Specification Uncertainty in the Instrumental Variable Regression Model: A Bayesian Model Averaging Approach

Gary Koop (University of Strathclyde, Scotland)

Roberto León-González (The National Graduate Institute for Policy Studies (GRIPS), Japan)

Rodney Strachan (University of Queensland, Australia)

Rimini, July 2009

All authors are fellows of The Rimini Centre for Economic Analysis

INTRODUCTION

Uncertainty on the validity of identification restrictions in the incomplete simultaneous equations model:

- Set of instruments and the rank condition for identification.
- Which regressors shall we include, and which of those are endogenous?

Researchers typically try different specifications until a set of restrictions passes a battery of mispecification tests.

Given the large number of possible models, the repeated application of diagnostic tests will result in distorted size and power properties.

However, estimates of structural estimates that rely on incorrect identification restrictions can result in large biases.

OBJECTIVES

- Develop methods for checking the validity of identifying restrictions.
  - This can be done by calculating the posterior probability of these restrictions.
  - Alternatively, we can use the BMA posterior distribution of discrepancy vectors and functions (Zellner, Bauwens and van Dijk, 1988) in order to shed light on the validity of instruments.

- Provide a method for inference about the structural parameters that, conditional on identification holding, accounts for model uncertainty when the number of models is large.
  - Build on the literature on Reversible Jump Markov Chain Monte Carlo algorithm (RJMCMC) similar to Holmes, Denison and Mallick (2002).
  - Borrow on simulated tempering ideas to avoid the algorithm getting stuck.
  - Algorithm allows use of any of the priors proposed in the Bayesian IV literature (e.g. Dreze (1976), Kleibergen and van Dijk (1998), Strachan and Inder, (2004))

THE INCOMPLETE SIMULTANEOUS EQUATIONS MODEL

We are interested in the structural model:

$$\begin{aligned} y_{1i} &= \gamma' y_{2i} + \beta' x_i + u_{1i} \\ y_{2i} &= \Pi_{2x} x_i + \Pi_{2z} z_i + v_{2i} \end{aligned}$$

With reduced form:

$$y_i = \Pi_x x_i + \Pi_z z_i + v_i$$

MODELLING CHOICES

The range of models we consider is:

- Over-identified models: These are the models in which $k_{2j} > m$ and Π_{2z} has full rank. Models in this category differ according to the following aspects:
  - Set of instruments: The variables in z_i are a subset of a larger group of potential instruments that we denote by Z. There is uncertainty as to which subset of Z should enter in the model.
  - Variables in x_i: x_i is a subset of $Z \cup X$, where X is the set of all potential regressors that are not allowed to be instruments.
  - Restrictions on the coefficients of endogenous regressors: some coefficients in γ might be restricted to be zero.

– Exogeneity: some of the covariances between u_{1i} and v_{2i} might be zero.

- Non-identified models: There are two categories within this group:

  – Full rank models: models in which Π_z has full rank. In this group x_i and z_i play the same role. Models differ only on the set of variables (x_i, z_i) that enter in the model. No restrictions are placed on the variance-covariance matrix of v_i.

  – Reduced rank models: models in which Π_{2z} has rank $(m-1)$. No restrictions are placed on the variance-covariance matrix of v_i. Models differ according to:

    * Variables in z_i. As before, $z_i \subseteq Z$.

    * Variables in x_i. As before $x_i \subseteq Z \cup X$.

STARTING POINT FOR ALGORITHM DESIGN

Building on RJMCMC, **H**olmes, **D**enison and **M**allick (2002, **HDM** hereafter) develop an algorithm for BMA in the SURE model.

If Dreze's prior is used, HDM can also be used for the INSEM.

However, the special nature of the IV regression implies this algorithm can perform poorly.

