Bayesian IV: the normal case with multiple endogenous variables*

Timothy Cogley and Richard Startz

revised January 2015

Abstract

We set out a Gibbs sampler for the linear instrumental-variable model with normal errors and normal priors, and we show how to compute the marginal likelihood.

1 Introduction

In this note we present a Gibbs sampler for a Bayesian instrumental-variable estimator with normal errors and priors, together with an algorithm for computing the marginal likelihood. While the seemingly unrelated regression formulation of instrumental variables is nonlinear in its parameters, Gibbs sampling is possible if parameters are blocked in the right way. For models with a single endogenous regressor, textbook expositions of Gibbs sampling include Lancaster (2004) and Rossi, et al. (2005). Here we describe a Gibbs sampler for a model with multiple instruments and endogenous right-hand variables.

The literature on substantive issues about Bayesian estimation of instrumental variable models is large. Two good places to start are Geweke (1996) and Kleibergen and Zivot (2003). Conley et. al. (2008) allow for a semi-parametric model for the error terms in the single endogenous regressor model of Rossi et. al. (2005). Hooerheide at. al. (2007) propose a natural conjugate prior for the IV problem. For an

*Tim Cogley: Department of Economics, New York University, 19 W. 4th St., 6FL, New York, NY 10012, email: tim.cogley@nyu.edu. Dick Startz: Department of Economics, 2127 North Hall, University of California, Santa Barbara, CA 93106, email: startz@ucsb.edu. Matlab implementations of the Gibbs sampler and marginal likelihood calculator are available from the second author. Our thanks to Peter Rossi, Galli Alain, and Gerdie Everaert for helpful comments.

exposition of MCMC with nonnormal priors, see Gao and Lahiri (2000). Hooerheide at. al. (2007b) discuss the properties of conditional distributions for IV estimators in MCMC settings. Zellner et. al. (2011) discuss an alternative to MCMC procedures with a flat prior and a single endogenous regressor.

2 Nonlinear SUR formulation and Gibbs sampling

Consider a model with a single structural equation

$$y = X\beta + \varepsilon, \tag{1}$$

where y is an $N \times 1$ vector of observations on a dependent variable, X is an $N \times k$ matrix of endogenous regressors, β is a $k \times 1$ vector of parameters, and ε is an $N \times 1$ vector of residuals. The "first-stage" equations are

$$X = Z\Gamma + v, \tag{2}$$

where Z is an $N \times q$ matrix of instruments, Γ is a $q \times k$ matrix of coefficients, and v is an $N \times k$ matrix of errors. Exogenous variables in the structural equation may be included in both the X and Z matrices without loss of generality. So that the model satisfies an order condition for identification, we assume that q is at least as large as k.

Substituting equations (2) into equation (1) gives a restricted reduced form,

$$y = Z\Gamma\beta + v_0, \tag{3}$$

where $v_0 = \varepsilon + v\beta$. Together, equations (2) and (3) yield a seemingly-unrelated regression with nonlinear cross-equation parameter restrictions,

$$\begin{bmatrix} y \\ x_1 \\ ... \\ x_k \end{bmatrix} = (I_{k+1} \otimes Z) \begin{bmatrix} \Gamma\beta \\ \gamma_1 \\ ... \\ \gamma_k \end{bmatrix} + \begin{bmatrix} v_0 \\ v_1 \\ ... \\ v_k \end{bmatrix}, \tag{4}$$

where x_i is the ith column of X and γ_j is the jth column of Γ.

Let $u_n = [v_{n0}, v_{n1}, ..., v_{nk}]'$ represent the vector of SUR errors for observation n. We assume that u_n is i.i.d normal with mean zero and covariance Σ. Although u_n is serially uncorrelated, its elements can be correlated contemporaneously. Hence Σ need not be diagonal and typically won't be.

3 Gibbs sampling

Gibbs sampling is accomplished in three blocks: $\Sigma|\Gamma,\beta$, $\beta|\Gamma,\Sigma$, and $\Gamma|\beta,\Sigma$. We assume that parameters are independent a priori across blocks,

$$p(\beta,\Gamma,\Sigma) = p(\beta)p(\Gamma)p(\Sigma), \tag{5}$$

Marginal priors are specified so that the conditional posterior for each block has a convenient form. The data are denoted $D = (y, X, Z)$.

