Imputation in GMM models with nonparametric missingness structure

Jason Abrevaya and Peter Toth

University of Texas at Austin, University of Nevada, Reno

University of Nevada, Reno

Introduction

Imputation

- We examine the case when the researcher
  - has 1 variable (X) with large number of missing values (2̃0%)
  - fully observed variables ($\mathbf{Z}$)
  - wants to infer the relationship between a LHS variable Y and the RHS variables $\mathbf{Z}$

  1. learn the relationship between X and $\mathbf{Z}$ using the fully observed cases
  2. recover some of the variation in X for the missing observations
  3. using the augmented data to infer the relationship between $\mathbf{Z}$ and Y
- Chamberlain (1982), Abrevaya and Donald (2017), Murris (2019), Coe (2019)
  - works well in linear models

Question

- Abrevaya and Donald (2017) offers a simple GMM solution for the case when we have
    - a linear model for $E[Y|X, \mathbf{Z}]$
    - another linear model connecting X and $\mathbf{Z}$
    - the sole exlusion restriction that the missingness is (mean-)independent of X, conditional on $\mathbf{Z}$
- Our question is: Is it possible to apply the framework for the case of
    - parametric nonlinear model for $E[Y|X, \mathbf{Z}]$,
    - no assumption on the relationship between X and $\mathbf{Z}$, other than X is not independent of $\mathbf{Z}$,
    - arbitrary missingness structure with the same exclusion restriction as in the linear case
- We aim to preserve the (relatively) simple nature of the GMM-framework

Today's results

- We provide a GMM estimator that allows for efficiency gains IF the dimension of the $\mathbf{Z}$ is at most 4 (including the constant)
  - you can have additional dimensions with discrete variables
- We derive the asymptotic properties of the estimator
- Highlight the trade-offs:
  - you do not want to do an imputation scheme if the estimates for the missing elements you use are very noisy
  - you have to do complicated schemes (including all $\mathbf{Z}$-s) if you do not allow for strict exclusion restrictions

The model

$$E[Y|X,\mathbf{Z},M] = E[Y|X,\mathbf{Z}] = h(\alpha X + \beta \mathbf{Z})$$
$$f_{X|\mathbf{Z},M}(x,z,m) = f_{X|\mathbf{Z}}(x,z)$$

- We know h, but the conditional distribution $f_{X|\mathbf{Z}}$ is unknown
  - h is smooth and well-behaved for identification (i.e. strictly increasing)
- M is the missingness indicator, taking the value 1 when the observation is missing (otherwise 0)
- We observe

$$M, \mathbf{Z}, M \cdot X, Y$$

Weakened exclusion restriction

- Directly from AD (2017)
- Weaker than the missingness-at-random assumption (standard), as it is allowed for M to depend on $\mathbf{Z}$
- Roughly translates to $M \perp X$
- Our two assumptions imply that

$$E(Y|\mathbf{Z}, M) = \int h(\alpha X + \beta \mathbf{Z}) f_{X|\mathbf{Z}}(x, \mathbf{Z}) dx$$

Estimator and asymptotic theory

Population moments

- Let us have $\mathbb{Z} \in \mathbb{R}^k$ (further, I will not emphasize that $\mathbf{Z}$ is a vector)

$$E[g(\alpha, \beta; E[y|z])] = E \begin{bmatrix} (1-m)x(y - h(\alpha x + \beta z)) \\ (1-m)z(y - h(\alpha x + \beta z)) \\ mz(y - E[y|z]) \end{bmatrix} = 0,$$

- Here g is a function whose co-domain is $\mathbb{R}^{1+k+k}$
  - the first $k+1$ moments are the basis of the usual GMM estimator (assumed to be well-behaved)
  - the argument $E[y|z]$ is the function of z itself (technically a parameter) with the true value

$$E[y|z] = \int h(\alpha x + \beta z) f_{X|Z}(x, z) dx$$

The imputation estimator (GMM)