Sample from the posterior of $(\boldsymbol{\Pi}, \boldsymbol{\Sigma}, M)$:

1. Draw $M|Y, \boldsymbol{\Sigma}$

2. Draw $\boldsymbol{\Pi}|Y, \boldsymbol{\Sigma}, M$

3. Draw $\boldsymbol{\Sigma}|Y, \boldsymbol{\Sigma}, \boldsymbol{\Pi}, M$

Step 1 draws a candidate M^* and accepts it with probability:

$$\min\left\{\frac{\pi(Y,\Sigma|M^*)}{\pi(Y,\Sigma|M)}\frac{\pi(M^*)}{\pi(M)},1\right\}$$

where:

$$\pi(Y,\Sigma|M)=\int\pi(\Pi,\Sigma|M)\pi(Y|\Pi,\Sigma,M)d\Pi$$

HDM AS A RJMCMC

Simple exposition in Clyde et al.:

Sample from $(\Pi, M)|\Sigma, Y$:

Current state: (Π_k, M_k) Candidate State: (Π^*, M^*)

Proposal density for (Π^*, M^*) : $q(M^*|M_k)q(\Pi^*|M^*)$

Accept with probability $\alpha = \min\{1, H\}$ where H is:

$$H=\frac{\pi(Y,\Sigma|\Pi^*,M^*)}{\pi(Y,\Sigma|\Pi_k,M_k)}\frac{\pi(\Pi^*|M^*)}{\pi(\Pi_k|M_k)}\frac{\pi(M^*)}{\pi(M_k)}\frac{q(M_k|M^*)}{q(M^*|M_k)}\frac{q(\Pi_k|M_k)}{q(\Pi^*|M^*)}$$

Choosing $q(\Pi^*|M^*)$ as the posterior conditional density of Π^* given Σ: Most efficient choice and ratio simplifies to the one in previous slide.

NUMERICAL EXAMPLE I

```
omegap=eye(3);
omegap[2,1]=0;
omegap[1,2]=0;
omegap[1,3]=0.8;
omegap[3,1]=0.8;
X=rndn(250,2);
Z=rndn(250,7);                              /* Potential Instruments */
epsilon=rndn(enep,mp)*chol(omegap);          /* Error terms */
gama={1, -1};
PIX2={1,2};
PIZ2={-1.3,10.5,1,1,1,10,5};
PIX3={3,-2};
PIZ3={-1.3,1.5,1,10,3,2,1};

beta={1,1,11,-10,0,0,0,12,13};

y2=Z*PIZ2+X*PIX2+epsilon[.,2];
y3=Z*PIZ3+X*PIX3+epsilon[.,3];
y1=(y2~y3)*gama+(X~Z)*beta+epsilon[.,1];
```

WHY THE ALGORITHM COULD GET STUCK AT A WRONG MODEL

Right model: $M_R = \{1, 1, 0, 0, 0, 1, 1\}$

Wrong model: $M_W = \{1, 1, 0, 0, 0, 1, 1\}$

ML estimate of Σ in M_R: $\widehat{\Sigma}_R$

ML estimate of Σ in M_W: $\widehat{\Sigma}_W$

$\pi(Y, \widehat{\Sigma}_W | M_R)$ =-7047.30

$\pi(Y, \widehat{\Sigma}_W | M_W)$ =-559.92

$\pi(Y, \widehat{\Sigma}_R | M_W)$ =-14436.22

$\pi(Y, \widehat{\Sigma}_R | M_R)$ =-529.52

NUMERICAL EXAMPLE II

Returns to Education: Card (1995), 14 potential instruments, 5 endogenous variables, 26 exogenous regressors, $N = 2040$.

$M_R = \{1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1\}$

$M_W = \{1, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 1\}$

$\pi(Y, \widehat{\Sigma}_W | M_R)$ =-6262.26

$\pi(Y, \widehat{\Sigma}_W | M_W)$ =-6181.10

$\pi(Y, \widehat{\Sigma}_R | M_W)$ =-6176.05

$\pi(Y, \widehat{\Sigma}_R | M_R)$ =-6144.32

ADAPT IDEAS FROM SIMULATED TEMPERING

Extend the model space with cold models.