3.1 Block 1: $p(\Sigma|\Gamma,\beta,D)$

Conditional on (Γ,β,D), the residuals $u = (v_0', v_1', ...v_k')'$ in equation (4) are observable. We assume the prior $p(\Sigma)$ is inverse Wishart with scale matrix $\underline{S}$ and degrees of freedom $df \geq k+1$. Since the conditional likelihood function is Gaussian, the posterior is also inverse Wishart with scale matrix $\bar{S} = \underline{S} + uu'$ and degrees of freedom $DF = df + N$. Hence Σ can be drawn by sampling from a $IW(\bar{S}, DF)$ distribution.

3.2 Block 2: $p(\beta|\Gamma,\Sigma,D)$

Since Γ is known, the restricted reduced form (equation 4) can be written as

$$\begin{bmatrix} y \\ x_1 - Z\gamma_1 \\ ... \\ x_k - Z\gamma_k \end{bmatrix} = \begin{bmatrix} Z\gamma_1 & ... & Z\gamma_k \\ 0 & ... & 0 \\ ... & ... & ... \\ 0 & ... & 0 \end{bmatrix} \beta + \begin{bmatrix} v_0 \\ v_1 \\ ... \\ v_k \end{bmatrix}, \tag{6}$$

or, letting $\hat{x}_i = Z\gamma_i$ be the analog of the 'fitted' values from the first-stage regression given Γ,

$$\begin{bmatrix} y \\ x_1 - \hat{x}_1 \\ ... \\ x_k - \hat{x}_k \end{bmatrix} = \begin{bmatrix} \hat{x}_1 & ... & \hat{x}_k \\ 0 & ... & 0 \\ ... & ... & ... \\ 0 & ... & 0 \end{bmatrix} \beta + \begin{bmatrix} v_0 \\ v_1 \\ ... \\ v_k \end{bmatrix}. \tag{7}$$

Since the regressors in rows 2 through k are zero, the residuals in those equations are conditionally observable. Given joint normality of the errors, the conditional mean and variance of v_0 can be found by projecting v_0 onto $v = (v_1, ..., v_k)$. Partition Σ as

$$\Sigma = \begin{bmatrix} \sigma_0^2 & \Sigma_{v_0 v} \\ \Sigma_{v v_0} & \Sigma_{vv} \end{bmatrix}, \tag{8}$$

where $\sigma_0^2 = var(v_0)$, $\Sigma_{vv} = var(v)$, and $\Sigma_{v_0 u} = cov(v_0, v)$. Then v_0 is conditionally normal with mean

$$E(v_0|v_1, ..., v_k) = \left[\begin{array}{ccc} v_1 & ... & v_k \end{array} \right] \Sigma_{vv}^{-1} \Sigma_{vv_0}, \tag{9}$$

and variance

$$var(v_0|v_1, ..., v_k) = \sigma_0^2 - \Sigma_{v_0 v} \Sigma_{vv}^{-1} \Sigma_{vv_0}. \tag{10}$$

The residual from this projection, $\eta_0 \equiv v_0 - E(v_0|v_1, ..., v_k)$, has conditional mean zero and is conditionally independent of $(v_1, ..., v_k)$. After subtracting $E(v_0|v_1, ..., v_k)$ from both sides of the first equation in (7), we find

$$\left[\begin{array}{c} y - E(v_0|v_1, ..., v_k) \\ x_1 - \hat{x}_1 \\ ... \\ x_k - \hat{x}_k \end{array} \right] = \left[\begin{array}{ccc} \hat{x}_1 & ... & \hat{x}_k \\ 0 & ... & 0 \\ ... & ... & ... \\ 0 & ... & 0 \end{array} \right] \beta + \left[\begin{array}{c} \eta_0 \\ v_1 \\ ... \\ v_k \end{array} \right]. \tag{11}$$

Because the transformed residual η_0 is independent of $(v_1, ..., v_k)$, the k bottom rows are irrelevant for estimating β.

