Journal of Econometrics 171 (2012) 237–250

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Bayesian model averaging in the instrumental variable regression model☆

Gary Koop [a], Roberto Leon-Gonzalez [b,*], Rodney Strachan [c]

[a] University of Strathclyde, United Kingdom
[b] National Graduate Institute for Policy Studies, Japan
[c] The Australian National University, Australia

ARTICLE INFO

Article history:
Available online 7 July 2012

JEL classification:
C11
C30

Keywords:
Bayesian
Endogeneity
Simultaneous equations
Reversible jump Markov chain Monte Carlo

ABSTRACT

This paper considers the instrumental variable regression model when there is uncertainty about the set of instruments, exogeneity restrictions, the validity of identifying restrictions and the set of exogenous regressors. This uncertainty can result in a huge number of models. To avoid statistical problems associated with standard model selection procedures, we develop a reversible jump Markov chain Monte Carlo algorithm that allows us to do Bayesian model averaging. The algorithm is very flexible and can be easily adapted to analyze any of the different priors that have been proposed in the Bayesian instrumental variables literature. We show how to calculate the probability of any relevant restriction such as exogeneity or over-identification. We illustrate our methods in a returns-to-schooling application.

1. Introduction

For the regression model where all potential regressors are exogenous, a large literature[1] has arisen to address the problems caused by a huge model space. That is, the number of models under consideration is typically 2^K, where K is the number of potential regressors. With such a huge model space, there are many problems with conventional model selection procedures (e.g. sequential hypothesis testing procedures run into pre-test problems). Bayesian model averaging (BMA) can be used to avoid some of these problems. However, the size of the model space means that carrying out BMA by estimating every model is typically computationally infeasible. Accordingly, an algorithm which simulates from the model space (e.g. the Markov chain Monte Carlo model composition algorithm of Madigan and York, 1995) must be used. In the case of the regression model with exogenous regressors, such methods are well-developed, well-understood and are increasingly making their way into empirical work. However, to our knowledge, there are no comparable papers for the empirically important case where regressors are potentially endogenous and, thus, instrumental variable (IV) methods are required.[2] The purpose of the present paper is to fill this gap.

Inference about structural parameters in the IV regression model requires the formulation of assumptions whose validity is often uncertain. A useful representation of the model is the incomplete simultaneous equations model (see, for example, Hausman, 1983). Within this representation, the most crucial assumptions relate to the set of instruments and the rank condition for identification (Greene, 2003, p. 392). In addition to these, one has to decide how many regressors to include, and which of these are potentially endogenous. This can lead to a huge model space and, thus, similar issues arise as for the regression model with exogenous regressors. In practice, researchers typically try different specifications until a set of restrictions (i.e. a particular choice of instruments, exogenous and endogenous regressors) passes a battery of misspecification tests (e.g. Anderson and Rubin, 1949, 1950; Hausman, 1983; Sargan, 1958). Given the large number of possible models, the repeated application of diagnostic tests will result in similar distorted size

☆ All authors are fellows of the Rimini Centre for Economic Analysis and Strachan is also a Fellow of The Centre for Applied Macroeconomic Analysis. The authors would like to thank Frank Kleibergen and other participants at the European Seminar on Bayesian Econometrics for helpful comments as well as the Leverhulme Trust for financial support under Grant F/00 273/J. Leon-Gonzalez also thanks the Japan Society for the Promotion of Science for financial support (start-up Grant #20830025).

* Corresponding author.
E-mail address: rlg@grips.ac.jp (R. Leon-Gonzalez).

[1] See, among many others, Fernandez et al. (2001) and the references cited therein.

[2] Two related papers are Cohen-Cole et al. (2009) and Lenkoski et al. (forthcoming) but the model space in these papers is small and, hence, simulation methods from the model space are not required. Furthermore, the approach of these papers (averaging of two-staged least squares estimates using BIC-based weights) does not have a formal Bayesian justification. Tobias and Li (2004) does BMA in a returns to schooling example similar to the one we use, but this paper does not address endogeneity concerns.

