_t, TAX_t and I_t are *gross domestic product*, *net taxes* and *investments and gov.exp.*
  - a - e are parameters of the macroeconomic model
  - ϵ_{Ct} and ϵ_{TAXt} are independent disturbances with *classical properties* conditional on I_t and C_{t-1}.

$$C_t = a + b(GDP_t - TAX_t) + cC_{t-1} + \epsilon_{Ct} \tag{28}$$

$$TAX_t = d + eGDP_t + \epsilon_{TAXt} \tag{29}$$

$$GDP_t = C_t + I_t \tag{30}$$

- C_t, GDP_t and TAX_t are endogenous, C_{t-1} is predetermined.
- Assume that I_t is strictly exogenous with $E(I_t) = \mu_I$ and $Var(I_t) = \sigma_I^2$. For simplicity, we will use

$$I_t = \mu_I + \epsilon_{It} \tag{31}$$

where is independent of ϵ_{Ct} and ϵ_{TAXt}, and has classical properties conditional on I_t and C_{t-1}.

Reduced form

- The reduced form equation for C_t and GDP_t

$$
\begin{aligned}
C_t &= \frac{a+bd}{(1-b(1-e))} + \frac{b(1-e)}{(1-b(1-e))} I_t + \frac{c}{(1-b(1-e))} C_{t-1} \\
&+ \frac{\epsilon_{Ct} - (be)\epsilon_{TAXt}}{(1-b(1-e))} \\
GDP_t &= \frac{a+bd}{(1-b(1-e))} + \frac{(1-be)}{(1-b(1-e))} I_t + \frac{c}{(1-b(1-e))} C_{t-1} \\
&+ \frac{\epsilon_{Ct} - (be)\epsilon_{TAXt}}{(1-b(1-e))}
\end{aligned}
$$

- Both C_t and GDP_t depend on ϵ_{Ct} (and ϵ_{TAXt}).

Structural parameters I

- Assume now that our parameter of interest is b, the 'marginal propensity to consume" . For exposition, we use a simplified and static structural model

$$C_t = a + b(GDP_t) + \epsilon_{Ct} \tag{32}$$
$$GDP_t = C_t + I_t \tag{33}$$
$$I_t = \mu_I + \epsilon_{It} \tag{34}$$

- Assumptions about structural disturbances:

$$E(\epsilon_{Ct} \mid I_t) = 0,\ \ Var(\epsilon_{Ct} \mid I_t) = \sigma_C^2 \tag{35}$$
$$E(\epsilon_{It}) = 0,\ \ Var(\epsilon_{Ct}) = \sigma_I^2 \tag{36}$$
$$Kov(\epsilon_{Ct}, \epsilon_{It}) = 0 \tag{37}$$

- In this model, the OLS estimator of b is inconsistent.

OLS estimator in the macro model example

We know that the OLS estimator is

$$\hat{b} = \frac{\sum_{t=1}^{T}(GDP_t - \overline{GDP})C_t}{\sum_{t=1}^{T}\left(GDP_t - \overline{GDP}\right)^2} = b + \frac{\sum_{t=1}^{T} GDP_t(\epsilon_{Ct} - \bar{\epsilon}_C)}{\sum_{t=1}^{T}\left(GDP_t - \overline{GDP}\right)^2}$$

To assess the probability limit of the bias term we use the reduced form:

$$GDP_t = \frac{a + \mu_I}{1-b} + \frac{\epsilon_{Ct}}{1-b} + \frac{\epsilon_{It}}{1-b}, \text{ for } 0 < b < 1$$

$$C_t = \frac{(a + b\mu_I)}{1-b} + \frac{\epsilon_{Ct}}{1-b} + \frac{b}{1-b}\epsilon_{It}$$

which together with the assumptions give the properties of the two random variables GDP_t and C_t

