Efficient GMM estimation using incomplete data

Chris Muris (SFU)

July, 2016

Outline

Incomplete data

- An observation is **incomplete** if not all variables are observed
- Examples:
  - missing instruments

	Data availability			
	1	2	3	4
Instrument 1	X	X	·	·
Instrument 2	X	·	X	·

  - unbalanced panels
- Incomplete observations may be informative
- An observation is **missing** if no moment function can be computed from the observed data

Contribution

This paper proposes a **framework** for handling incomplete data in a moment condition setting. Under a MAR assumption, I obtain:

1. A set of moment conditions for incomplete data
2. The **efficiency bound** for those moment conditions
3. An **efficient estimator**

Example: Linear IV

- Endogenous variables $X = (X_1, X_2)$
- Instruments W_1 and W_2
- Regression coefficient β_0 satisfies:

$$E\begin{pmatrix} W_1\,(y - X\beta_0) \\ W_2\,(y - X\beta_0) \end{pmatrix} = 0 \tag{1}$$

- **Missing data**: both components of (1) available, or none

Example: Linear IV

- Three **strata** based on data availability.

Stratum 1 Both instruments W_1 and W_2 are observed
Stratum 2 Only the instrument W_1 is observed
Stratum 3 Only W_2 is observed

- Efficient estimator optimally combines the information from all strata.
- No **identification** in stratum 2, but $E\left[W_1\left(y - X\beta_0\right)\right] = 0$ is informative
- The approach is **general**:
  - arbitary number of strata
  - arbitrary set of moment conditions

Example: Dynamic panels

- Panel data AR(1) model:

$$y_{i,t} = \alpha_i + \rho y_{i,t-1} + u_{i,t},\ 2 \leq t \leq T \tag{2}$$

- Arellano and Bond (1991) based on moment conditions

$$E(y_{i,t-s}\Delta u_{i,t}) = 0,\ \ t \geq 3,\ s \geq 2 \tag{3}$$

- If $y_{i,t}$ is not observed, then several components are unavailable

Example: Dynamic panels

	Unavailable components				
	None	$y_{i,1}$	$y_{i,4}$	$(y_{i,1}, y_{i,4})$	$y_{i,2}$
$y_{i,1}\Delta u_{i,3}$	X	·	X	·	·
$y_{i,1}\Delta u_{i,4}$	X	·	·	·	X
$y_{i,1}\Delta u_{i,5}$	X	·	·	·	X
$y_{i,2}\Delta u_{i,4}$	X	X	·	·	·
$y_{i,2}\Delta u_{i,5}$	X	X	·	·	·
$y_{i,3}\Delta u_{i,5}$	X	X	·	·	X

Table: Strata for dynamic panels.

Literature

1. **Missing data.** Data is either complete, or completely missing. Robins et al. (1994), Hirano et al. (2003), Wooldridge (2007), Chen et al. (2008), Prokhorov and Schmidt (2009), **Graham (2011)**
   - My approach allows for an arbitrary number of strata
2. **Model-specific solutions.** Methods for attrition in panels (Verbeek and Nijman, 1992; Hirano et al., 2001, Abrevaya, 2016), dynamic panels (Pacini and Windmeijer, 2015), partially observed instruments (Mogstad and Wiswall, 2012; Abrevaya and Donald, 2015)
   - My approach allows for an arbitrary set of moment conditions
3. **Imputation-based methods.** Dagenais (1973), Gourieroux and Monfort (1981), Dardanoni et al. (2011)
   - I do not rely on imputation

Outline

Model: complete data moments

- $Z = \left(Y_1', X'\right)$ is a random vector
- β is an unknown parameter vector $(K \times 1)$
- $\psi(Z, \beta)$ is a vector of moment functions $(p \times 1,\ p \geq K)$

Assumption 1

$E(\psi(Z, \beta)) = 0 \Leftrightarrow \beta = \beta_0$

Model: data availability

- Not all elements of ψ are always observable
- $J + 1$ strata of incompleteness
- D is an incomplete data indicator with $J + 1$ outcomes $\{d_1, \cdots, d_{J+1}\}$
  - d_j is an $r \times r$ selection matrix that selects the elements of ψ that are observable
  - $d_{J+1} = O_r$
- Researcher observes $D\psi(Z, \cdot)$

Example: Linear IV

- Moment conditions

$$E\begin{pmatrix} W_1(y - X\beta_0) \\ W_2(y - X\beta_0) \end{pmatrix} = 0 \tag{4}$$

- D takes one of $J + 1 = 4$ values

$$\left\{ d_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, d_2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, d_3 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, d_4 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \right\}$$

