Sheffield Economic Research Paper Series

SERP Number: 2003008

Roberto León González

Sampling the Variance-Covariance Matrix in the Bayesian Multivariate Probit Model

September 2003

Department of Economics
University of Sheffield
9 Mappin Street
Sheffield
S1 4DT
United Kingdom
www.shef.ac.uk/economics

Sampling the Variance-Covariance Matrix in the Bayesian Multivariate Probit Model*

Roberto León González

Department of Economics

University of Sheffield

September 2003

Abstract

This paper is concerned with the Bayesian estimation of a Multivariate Probit model. In particular, this paper provides a method to sample the restricted variance-covariance matrix directly from its conditional posterior density. The method allows the application of a standard Gibbs sampling algorithm to sample from the posterior density of the parameters, and hence it avoids the use of a Metropolis step. The method uses a decomposition of the Inverted Wishart density and alternative identification restrictions.

*This paper was circulated before as part of the discussion paper 09/02 of the Department of Economics at the University of York

1 Introduction

A common feature in qualitative-dependent variable models is that the scale of the latent variable is not identified. As a normalization, restrictions are usually placed upon the variance parameters. In the context of the Bayesian multinomial probit, just one of the elements in the variance-covariance matrix is restricted to be one, and several algorithms to sample the variance-covariance matrix directly are available (Cowles 1996, McCulloch et al. 2000 and Nobile 2000).

In the multivariate probit model, a larger number of normalization restrictions are necessary in order to make the model identifiable. In a Bayesian context, Chib and Greenberg (1998) propose to restrict the diagonal elements of the variance-covariance matrix (Σ) to be equal to one. However, there is not any method to obtain a draw from the posterior conditional density of Σ subject to such restriction. In order to overcome this difficulty, Chib and Greenberg (1998) use a Metropolis step (Metropolis et al. 1953). However, a Metropolis step requires a proposal density that is a good approximation of the posterior density. For this reason, Chib and Greenberg (1998) proposed to incorporate a maximization algorithm within the MCMC algorithm in order to calculate the mode and hessian of the conditional posterior density of Σ. However, even with this additional computational and programming burden, a Metropolis step will probably not work well when the number of variables is large.

Lui (2001) proposes an algorithm to sample Σ directly from its conditional posterior. However, the algorithm relies on the specification of an improper prior for the covariance matrix and restricts each equation to contain the same set of regressors. Unfortunately, there are no results that ensure that the posterior density would exist if an improper prior is specified, and hence the specification of a proper prior seems to be necessary. In addition, proper priors are often required for Bayes factors to be well defined, specially when testing hypothesis about model specific parameters.

This paper proposes a simple method to sample Σ directly from its conditional posterior density, using a proper prior. The method relies on an alternative normalization and on a decomposition of the inverted Wishart density. Despite of the use of an alternative normalization, the method can also yield results according to the normalization used by Chib and Greenberg (1998).

The plan of the paper is as follows. Section 2 describes the notation for the model, and for simplicity in the exposition concentrates on the bivariate probit model (i.e. two equations). Section 3 looks at the more general case. Section 4 describes how the algorithm can be adapted when the normalization of Chib and Greenberg (1998) is used. Section 5 summarizes the paper.

2 Sampling the Variance-Covariance Matrix in the Bivariate Probit Model

Let Y_i be a $(T \times 1)$ vector of zeros and ones. In the multivariate probit model, each component y_{it} of Y_i is determined by a continuous unobserved latent variable y_{it}^* generated according to the following process,

$$y_{it}^* = X_{it}\beta_t + e_{it} \qquad i = 1, ..., N \qquad t = 1, ..., T \tag{1}$$

the vector $e_i = (e_{i1}, ..., e_{iT})^T$ is normally distributed with zero mean and covariance matrix $\Sigma = (\sigma_{jk})$. The binary variable y_{it} is equal to one if and only if $y_{it}^* \geq 0$, and is equal to zero otherwise. X_{it} is a $1 \times k_t$ vector of regressors and β_t is a vector of parameters.

2.1 Identification in the Bivariate Probit Model

The likelihood contribution of an observation $(0, 1)$ is,

$$\Pr\{y_{i1} = 0, y_{i2} = 1\} = \Pr\left\{\begin{array}{c} X_{i1}\beta_1 + e_{i1} \leq 0 \\ X_{i2}\beta_2 + e_{i2} > 0 \end{array}\right\} \tag{2}$$

Let $\Delta = (\delta_{jk})$ be the lower triangular Cholesky decomposition of Σ, so that $\Sigma = \Delta\Delta^T$.

$$\Delta = \begin{pmatrix} \sqrt{\sigma_{11}} & 0 \\ \sigma_{12}/\sqrt{\sigma_{11}} & \sqrt{\sigma_{22} - (\sigma_{12})^2/\sigma_{11}} \end{pmatrix} \tag{3}$$

Then the vector e_i can be seen as a transformation of a random vector ε_i that contains independent standard normal variables. That is, $e_i = \Delta\varepsilon_i$, where ε_i follows a $N(0, I)$.