Simultaneous equations bias in macro example

$$plim(\hat{b} - b) = plim \frac{\sum_{t=1}^{T} GDP_t(\epsilon_{Ct} - \bar{\epsilon}_C)}{\sum_{t=1}^{T} \left(GDP_t - \overline{GDP}\right)^2}$$

$$plim(\hat{b} - b) = plim \frac{\sum_{t=1}^{T} \left(\frac{a+\mu_I}{1-b} + \frac{\epsilon_{Ct}}{1-b} + \frac{\epsilon_{It}}{1-b}\right)(\epsilon_{Ct} - \bar{\epsilon}_C)}{\sum_{t=1}^{T} \left(\frac{\epsilon_{Ct}-\bar{\epsilon}_C}{1-b} + \frac{\epsilon_{It}-\bar{\epsilon}_I}{1-b}\right)^2}$$

From the assumptions of the model:

$$plim(\hat{b} - b) = \frac{\sigma_C^2}{\sigma_C^2 + \sigma_I^2} > 0 \tag{38}$$

- OLS is overestimating the structural parameter b in the model given by (32)-(37)
- Trygve Haavelmo: *The statistical implications of system of simultaneous equations*, Econometrica (1943)
- Inconsistency of OLS is true for *all* simultaneous equations

Instrumental variables matrix I

- Since (22) is assumed to be exactly identified, it is logically consistent to assume that the SEM defines a matrix $\mathbf{W}_1$ with the properties

$$plim(\frac{1}{n}\mathbf{W}_1'\mathbf{X}_1) = \mathbf{S}_{W_1'X_1} \text{ (invertible)} \tag{39}$$

$$plim(\frac{1}{n}\mathbf{W}_1'\epsilon_1) = \mathbf{0} \text{ (independence)} \tag{40}$$

$$plim(\frac{1}{n}\mathbf{W}_1'\mathbf{W}_1) = \mathbf{S}_{W_1'W_1} \text{ (positive definite)} \tag{41}$$

- Since (39) requires invertiability, the number of columns in $\mathbf{W}_1$ must k_1.

Instrumental variables matrix II

- The k_{11} predetermined variables included in eq # 1 are of course in $\mathbf{W}_1$. In addition we need k_{12} instrumental variables for the included endogenous explanatory variables in eq # 1.
- The k_{12} instrumental variables must be "taken from" the predetermined variables in the SEM that are excluded from the first equation.
- But if structural equation # 1 is **just-identified**, the number of excluded predetermined variables is exactly equal to the number of included endogenous variables in the equation minus one. Hence we can write $\mathbf{W}_1$ as

$$\mathbf{W}_1 = \begin{pmatrix} \mathbf{Z}_1 & \mathbf{X}_{01} \end{pmatrix} \tag{42}$$

where $\mathbf{X}_{01}$ is $n \times k_{11}$(number of included endogenous variables minus one).

Instrumental variables matrix III

- We see that, defined in this way, $\mathbf{W}_1$ will satisfy both
  - **instrument relevance** (39), and
  - **instrument validity** in (40)
- In the absence of perfect collinearity among instruments, also (41).

The IV estimator I

If we apply the method of moments to the structural equation $\mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}_1$, but with $\mathbf{W}_1$ instead of $\mathbf{X}_1$:

$$\mathbf{W}_1' \left[\mathbf{y}_1 - \mathbf{X}_1\hat{\boldsymbol{\beta}}_{1,IV}\right] = \mathbf{0} \tag{43}$$

we obtain the IV-estimator

$$\hat{\boldsymbol{\beta}}_{1,IV} = (\mathbf{W}_1'\mathbf{X}_1)^{-1}\mathbf{W}_1'\mathbf{y}_1 \tag{44}$$

$\hat{\boldsymbol{\beta}}_{1,IV}$ is clearly a method-of-moments estimator. The only difference from OLS is that $\mathbf{W}_1'$ takes the place of $\mathbf{X}_1'$ in the IV "normal equations", or orthogonality conditions, (43).
This means that, by construction, the IV-residuals