- $D = d_1$ corresponds to observing all variables:

$$d_1\psi(Z, \beta) = \begin{pmatrix} W_1(y - X\beta) \\ W_2(y - X\beta) \end{pmatrix}$$

- $D = d_2$ corresponds to observing only the first instrument

$$d_2\psi(Z, \beta) = \begin{pmatrix} W_1(y - X\beta_0) \\ 0 \end{pmatrix}$$

Model: MAR assumption

Assumption 2

Consider the following assumptions on the joint ‘distribution of $Z = (Y_1, X)$ and D, and on the sampling process:

1. *Random sampling:* $\{(Z_i, D_i), i = 1, \cdots, n\}$ *is an independent and identically distributed sequence*
2. *Observed data: The researcher observes* D_i, X_i, *and* $D_i\psi(Z_i, \beta)$ *for all* $\beta \in \mathcal{B}$
3. *Missing at random:* $Y_1 \perp D | X$
4. *Overlap: There exists a* $\kappa > 0$ *such that*

$$p_{j,0}(x) = P(D = d_j | X = x) \geq \kappa$$

for all $j = 1, \cdots, J + 1$ *and for all* $x \in supp(X)$

Model: MAR assumption

- For $J = 1$ and $d_1 = I_p$, Assumptions1+2 correspond to the standard MAR setup
- Missing at random (MAR):
    - data availability is randomly determined within subpopulations determined by X
    - point-identification of β_0 in missing data and program evaluation settings
- Missing completely at random (MCAR)
    - Special case with $X = 1$, $Z \perp D$

Model: identification

Assumption 3

Every component of ψ is observable in at least one stratum, i.e. matrix $\sum_{j=1}^{J} d_j$ has full rank.

- Example: bivariate mean estimation

$$E\left(\left[\begin{array}{c} X_1 \\ X_2 \end{array}\right] - \left[\begin{array}{c} \mu_1 \\ \mu_2 \end{array}\right]\right) = 0 \qquad (5)$$

when only one variable is available for any observation:

$$D \in \left\{ d_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, d_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, d_3 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \right\}$$

- Differentiates the incomplete data model from a multi-valued treatment setup (e.g. Cattaneo, 2010)

Model: summary

- Moment condition setup
- Incomplete data indicator
- Missing at random + overlap
- Every component is observed at least once

Outline

Moment conditions (missing data)

- Graham (2011): missing data ($J = 1$ and $d_1 = I_p$)
- Assumptions 1-3 informationally equivalent to

$$E\left[\frac{1\{D = I_p\}}{p_{1,0}(X)}\psi(Z,\beta_0)\right] = 0 \quad (6)$$

$$E\left[\left.\frac{1\{D = I_p\}}{p_{1,0}(X)} - 1\right| X\right] = 0, \quad (7)$$

where

$$p_{1,0}(x) = P(D = d_1 | X = x)$$

- With incomplete data, (6) and (7) hold for each stratum

Moment conditions: selection probabilities

- For each $j \in \{1, \cdots, J\}$, define the stratum indicator

$$s_j = 1\{D = d_j\}. \tag{8}$$

- Selection probabilities are given by

$$\begin{aligned} p_{j,0}(x) &= P(D = d_j | X = x) \quad (9) \\ &= E[s_j | X = x] \quad (10) \end{aligned}$$

- Analog of moment condition (7) rewrites that definition:

$$E\left[\frac{s_j}{p_{j,0}(X)} - 1 \middle| X \right] = 0 \tag{11}$$

Moment conditions: IPW

- Analog of moment condition (6):

$$E\left[\frac{s_j}{p_{j,0}(X)} d_j \psi(Z, \beta_0)\right] = 0, \tag{12}$$

- Valid under Assumptions 1+2:

$$\begin{aligned}
E\left[\frac{s_j}{p_{j,0}(X)} \psi(Z, \beta_0)\right] &= E_X\left[E\left[\frac{s_j}{p_{j,0}(X)} \psi(Z, \beta_0)\right] \middle| X\right] \\
&= E_X\left[E\left[\frac{s_j}{p_{j,0}(X)} \middle| X\right] E\left[\psi(Z, \beta_0) \middle| X\right]\right] \\
&= E_X\left[E\left[\psi(Z, \beta_0) \middle| X\right]\right] \\
&= E\left[\psi(Z, \beta_0)\right] = 0.
\end{aligned}$$

Stacked moment conditions

To facilitate a GMM analysis, stack moment conditions across j.

