Lecture 13

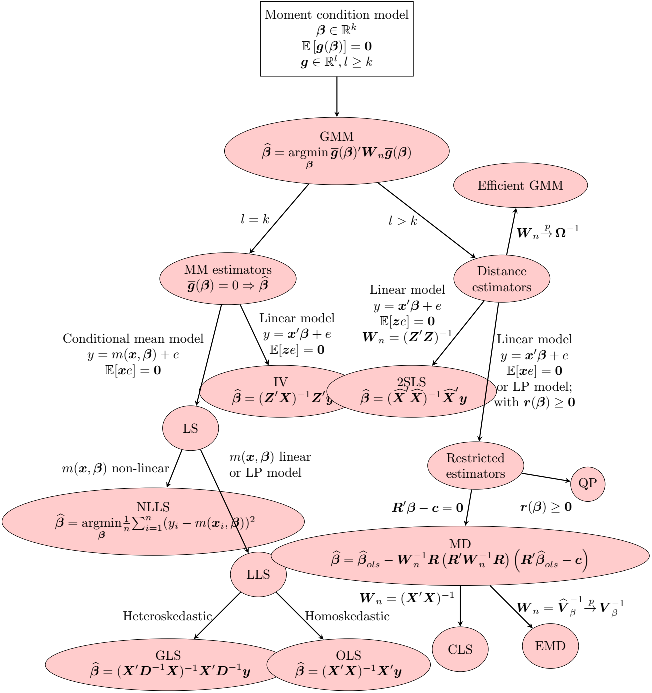
Concluding remarks

Lectures in SDPE: *Econometrics I* on March 11, 2024

Markus Jäntti
Swedish Institute For Social Research
Stockholm University

Purpose of these concluding remarks

- Summarize in a “mind map” how the “moment condition model” relates to different special cases we have examined.
- Summarize in broad terms what we should have learned.
- Point out important areas which we have not covered.
- Discuss good empirical research practicalities.



Moment condition model
$\boldsymbol{\beta} \in \mathbb{R}^k$
$\mathbb{E}\left[\boldsymbol{g}(\boldsymbol{\beta})\right] = \mathbf{0}$
$\boldsymbol{g} \in \mathbb{R}^l, l \geq k$
GMM
$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \overline{\boldsymbol{g}}(\boldsymbol{\beta})'\boldsymbol{W}_n\overline{\boldsymbol{g}}(\boldsymbol{\beta})$
Efficient GMM
$l = k$
$l > k$
$\boldsymbol{W}_n \xrightarrow{p} \boldsymbol{\Omega}^{-1}$
MM estimators
$\overline{\boldsymbol{g}}(\boldsymbol{\beta}) = 0 \Rightarrow \widehat{\boldsymbol{\beta}}$
Distance estimators
Linear model
$y = \boldsymbol{x}'\boldsymbol{\beta} + e$
$\mathbb{E}[\boldsymbol{z}e] = \mathbf{0}$
$\boldsymbol{W}_n = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}$
Linear model
$y = \boldsymbol{x}'\boldsymbol{\beta} + e$
$\mathbb{E}[\boldsymbol{z}e] = \mathbf{0}$
Conditional mean model
$y = m(\boldsymbol{x}, \boldsymbol{\beta}) + e$
$\mathbb{E}[\boldsymbol{x}e] = \mathbf{0}$
Linear model
$y = \boldsymbol{x}'\boldsymbol{\beta} + e$
$\mathbb{E}[\boldsymbol{x}e] = \mathbf{0}$
or LP model;
with $\boldsymbol{r}(\boldsymbol{\beta}) \geq \mathbf{0}$
IV
$\widehat{\boldsymbol{\beta}} = (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\boldsymbol{y}$
2SLS
$\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}})^{-1}\widehat{\boldsymbol{X}}'\boldsymbol{y}$
LS
$m(\boldsymbol{x}, \boldsymbol{\beta})$ non-linear
$m(\boldsymbol{x}, \boldsymbol{\beta})$ linear
or LP model
Restricted estimators
QP
NLLS
$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n}(y_i - m(\boldsymbol{x}_i, \boldsymbol{\beta}))^2$
$\boldsymbol{R}'\boldsymbol{\beta} - \boldsymbol{c} = \mathbf{0}$
$\boldsymbol{r}(\boldsymbol{\beta}) \geq \mathbf{0}$
MD
$\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}_{ols} - \boldsymbol{W}_n^{-1}\boldsymbol{R}\left(\boldsymbol{R}'\boldsymbol{W}_n^{-1}\boldsymbol{R}\right)\left(\boldsymbol{R}'\widehat{\boldsymbol{\beta}}_{ols} - \boldsymbol{c}\right)$
LLS
$\boldsymbol{W}_n = (\boldsymbol{X}'\boldsymbol{X})^{-1}$
$\boldsymbol{W}_n = \widehat{\boldsymbol{V}}_{\beta}^{-1} \xrightarrow{p} \boldsymbol{V}_{\beta}^{-1}$
Heteroskedastic
Homoskedastic
CLS
EMD
GLS
$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{D}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{D}^{-1}\boldsymbol{y}$
OLS
$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$