0304-4076/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2012.06.005

properties as arise in the regression model with exogenous regressors. Since estimates of structural parameters that rely on incorrect identification restrictions can result in large biases, the consequences of these problems can be substantive. BMA can be used to mitigate such problems. But the size of the model space often precludes estimation of all models. This leads to a need for computational methods which simulate from the model space. A contribution of the present paper is to design a reversible jump Markov chain Monte Carlo algorithm (RJMCMC, see Green, 1995 or Waagepetersen and Sorensen, 2001) that explores the joint posterior distribution of parameters and models and thus allows us to do BMA. This allows us to carry out inference on the structural parameters that, conditional on identification holding, accounts for model uncertainty. Furthermore, our algorithm allows for immediate calculation of the posterior probability associated with any restriction, model or set of models. Thus, we can easily check the validity of identifying restrictions (or exogeneity restrictions, etc.) by calculating the posterior probability of these restrictions.

In our applications, we find that standard versions of RJMCMC algorithms (e.g. adapting the RJMCMC methods for seemingly related regression, SUR, models developed by Holmes et al., 2002, to the IV case) can perform poorly, remaining stuck for long periods in models with low posterior probability. To improve the performance of our RJMCMC algorithms, we borrow an idea from the simulated tempering literature and augment our model space with so-called cold models.[3] The cold models are similar to the models of interest (called hot models) but are simplified in such a way that the RJMCMC algorithm makes very rapid transitions between cold models. As suggested by the simulated tempering literature, we find that this strategy helps the algorithm escape from local modes in the posterior.

The RJMCMC algorithm we develop is very flexible and can be easily adapted to handle any of the popular approaches to Bayesian inference in IV models such as Drèze (1976), Kleibergen and Van Dijk (1998) and Strachan and Inder (2004). We describe in detail how the algorithm works in the context of two popular Bayesian approaches to instrumental variables and reduced rank regression. These are the classic approach of Drèze (1976) and the modern approach of Strachan and Inder (2004).[4] We also show how, if desired, the RJMCMC algorithm can be easily coded to produce results for several different priors by running the algorithm just once.

Section 2 describes the model space we consider. Section 3 describes the algorithm with complete details being included in a Technical Appendix. Section 4 applies our methods to a returns-to-schooling example based on Card (1995) and Section 5 concludes.

2. Modelling choices in the incomplete simultaneous equations model

We will work with the incomplete simultaneous equations model, which takes the form:

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

where $y_{1i} : 1 \times 1, y_{2i} : m \times 1, x_i : k_{1j} \times 1, z_i : k_{2j} \times 1, i = 1, \ldots, N$. The errors are normal with zero means and are uncorrelated over i. We assume

$$E\left(x_i \begin{pmatrix} u_{1i} \\ v_{2i} \end{pmatrix}'\right) = 0 \quad \text{and} \quad E\left(z_i \begin{pmatrix} u_{1i} \\ v_{2i} \end{pmatrix}'\right) = 0.$$

The reduced form version of this model can be written as:

$$y_i = \Pi_x x_i + \Pi_z z_i + v_i \tag{2}$$

where $y_i = (y_{1i}, y_{2i}')'$, $v_i = (v_{1i}, v_{2i}')'$ and:

$$\Pi_x = \begin{pmatrix} \pi_{1x} \\ \Pi_{2x} \end{pmatrix} = \begin{pmatrix} \gamma' \Pi_{2x} + \beta' \\ \Pi_{2x} \end{pmatrix}$$

$$\Pi_z = \begin{pmatrix} \pi_{1z} \\ \Pi_{2z} \end{pmatrix} = \begin{pmatrix} \gamma' \\ I_m \end{pmatrix} \Pi_{2z}$$

$$\Sigma = E\left(\begin{pmatrix} u_{1i} \\ v_{2i} \end{pmatrix} \begin{pmatrix} u_{1i} & v_{2i}' \end{pmatrix}\right) = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \Sigma_{22} \end{bmatrix}$$

$$\Omega = E(v_i v_i') = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \Omega_{22} \end{bmatrix} = \begin{pmatrix} 1 & \gamma' \\ 0 & I_m \end{pmatrix} \Sigma \begin{pmatrix} 1 & 0 \\ \gamma & I_m \end{pmatrix}$$

$$\Pi_x : (m+1) \times k_{1j} \qquad \Pi_z : (m+1) \times k_{2j}.$$

The subindex j stands for the jth model, and j varies from 1 to N^{mod}, where N^{mod} is the total number of models. To avoid notational clutter, we will not attach j subindices to parameter matrices although, of course, these will vary over models.