$$E\left[\begin{array}{c} \frac{s_1}{p_{1,0}(X)} - 1 \\ \vdots \\ \frac{s_J}{p_{J,0}(X)} - 1 \end{array} \middle| X \right] = 0 \tag{13}$$

$$E\left[\left[\begin{array}{c} \frac{s_1}{p_{1,0}(X)} d_1 \\ \vdots \\ \frac{s_J}{p_{J,0}(X)} d_J \end{array}\right] \psi(Z, \beta_0)\right] = 0 \tag{14}$$

Main result (notation)

- Moment function ψ has conditional expectation $q(X)$, and conditional variance $\Sigma_0(X)$
- Selection probabilities are stacked into $R_0(X)$
- Λ_0 is the variance of stacked weighted moments
- Δ_2 selects the observed components

$$q(X) = E[\psi(Z,\beta_0)|X] \tag{15}$$

$$\Sigma_0(X) = V[\psi(Z,\beta_0)|X] \tag{16}$$

$$R_0(X) = \operatorname{diag}(p_{1,0}(X), \cdots, p_{J,0}(X)) \tag{17}$$

$$\Lambda_0 = E\left[R_0^{-1}(X) \otimes \Sigma_0(X) + \iota_J \iota_J' \otimes q(X)\, q(X)'\right] \tag{18}$$

$$\Delta_2 = \begin{bmatrix} d_1 & \cdots & d_J \end{bmatrix}' \tag{19}$$

Main result

Theorem 4

Assume that (i) the distribution of Z has known, finite support; (ii) the moment conditions (13) and 14 hold; (iii) Λ_0 is invertible and Γ_0 has full rank; (iv) other regularity conditions hold (see e.g. Chamberlain (1989, Section 2). Then the Fisher information for β_0 is given by

$$I(\beta_0) = \Gamma_0' \Delta_2' \Lambda_0^{-1} \Delta_2 \Gamma_0. \tag{20}$$

Main result: extension of missing data

- Missing data bound is well-known:

$$I_m(\beta_0) = \Gamma_0' \Lambda_{m,0}^{-1} \Gamma_0, \tag{21}$$

where

$$\Lambda_{m,0} = E\left[\frac{\Sigma_0(X)}{p_{1,0}(X)} + q(X)\, q(X)'\right], \tag{22}$$

- Obtained as a special case of bound in Theorem 4 with $J = 1$, $d_1 = I$

Main result: MCAR

- Special case:
  - MCAR: $X = 1$
  - Identification in each stratum
  - $d_1 = I_p$
- Information in each stratum is:

$$I_{j,mcar}(\beta_0) = p_{j,0}\Gamma_0' d_j \Sigma_0^{-1} d_j \Gamma_0 \tag{23}$$

- Total information is:

$$I(\beta_0) = \sum_{j=1}^{J} I_{j,mcar}(\beta_0) \tag{24}$$

- Information when discarding incomplete information:

$$I_m(\beta_0) = I_{1,mcar}(\beta_0), \tag{25}$$

- Incomplete observations allow for more efficient estimation:
  - $I(\beta_0) - I_m(\beta_0)$ is positive definite iff $J > 1$.

Efficiency (remarks)

1. Known finite support is not crucial, but used to derive the bound. Results in (Chamberlain, 1989, Section 3) establish that the bound applies to the general case.
2. The efficiency bound is for the derived moment conditions
   - 2.1 Is it the efficiency bound for Assumptions 1-3? Probably yes, no proof yet.
   - 2.2 additional moment conditions may be available, that are redundant with full data (Pacini and Windmeijer, 2015)

Outline

Estimation: selection probabilities

- Assume discrete X
- Estimator for the selection probabilities $p_{j,o}(x_l) = P(D = d_j | X = x_l)$ is

$$\hat{p}_j(x) = \frac{\sum_{i=1}^{n} 1\{D_i = d_j, X_i = x\}}{\sum_{i=1}^{n} 1\{X = x\}}. \tag{26}$$

- Stack all selection probabilities in p_0, with estimator $\hat{p}$

Estimation: parameter of interest

- Use the optimal GMM estimator with $\hat{p}$ plugged in.
- $\bar{m}_n(p, \beta)$ is the sample analog of the available moment conditions:

$$\bar{m}_n(p, \beta) = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} \frac{1\{D_i = d_1\}}{\hat{p}_1(X_i)} \tilde{d}_1 \psi(Z_i, \beta) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} \frac{1\{D_i = d_J\}}{\hat{p}_J(X_i)} \tilde{d}_J \psi(Z_i, \beta) \end{bmatrix}, \qquad (27)$$

  where $\tilde{d}_j$ is the rectangular version of d_j (no zero rows)
- Plug-in GMM estimator $\hat{\beta}$ is

$$\hat{\beta} = \text{argmin}_{\beta \in \mathcal{B}} \bar{m}_n(\hat{p}, \beta)' W_n \bar{m}_n(\hat{p}, \beta) \qquad (28)$$

  for a weight matrix W_n

Estimation: large-sample distribution

Assumption 5

(i) The parameter space $\mathcal{B}$ *is compact, and* β_0 *is in the interior of* $\mathcal{B}$*; (ii) the sequence of matrices* W_n *converges to* $I(\beta_0)$*; (iii) the moment function* ψ *is continuously differentiable on* $\mathcal{B}$.