The moment condition model-family

- minimization of the quadratic form yields the estimator
  - if the number of parameters k equals the number of moment conditions l, and estimator is linear, the minimum can be solved directly
  - if not, it is solved by least squares or by an iterative algorithm
- estimator properties can be studied by
  - substituting the regression equation into the estimator, ...
  - ... and studying/using the limiting behaviour of functions of sampled random variables and regression errors

The moment condition model-family

- we rely on asymptotic properties of our estimators
  - we assume (in general) the data have finite fourth moments. . .
  - . . . and that key empirical moment matrices converge in probability to their population expectations (positive definite!)
  - the estimators' asymptotic properties are derived by applying the weak law of large numbers, the central limit theorem, and the continuous mapping theorem
  - the asymptotic distribution of all the estimators in the family is the multivariate normal
  - the estimators are in general consistent and have an asymptotic finite, positive definite variance matrix
  - unbiasedness is not a general property of these estimators and is often hard to show
- from asymptotic normality, it follows that the variance-standardized quadratic form converges to the χ^2-distribution
  - which gives the limiting distribution of multivariate test statistics

The regression error

- we have, for the most part, assumed the regression error has a zero covariance with the regressors, $\mathrm{E}[\boldsymbol{X}e] = 0$
- on occasion, we have made the stronger conditional mean (aka mean independence) assumption that $\mathrm{E}[e|\boldsymbol{X}] = 0$. This implies, but is not implied by, $\mathrm{E}[\boldsymbol{X}e] = 0$
- $\mathrm{E}[e|\boldsymbol{X}] = 0$ allows us to think in terms of conditional expectations so we have
  - unbiasedness of $\widehat{\beta}$
  - the conditional variance of $\widehat{\beta}$
  - the conditional variance of e, $\sigma^2(\boldsymbol{X})$

  But outside of normal regression, this assumption does not give us the *distribution* of $\widehat{\beta}$.
- we rely on $\mathrm{E}[\boldsymbol{X}e] = 0$ for asymptotic properties of $\widehat{\beta}$:
  - consistency of $\widehat{\beta}$
  - asymptotic normality of $\widehat{\beta}$

  Note that we allow ourselves to think implicitly or explicitly of the conditional variance of e, $\sigma^2(\boldsymbol{X})$ even in this case, although it mostly enters only through the asymptotic distribution of $\widehat{\beta}$.

On the down side...

- we have only covered cases of continuous dependent variables – in applied work, limited dependent variables of various types are important

  However, in many cases, these can be studied using **generalized linear models**, a straightforward extension of the linear regression model into a large set of limited dependent variables.
  When not, maximum likelihood in various guises is an option. As (by an extension of sorts) are Bayesian methods.

On the down side...

- we have relaxed the “identical” part of the iid assumption (mostly assuming heteroscedasticity, treating homoscedasticity as a rare special case), and to some extent the “independence” assumption by looking at clustered data (and taken a brief excursion into time series)

  Non-independence is in practice very important, and can be dealt with by generalizing the heteroscedasticity-consistent variance estimators. Or modeled explicitly, as in spatial econometrics, panel data, clustered data and so on.

On the down side...

- Three important approach to estimation have been (mostly) overlooked, namely Bayesian, simulation-based, and non-parametric methods.

  All three are important in applied and theoretical econometrics. Simulation and Bayesian methods are important approaches to the estimation of otherwise intractable problems. Non-parametric methods allow a flexible examination of how the dependent variable relates to the explanatory variables

Research practicalities

- keep a research log
- make frequent (eg daily), preferably automated backups
- comment your code (so that someone else than you can understand what you do and why; that includes *you* X months in the future!)
- strive for reproducible research
    - consider “weaving” methods
    - at the very least, keep a close connection between your research report and your coding; all tables and figures should be produced from input data with minimal intervention from you
- use version control software such as git and be sure to comment also the versions properly (real version control, not “accidental/incidental” ones)
- computers are useful for research and not elaborate typing machines:
    - weaving
    - version control
    - bibliographic databases