The probability in (2) can be rewritten as,

$$\Pr\left\{y_{i1}=0,y_{i2}=1\right\}=\Pr\left\{\begin{array}{c} X_{i1}\beta_1+\sqrt{\sigma_{11}}\varepsilon_{i1}\leq 0, \\ X_{i2}\beta_2+\frac{\sigma_{12}}{\sqrt{\sigma_{11}}}\varepsilon_{i1}+\sqrt{\sigma_{22}-\left(\sigma_{12}\right)^2/\sigma_{11}}\varepsilon_{i2}>0 \end{array}\right\} \quad (4)$$

From expression (4), different values for (β, Σ) give the same value for the probability. In particular, for two arbitrary positive constants (c, d), the value of the parameters $\{c(\beta_1, \delta_{11}), d(\beta_2, \delta_{21}, \delta_{22})\}$ give the same value for the probability as $\{\beta_1, \delta_{11}, \beta_2, \delta_{21}, \delta_{22}\}$. Hence, the model is not identified.

The most common normalization in the literature is to fix $\sigma_{11} = \sigma_{22} = 1$ (e.g. Chib and Greenberg 1998). However, the following section shows that it is more convenient from the point of view of computational tractability to choose $\sigma_{11} = \sigma_{22} - (\sigma_{12})^2/\sigma_{11} = 1$. From expression (4), both normalizations make the model identified without imposing any unnecessary restrictions upon the parameters.

2.2 The Sampling Method

In the Bayesian approach, model specification is completed by providing a prior distribution for the parameters. Let the prior for Σ be an Inverted Wishart $IW_2(df_0, K_0)$ distribution conditional to the restriction that $\sigma_{11} = \sigma_{22} - (\sigma_{12})^2/\sigma_{11} = 1$. That is, given a matrix Σ that satisfies the restriction, the kernel of the prior is:

$$|\Sigma|^{-df_0/2}\exp\left(-1/2tr\left(\Sigma^{-1}K_0\right)\right) \quad (5)$$

The expected value of the unrestricted prior is $\frac{1}{df_0-2T-2}K_0$, where T is the dimension of Σ. The definition of inverted Wishart distribution used here is the one described in Press (1986, pp. 117).

The conditional posterior of Σ given parameters β and latent data $\{y^*_{it} : t = 1, \ldots, T\}_{i=1}^T$ is an inverted Wishart $IW_2(df, K)$ with the restriction that $\sigma_{11} = \sigma_{22} - (\sigma_{12})^2/\sigma_{11} = 1$. The parameters of this inverted Wishart are $df = df_0 + N$ and $K = K_0 + \sum_{i=1}^N e_i e_i^T$.

The following Theorem is useful to sample Σ conditional on the normalisation $\delta_{11} = \delta_{22} = 1$. Cowles et al (1996) sample from an Inverted Wishart distribution with the restriction that one element in the diagonal is equal to one. They made use of the following theorem, which can be found in Bauwens et al. (1999, pages 305-306).

Theorem 1 *Let* Σ *be distributed as an* $IW_d(n, G)$, *and be partitioned as* $\Sigma = (\Sigma_{ij})$, *i,j=1,2, where* Σ_{11} *is a* $q \times q$ *matrix. Define* $\Sigma_{22\cdot 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$, *then*

$$\begin{aligned} &1) \qquad \Sigma_{22\cdot 1}|\Sigma_{11} \sim IW_{d-q}(n-q, G_{22\cdot 1}) \\ &2) \quad \Sigma_{12}|(\Sigma_{22\cdot 1}, \Sigma_{11}) \sim \quad MN\left(\Sigma_{11}(G_{11})^{-1}G_{12}, \Sigma_{11}(G_{11})^{-1}\Sigma_{22\cdot 1}\Sigma_{11}^{T}\right) \end{aligned}$$

where MN *refers to a matrix normal distribution.*

Theorem 1 states that the conditional posterior of σ_{12} given β and all latent data y_{it}^* is a normal distribution and hence it can be sampled directly.