- We take the sample analogue of the population moments

$$\hat{g}(a,b;\hat{E}[y|z]) = n^{-1}\sum_{i=1}^{n}\begin{bmatrix}(1-m_i)x_i(y_i-h(ax_i+bz_i))\\(1-m_i)z_i(y_i-h(ax_i+bz_i))\\m_i z_i(y_i-\hat{E}[y_i|z_i])\end{bmatrix}$$

$$\hat{E}[y_i|z_i] = \int h(ax+bz_i)\hat{f}_{x|z}(x,z_i)dx$$

- $\hat{f}_{x|z}$ is a linear estimator of the conditional pdf $f_{x|z}$ (Nadaraya-Watson for us)

$$[\hat{\alpha},\hat{\beta}] = \operatorname*{argmin}_{a,b}\hat{g}(a,b;\hat{E}[y|z])'\hat{W}\hat{g}(a,b;\hat{E}[y|z])$$

Weighting Matrix

- The optimization of the weighting matrix seems to be important to achieve good results for imputation
- The optimal weighting matrix that minimizes the MSE (not the variance!) is the usual

$$W = \left(E[g(\alpha, \beta; E[y|z])' g(\alpha, \beta; E[y|z])]\right)^{-1}$$

- We always going to take the sample analogue of this population moment for our calculations ($\hat{W}$)

Results

- Assume that the estimator $\hat{E}[y|z]$ converges uniformly and the bias is $o_p(\sqrt{n}^{-1})$
- Under usual regulatory assumptions, given that $\hat{W}$ is the sample analogue of W,
  1. The imputation estimator is root-n consistent
  2. The asymptotic variance-covariance matrix is $(G'WG)^{-1} + o_p(n^{-1/2})$
  3. Asymptotically, $MSE = (G'WG)^{-1}$
  4. (Asymptotic normality holds - not proven yet)

  - where G is the Jacobian matrix of g w.r.t. the finite dim'l parameters at the true values

Take-away

- **Given that the optimal weighting matrix puts non-zero weights on the imputation moments**, the MSE of the imputation estimator is strictly smaller than that of the optimally weighted GMM estimator that discards the observations with missing values
- When there is no convergent nonparametric estimator for which the bias vanishes fast enough, calculating these additional imputation moments gives more noise to the GMM estimator than they are worth
- In the Nadaraya-Watson case, we need that the rate of the bandwidth

$$-\frac{1}{k-1} < h < -1/4,$$

  - where $k-1$ is the number of non-constant Z elements

Monte Carlo simulations

The data generating process

$$E[Y|X = x, Z = z] = \Phi(\alpha x + \beta z)$$

- The X is a nonlinear function of Z and some exogenous randomness
- The missingness is based on another probit model and truncation

$$M = \mathbb{1}[|\gamma z > \epsilon_i| < 0.8],\ \epsilon_i \sim N[0, 1]$$

  - this gives missingness rates around 55%
- We implemented optimal weighting with k=2 ($h = -1/3$)
- Three estimators: 1. Full-data set GMM (infeasible) 2. Completely-observed GMM 3. Imputation GMM

Monte Carlo Results (true coefficients are $[1, 0.5, -2]$)

$n = 4000$	Full-data	Completely-observed	Imputation
α	0.995 (0.051)	0.989 (0.076)	0.990 (0.076)
β_0	0.497 (0.047)	0.516 (0.068)	0.510 (0.054)
β_1	-2.007 (0.069)	-2.009 (0.099)	-2.012 (0.078)
$n = 16000$	Full-data	Completely-observed	Imputation
α	0.996 (0.026)	0.998 (0.038)	0.998 (0.038)
β_0	0.499 (0.025)	0.497 (0.036)	0.501 (0.028)
β_1	-1.998 (0.039)	-1.998 (0.055)	-2.003 (0.04)

Further questions

1 Introduction

2 Estimator and asymptotic theory

3 Monte Carlo simulations

4 Further questions

More exclusion restrictions, marginalized estimators (speculation)

- "Simple" imputation is not going to yield better results when the dimension of the Z vector is higher than 4
- There are two ways to remedy this
  1. getting closer to missing-at random assumptions by adding exclusion restrictions like

$$M \perp (X, Z_i) | \mathbf{Z_{-i}}$$

  2. MAYBE we can devise clever reweighting-schemes to increase the number of moments but decrease the number of variables we condition on in $E[y|z]$ (but the weighting scheme may be just as noisy to calculate, it urns out)

Thank you for your attention