The IV estimator II

$$\hat{\epsilon}_{\mathbf{IV,1}} = \mathbf{y}_1 - \mathbf{X}_1 \hat{\beta}_{1,IV} \tag{45}$$

are uncorrelated with (all the instruments in) $\mathbf{W}_1$:

$$\mathbf{W}_1' \hat{\epsilon}_{\mathbf{IV,1}} = \mathbf{0}. \tag{46}$$

Among the results for the IV-estimator are:

- $\hat{\beta}_{1,IV}$ is a consistent estimator ,but it is biased in a finite sample
- Asymptotic inference based on "t-ratios" and Chi-squared statistics for joint hypotheses are valid
- The estimated standard error of $\hat{\beta}_{1,IV}$ can be considerably larger than $\hat{\beta}_{1,OLS}$ if the instruments are **weak**, meaning that they are almost independent from the endogenous variables that they act as instruments for.

- DM show in equation (8.17) that the asymptotic covariance matrix of the vector $(\hat{\boldsymbol{\beta}}_{1,IV} - \boldsymbol{\beta}_1)$ is

$$Var(\hat{\boldsymbol{\beta}}_{1,IV} - \boldsymbol{\beta}_1) = \sigma_1^2 plim(n^{-1}\mathbf{X}_1'\mathbf{P}_{W_1}\mathbf{X}_1)^{-1} \tag{47}$$

  where:

$$\mathbf{P}_{W_1} = \mathbf{W}_1(\mathbf{W}_1'\mathbf{W}_1)^{-1}\mathbf{W}_1$$

  our old friend the *prediction maker*.
- Which role does it take here?

The IV residual maker I

- If we want, we can define the IV-residual maker as

$$\mathbf{M}_{IV,1} = \left[\mathbf{I} - \mathbf{X}_1(\mathbf{W}_1'\mathbf{X}_1)^{-1}\mathbf{W}_1'\right]$$

(Check that

$$\hat{\epsilon}_{\mathbf{IV,1}} = \mathbf{M}_{IV,1}\mathbf{y}_1)$$

- If $\mathbf{M}_{IV,1}$ is a proper residual maker, regression of $\mathbf{W}_1$ on $\mathbf{W}_1$ should result in zero-residuals.

The IV residual maker II

- Check:

$$\mathbf{M}_{IV,1}\mathbf{W}_1 = \left[\mathbf{I} - \mathbf{X}_1(\mathbf{W}_1'\mathbf{X}_1)^{-1}\mathbf{W}_1'\right]\mathbf{W}_1$$

Use that

$$\mathbf{M}_{IV,1}^{'} = \mathbf{M}_{IV,1}$$

(show for $k_1 = 2$ for example), then

$$\begin{aligned}
\mathbf{M}_{IV,1}\mathbf{W}_1 &= \mathbf{M}_{IV,1}^{'}\mathbf{W}_1 \\
&= \left[\mathbf{I} - \mathbf{W_1}\left\{(\mathbf{W}_1'\mathbf{X}_1)^{-1}\right\}'\mathbf{X}_1^{'}\right]\mathbf{W}_1 \\
&= \left[\mathbf{I} - \mathbf{W_1}\left\{(\mathbf{W}_1'\mathbf{X}_1)'\right\}^{-1}\mathbf{X}_1^{'}\right]\mathbf{W}_1 \\
&= \left[\mathbf{I} - \mathbf{W_1}\left\{\mathbf{X}_1'\mathbf{W}_1\right\}^{-1}\mathbf{X}_1^{'}\right]\mathbf{W}_1 \\
&= \mathbf{0}
\end{aligned}$$

The IV residual maker III

so that the orthogonality condition (43) can be interpreted as:

$$\mathbf{W}_1^{'}\hat{\epsilon}_{IV,1} = (\mathbf{M}_{IV,1}^{'}\mathbf{W}_1)'\mathbf{y}_1 = \mathbf{0} \tag{48}$$

confirming (46), and showing that the "only" difference compared to OLS is that the set of instruments used to form orthogonality conditions (normal equations) has been changed from "**X**" to "**W**".