Cold models are simplified versions of hot models: we have as many cold models as hot ones.

Marginal likelihood $\pi(Y|M)$ can be computed analytically in cold models.

Use approximation to the likelihood (Zellner, p. 274) for cold models.

We then construct an algorithm that samples over the model space containing both hot and cold models.

Transitions between cold models are very efficient, even if they have different instruments, rank or identifying restrictions. Efficiency like in MC^3 (e.g. Madigan and York, 1995 or Fernandez, Ley and Steel, 2001).

Transitions between cold and hot models are also efficient, because they share a similar posterior distribution.

In the numerical examples above this ST algorithm does not get stuck in models of low posterior probability.

SOME TECHNICAL DETAILS

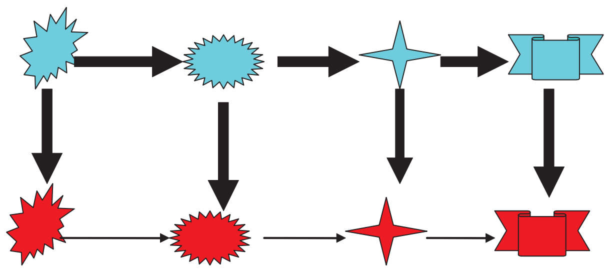
Transitions between cold models are marginal on both Π and Σ.

Transitions between cold models and hot ones are conditional on Π (but not on Σ)

Transitions between hot models with different set of instruments are marginal on Π but conditional on Σ.(like in HDM)

Transitions between hot models with different exogeneity restrictions are marginal on Σ but conditional on Π.

A GRAPHICAL ILLUSTRATION



OTHER PRIORS

Suppose we want to use another prior π^{SH} for (Π, Σ) different from Dreze (1976).

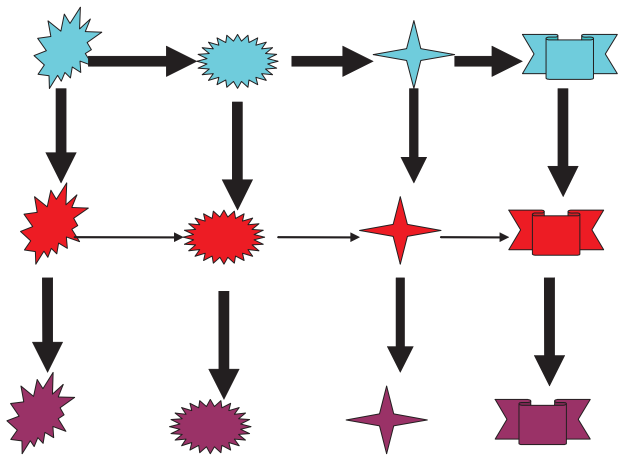
Extend the model space with super-hot models: those that use prior π^{SH}

- Cold models: use approximation to likelihood
- Hot models: use Dreze (1976) prior
- Super-hot models: use π^{SH} as prior

Transitions between hot and super-hot models are conditional on both Π and Σ

No need to make direct transitions between super-hot models subject to different set of restrictions.

A GRAPHICAL ILLUSTRATION



DIAGNOSTICS

Probability of over-identification: Relative number of visits to over-identified models

Diagnostics for over-identification: BMA posterior of discrepancy measures (Zellner, Bauwens and van Dijk, 1988). Some of these measures are bounded between 0 and 1.

Diagnostics for partial-identification: BMA posterior of a function of smallest singular value of Π_{2z}. This function can be constructed to be bounded between 0 and 1.

SUMMARY

Develop algorithm for BMA in the Instrumental Variable regression model when there is a large amount of competing models.

Show with simulated and real data that it is important to incorporate ideas from simulated tempering so that algorithm does not get stuck

It can be adapted for the use of most priors proposed in the IV literature

Diagnostics for over and partial identification: posterior probability and BMA posterior of discrepancy measures.