When using this model, there are many sources of uncertainty over identification that arise. Assuming $\sigma_{12} \neq 0$, we can solve for the parameters (β', γ') from the reduced form matrix

$$\widetilde{\Pi} = [\Pi_x \quad \Pi_z]$$

through the relations

$$\pi_{1x} - \gamma' \Pi_{2x} = \beta' \quad \text{and} \tag{3}$$
$$\pi_{1z} - \gamma' \Pi_{2z} = 0. \tag{4}$$

If we are able to solve (4) for γ, we can subsequently solve for β using (3). Solving for γ depends upon the rank of the matrix Π_z. If $k_{2j} = m$ and rank $(\Pi_z) = m$ then there is a unique solution $\gamma' = \pi_{1z} \Pi_{2z}^{-1}$ and the equation is just identified. If $k_{2j} > m$ and rank $(\Pi_z) = m$ then there are more equations than we need to identify γ and so the equation is over-identified. If $k_{2j} < m$ so rank $(\Pi_z) < m$, there are fewer equations than necessary to identify γ and thus the equation is under-identified.

Uncertainty over identification can also result from uncertainty over what variables in y_{2i} are endogenous and what variables in z_i are not valid instruments. If we relax the earlier assumption on σ_{12} to allow for $\sigma_{12} = 0$, which implies y_{2i} is exogenous, then we have additional solutions for γ from $\gamma' = \omega_{12} \Sigma_{22}^{-1}$ and the condition $\sigma_{12} = 0$ needs to be taken into account when determining whether (β', γ') is just or over-identified. A further complication arises if elements of γ or σ_{12} are zero, as these restrictions imply elements of y_{2i} are exogenous. This effectively changes the value of m, increasing the number of identifying restrictions in (4) and, hence, the conditions for under, just and over identification. Note also that, if $k_{2j} > m$ and $(k_{2j} - m_j)$ columns of the $m \times k_{2j}$ matrix Π_{2z} are zero, or, if rank $(\Pi_{2z}) = m_j < m$, then not all elements of z_i may be regarded as valid instruments. In this case, we can then represent Π_{2z} as the product of two lower dimensional matrices, $\Pi_{2z} = \underline{\Pi}_{2z} \varrho$, where $\underline{\Pi}_{2z}$ is $m \times m_j$ and ϱ is $m_j \times k_{2j}$ both full rank. The valid instruments are then ϱz_i.

Furthermore, if elements of β are zero, then this gives us more equations of the type (4) and fewer equations of the type (3), again affecting the identification status of (β', γ').

In this paper, we consider a model space which includes all the over-identified and just-identified models (see below for a discussion of non-identified models). These are the models in which $k_{2j} \geq m$ and Π_{2z} has full rank. Models in this category differ according to the following aspects:

[3] To avoid confusion, note that some of the related literature uses different terminology where the space of distributions to be simulated from is augmented with hot distributions, while the actual target distribution is the distribution with the lowest temperature. See for example Kou et al. (2006) and Hoogerheide et al. (2011).

[4] We use a proper prior version of the improper prior used by Drèze (1976), as in the subsequent papers of Drèze and Richard (1983) and Zellner et al. (1988). With respect to the prior by Strachan and Inder (2004), we will use a parameter-augmented version of it similar to that used by Koop et al. (2010). The working paper version of this paper, available on Gary Koop's website, provides full details of how Kleibergen and Van Dijk (1998)'s prior can be used with our algorithm.