Optimal instruments in the overidentified case I

- In the case of *overidentification*, $\mathbf{W}_1$ is $n \times l_1$ where $l_1 > k_1 = k_{11} + k_{12}$,
  $\Longrightarrow \mathbf{W}_1'\mathbf{X}_1$ is no longer quadratic.
- There is more that one moment-matrix (based on $\mathbf{W}_1'\mathbf{X}_1$) that are quadratic and invertible
- Each one defines a consistent IV-estimator of $\boldsymbol{\beta}_1$. We call this **over-identification**

Optimal instruments in the overidentified case II

- To solve this "luxury problem" we can define another IV matrix $\widehat{\mathbf{W}}_1$ that has dimension $n \times k_1$:

$$\widehat{\mathbf{W}}_1 = \begin{pmatrix} \mathbf{Z}_1 & \widehat{\mathbf{Y}}_1 \end{pmatrix}, \tag{49}$$

where $\widehat{\mathbf{Y}}_1$ is $n \times k_{12}$ and is made up of the best linear predictors of the k_{12} endogenous variables included in the first equation:

$$\widehat{\mathbf{Y}}_1 = \begin{pmatrix} \widehat{\mathbf{y}}_2 & \widehat{\mathbf{y}}_3 & \dots \end{pmatrix}_{n \times k_{12}} \tag{50}$$

Optimal instruments in the overidentified case III

- Where does the optimal predictors come from? Since we are looking at a single equation in a system-of-equations, they must come from the reduced form equations for the endogenous variables:

$$\widehat{\mathbf{y}}_j = \mathbf{W}_1 \widehat{\boldsymbol{\pi}}_j, \quad j = 2, ..., k_{12} + 1. \tag{51}$$

where

$$\widehat{\boldsymbol{\pi}}_j = (\mathbf{W}_1' \mathbf{W}_1)^{-1} \mathbf{W}_1' \mathbf{y}_j, \tag{52}$$

are the OLS estimators of the regression coefficients in the conditional expectation function for each included endogenous variables in the first equation, conditional on the full set of

Optimal instruments in the overidentified case IV

predetermined variables in the system of equations. We can write $\widehat{\mathbf{Y}}_1$ as:

$$\widehat{\mathbf{Y}}_1 = \begin{pmatrix} \mathbf{W}_1(\mathbf{W}_1'\mathbf{W}_1)^{-1}\mathbf{W}_1'\mathbf{y}_2 & \dots & \mathbf{W}_1(\mathbf{W}_1'\mathbf{W}_1)^{-1}\mathbf{W}_1'\mathbf{y}_{(k_{21}+1)} \end{pmatrix},$$

and more compactly:

$$\widehat{\mathbf{Y}}_1 = \mathbf{W}_1(\mathbf{W}_1'\mathbf{W}_1)^{-1}\mathbf{W}_1'\mathbf{Y}_1 = \mathbf{P}_{W_1}\mathbf{Y}_1$$

in terms of the prediction-maker:

$$\mathbf{P}_{W_1} = \mathbf{W}_1(\mathbf{W}_1'\mathbf{W}_1)^{-1}\mathbf{W}_1' \tag{53}$$

The GIV-estimator (GIVE) I

- We define the **Generalized IV estimator** as

$$\hat{\boldsymbol{\beta}}_{1,GIV} = (\widehat{\mathbf{W}}_1'\mathbf{X}_1)^{-1}\widehat{\mathbf{W}}_1'\mathbf{y}_1. \tag{54}$$

with

$$\widehat{\mathbf{W}}_1 = \begin{pmatrix} \mathbf{Z}_1 & \widehat{\mathbf{Y}}_1 \end{pmatrix},$$

and

$$\widehat{\mathbf{Y}}_1 = \mathbf{P}_{W_1}\mathbf{Y}_1$$

- $\hat{\boldsymbol{\beta}}_{1,GIV}$ is also known as the 2-stage least squares estimator of $\boldsymbol{\beta}_1$ in (26) that you have seen in operation in CC #4.