Theorem 6

Assume that the conditions of Theorem 5 are satisfied, and that Assumptions 1, 2, 3, and 5 hold. Then the limiting distribution of the two-step GMM estimator $\hat{\beta}$ *in (28) is given by*

$$\sqrt{n}\left(\hat{\beta} - \beta_0\right) \xrightarrow{p} \mathcal{N}\left(0, I^{-1}\left(\beta_0\right)\right). \tag{29}$$

Outline

Main result: linear IV

- Linear IV model:

$$E\left[W_1\left(y - X\beta_0\right)\right] = E\left[W_2\left(y - X\beta_0\right)\right] = 0 \tag{30}$$

- Three outcomes for the missing data indicator:

$$D \in \left\{d_1 = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}, d_2 = \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}, d_3 = \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right\}$$

- Efficient GMM estimator sets
  - $\hat{p}$ as above
  - $\hat{\beta}$ from 4 weighted instruments

$$\begin{aligned}
E\left[\frac{s_1}{\hat{p}_1(X)} W_1\left(y - X\beta_0\right)\right] &= 0\\
E\left[\frac{s_1}{\hat{p}_1(X)} W_2\left(y - X\beta_0\right)\right] &= 0\\
E\left[\frac{s_2}{\hat{p}_2(X)} W_1\left(y - X\beta_0\right)\right] &= 0\\
E\left[\frac{s_3}{\hat{p}_3(X)} W_2\left(y - X\beta_0\right)\right] &= 0
\end{aligned}$$

Main result: linear IV

- Set of instruments depends on observed data availability
- If strata are

$$D \in \left\{ d_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, d_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\},$$

  then use instruments:

  - $\frac{s_1}{\hat{p}_1(X)} W_1$
  - $\frac{s_2}{\hat{p}_2(X)} W_2$

Numerical results: setup

- Linear IV, one regressor X, two instruments (W_1, W_2)
- Data availability:
  1. Both instruments.
  2. Only W_1
  3. Only W_2
- Instruments both have a 50% of being missing, and

$$\text{Var}\begin{pmatrix} X \\ W_1 \\ W_2 \end{pmatrix} = \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & \rho \\ 0.5 & \rho & 1 \end{pmatrix}.$$

- Instruments have correlation $\rho = Cov(W_1, W_2)$.

Numerical results

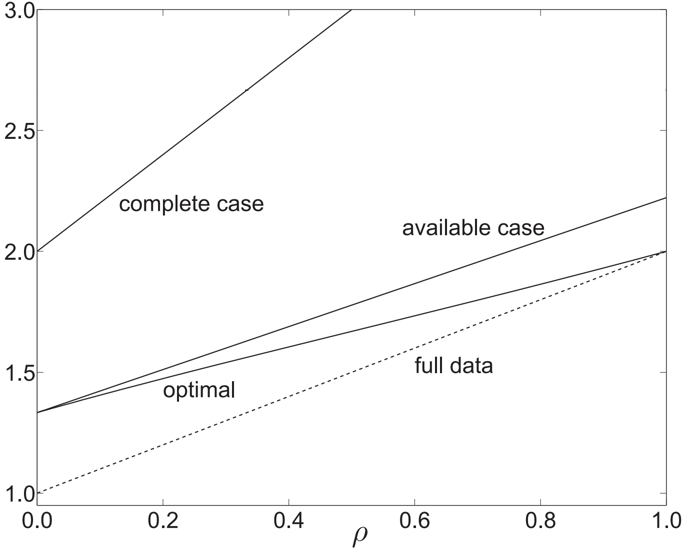

Figure: Asymptotic variance for various estimators of β_0 as a function of $\rho,\ p_1 = 0.5$.

Outline

Conclusion

- Framework for estimation using incomplete data
- Moment conditions and efficiency bound generalize those for missing data
- Efficient estimators are easy to implement
- Identification can be achieved even if it fails in each stratum of incompleteness