Assume a normal prior $p(\beta) = N(\beta_0, \underline{V}_\beta)$. Let $\bar{V}_\beta = \sigma_{\eta_0}^2 \left(\sigma_{\eta_0}^2 \underline{V}_\beta^{-1} + \hat{X}' \hat{X} \right)^{-1}$ and $\bar{\beta} = \bar{V}_\beta \left[\underline{V}_\beta^{-1} \beta_0 + \hat{X}' \left(y - E(v_0|v_1, ..., v_k) \right) \left(\sigma_{\eta_0}^2 \right)^{-1} \right]$. Then the conditional posterior is

$$p(\beta|\Gamma, \Sigma, D) = N(\bar{\beta}, \bar{V}_\beta). \tag{12}$$

3.3 Block 3: $p(\Gamma|\beta, \Sigma, D)$

Since β is known, the restricted reduced form (equation 4) can be written as a seemingly unrelated regression that is linear in the unknown Γ parameters,[1]

$$\left[\begin{array}{c} y \\ x_1 \\ ... \\ x_k \end{array} \right] = \left[\begin{array}{c} \beta' \otimes Z \\ I_k \otimes Z \end{array} \right] \left[\begin{array}{c} \gamma_1 \\ ... \\ \gamma_k \end{array} \right] + \left[\begin{array}{c} v_0 \\ v_1 \\ ... \\ v_k \end{array} \right]. \tag{13}$$

Call the left-hand side variables $\tilde{y}_i$ and the right hand side variables $\tilde{X}_i$. Assuming a normal prior $p(vec(\Gamma)) = N(\gamma_0, \underline{V}_\gamma)$, the conditional posterior $p(\Gamma|\beta, \Sigma, D)$ is normal with variance

$$\bar{V}_\gamma = \left(\underline{V}_\gamma^{-1} + \sum_{i=1}^{k+1} \tilde{X}_i' \Sigma^{-1} \tilde{X}_i \right), \tag{14}$$

[1] To verify the equivalence with (4), write $\beta' \otimes Z$ longhand and rearrange terms.

and mean

$$\bar{\gamma} = \bar{V}_\gamma \left(\underline{V}_\gamma^{-1} \gamma_0 + + \sum_{i=1}^{k+1} \tilde{X}_i' \Sigma^{-1} \tilde{y}_i \right). \tag{15}$$

4 Marginal Likelihood Calculation

The marginal likelihood of the model can be computed by applying Chib's (1995) method. Let θ be the three blocks $\{\beta, \Gamma, \Sigma\}$. The "basic marginal likelihood identity" is

$$p(y, X|Z) = \frac{p(y, X|\theta^*, Z) p(\theta^*)}{p(\theta^*|D)}. \tag{16}$$

While the identity in equation (16) holds for any value of θ^*, we want to be sure that Σ^* is positive definite and should also recognize that, since the frequentist estimate of β does not have a finite first moment in the just-identified case, the MCMC mean of β may not be a wise choice for β^*. As a recommendation, let $\beta^* = median(\beta^{(s)})$ and $\Gamma^* = median(\Gamma^{(s)})$, where the superscript (s) indicates draws from the MCMC distribution. Then, using the residuals $e_i(\beta^*, \Gamma^*)$, let $\Sigma^* = \sum_{i=1}^N e_i e_i'/N$.

The values of the likelihood and prior in equation (16) can be computed directly. Note that the log likelihood is given by

$$\log p(y, X|\theta^*, Z) = -\frac{N(k+1)}{2} \log(2\pi) - \frac{N}{2} \log \det \Sigma^* - \frac{1}{2} \sum_{i=1}^N e_i'(\Sigma^*)^{-1} e_i.$$

The third term simplifies as follows,

$$\begin{aligned} \sum_{i=1}^N e_i'(\Sigma^*)^{-1} e_i &= \sum_{i=1}^N tr\left(e_i'(\Sigma^*)^{-1} e_i \right), \\ &= tr\left((\Sigma^*)^{-1} \sum_{i=1}^N e_i e_i' \right), \\ &= tr\left(N I_{k+1} \right), \\ &= N(k+1). \end{aligned} \tag{17}$$

Hence

$$\log p(y, X|\theta^*, Z) = -\frac{N(k+1)}{2} \left(1 + \log(2\pi)\right) - \frac{N}{2} \log \det \Sigma^*. \tag{18}$$

Break up the posterior density for θ^* as follows.

$$p(\beta^*, \Gamma^*, \Sigma^*|D) = p(\Gamma^*|D) \times p(\beta^*|D, \Gamma^*) \times p(\Sigma^*|D, \beta^*, \Gamma^*). \tag{19}$$

The order of parameters in equation (19) reflects computational efficiency rather than anything more fundamental. The evaluation of (19) proceeds in three steps. If the

draws from the MCMC are indexed $(s) = 1, ...S$, then

$$p(vec(\Gamma^*)|D) \approx S^{-1} \sum_{(s)=1}^{S} f_N(vec(\Gamma^*); \bar{\beta}^{(s)}, \bar{V}_{\beta}^{(s)}). \tag{20}$$

where $f_N(\cdot)$ is the density of a normal pdf and $\bar{\gamma}^{(s)}, \bar{V}_{\gamma}^{(s)}$ have already been calculated in Block 3 above in the original MCMC.

Calculation of $p(\beta^*|D, \Gamma^*)$ – the second term in (19) – requires a second run of the MCMC procedure outlined above, except that Step 3 is omitted and Γ^* replaces draws of Γ. Index the draws of this auxiliary sampler by $(s_\beta) = 1, ...S_\beta$. Then

$$p(\beta^*|D, \Gamma^*) \approx S_{\beta}^{-1} \sum_{(s_\beta)=1}^{S_\beta} f_N(vec(\beta^*); \bar{\gamma}^{(s_\beta)}, \bar{V}_{\gamma}^{(s_\beta)}), \tag{21}$$

where $\bar{\gamma}^{(s)}, \bar{V}_{\gamma}^{(s)}$ are calculated as in Block 2 above.

The conditional posterior $p(\Sigma|D, \beta, \Gamma)$ is simulated in Block 1 above. Hence the final term in (19) can be evaluated as

$$p(\Sigma^*|D, \beta^*, \Gamma^*) = f_{IW}(\Sigma^*|DF, \bar{S}), \tag{22}$$

where the scale matrix $\bar{S}$ is evaluated at β^*, Γ^*,as in Block 1 above.

Collecting the results of (20), (21), and (22) and multiplying them together delivers the denominator of (16). Dividing the posterior kernel $p(y, X|\theta^*, Z)p(\theta^*)$ by the result delivers the marginal likelihood $p(y, X|Z)$.

5 References

Chib, S., 1995. Marginal Likelihood From the Gibbs Output, *Journal of the American Statistical Association* 90, 1313-1321.

Conley, T., C. Hansen, R. McCulloch, and P. Rossi, 2008. A semi-parametric bayesian approach to the instrumental variable problem, *Journal of Econometrics*, 144, 276-305.

Geweke, J., 1996. Bayesian reduced rank regression in econometrics. *Journal of Econometrics* 75, 121–146.

Gao, C. and K. Lahiri, 2000. MCMC algorithms for two recent Bayesian limited information estimators. *Economics Letters* 66, 121-126.

Hoogerheide L., Kleibergen, F., and van Dijk, H. 2007 Natural conjugate priors for the instrumental variables regression model applied to the Angrist–Krueger data. *Journal of Econometrics* 138, 63–103.

_____, Kaashoek, J., and van Dijk, H. 2007b. On the shape of posterior densities and credible sets in instrumental variable regression models with reduced rank: An application of flexible sampling methods using neural networks. *Journal of Econometrics* 139, 154-180.

Kleibergen, F., and Zivot, E., 2003. Bayesian and classical approaches to instrumental variable regression. *Journal of Econometrics* 114, 29–72.

Lancaster, T., 2004. *An Introduction to Modern Bayesian Econometrics*, Blackwell Publishing.

Rossi, P., G. M. Allenby, and R. McCulloch, 2005. *Bayesian Statistics and Marketing*, John Wiley & Sons.

Zellner, A., T. Ando, N. Basturk, L. Hoogerheide, and H. van Dijk, 2011. Instrumental variables, errors in variables, and simultaneous equations models: applicability and limitations of direct monte carlo. Tinbergen Institute Discussion Paper 2011-137/4.