3 Dimension larger than 2

The proposed normalization in the bivariate probit is to set the variance of e_{i1} (σ_{11}) and the conditional variance of e_{i2} given e_{i1} $\left(\sigma_{22} - (\sigma_{12})^2/\sigma_{e11}\right)$ both equal to one. Or in other words, to fix the elements in the diagonal of the cholesky decomposition equal to one. Similarly, in the case of $T \geq 2$ the proposed normalization is

$$Var(e_{i1}) = Var(e_{i2}|e_{i1}) = Var(e_{i3}|e_{i2}, e_{i1}) = ... = Var\left(e_{iT}|e_{i(T-1)}, ..., e_{i1}\right) = 1 \quad (6)$$

This normalization can also be shown to be the same as fixing the elements in the diagonal of the Cholesky decomposition equal to one. The following lemma gives an statistical interpretation to each of the elements of the Cholesky decomposition.

Lemma 1 *If* e_i *follows a* $N(0, \Sigma)$, *and* Δ *is the lower triangular cholesky decomposition, then*

$$\Delta = \begin{pmatrix} \sqrt{\sigma_{11}} & 0 & 0 & 0 & & 0 \\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}}} & \sqrt{\sigma_{22\cdot 1}} & 0 & 0 & & 0 \\ \frac{\sigma_{13}}{\sqrt{\sigma_{11}}} & \frac{\sigma_{23\cdot 1}}{\sqrt{\sigma_{22\cdot 1}}} & \sqrt{\sigma_{33\cdot 12}} & 0 & & 0 \\ \frac{\sigma_{14}}{\sqrt{\sigma_{11}}} & \frac{\sigma_{24\cdot 1}}{\sqrt{\sigma_{22\cdot 1}}} & \frac{\sigma_{34\cdot 12}}{\sqrt{\sigma_{33\cdot 12}}} & \sqrt{\sigma_{44\cdot 123}} & & 0 \\ & & & & \cdots & \\ \frac{\sigma_{1T}}{\sqrt{\sigma_{11}}} & \frac{\sigma_{2T\cdot 1}}{\sqrt{\sigma_{22\cdot 1}}} & \frac{\sigma_{3T\cdot 12}}{\sqrt{\sigma_{33\cdot 12}}} & \frac{\sigma_{4T\cdot 123}}{\sqrt{\sigma_{44\cdot 123}}} & \cdots & \sqrt{\sigma_{TT\cdot 123...(T-1)}} \end{pmatrix}$$

where $\sigma_{tt\cdot vwz} = Var(e_{it}|e_{iv}, e_{iw}, e_{iz})$ *and* $\sigma_{th\cdot vwz} = Cov(e_{it}, e_{ih}|e_{iv}, e_{iw}, e_{iz})$.

Proof. If ε_i follows a $N(0, I)$, then the linear combination $\Delta\varepsilon_i$ also follows a normal $N(0, \Sigma)$. Hence, e_i can be expressed as $e_i = \Delta\varepsilon_i$, that is:

$$
\begin{aligned}
&e_{i1} = \delta_{11}\varepsilon_{i1} \\
&e_{i2} = \delta_{21}\varepsilon_{i1} + \delta_{22}\varepsilon_{i2} \\
&e_{i3} = \delta_{31}\varepsilon_{i1} + \delta_{32}\varepsilon_{i2} + \delta_{33}\varepsilon_{i3} \\
&e_{i4} = \delta_{41}\varepsilon_{i1} + \delta_{42}\varepsilon_{i2} + \delta_{43}\varepsilon_{i3} + \delta_{44}\varepsilon_{i4} \\
&\ldots \\
&e_{iT} = \delta_{T1}\varepsilon_{i1} + \delta_{T2}\varepsilon_{i2} + \delta_{T3}\varepsilon_{i3} + \delta_{T4}\varepsilon_{i4} + \ldots + \delta_{TT}\varepsilon_{iT}
\end{aligned}
\tag{7}
$$

Using these equations, Σ can be related to Δ:

$$
\begin{aligned}
&Var(e_{i1}) = \sigma_{11} = Var(\delta_{11}\varepsilon_{i1}) = (\delta_{11})^2 \\
&Cov(e_{i1}, e_{i2}) = \sigma_{12} = Cov(\delta_{11}\varepsilon_{i1}, \delta_{21}\varepsilon_{i1} + \delta_{22}\varepsilon_{i2}) = \delta_{11}\delta_{21} \\
&Var(e_{i2}|e_{i1}) = \sigma_{22\cdot 1} = Var(\delta_{21}\varepsilon_{i1} + \delta_{22}\varepsilon_{i2}|\delta_{11}\varepsilon_{i1}) = (\delta_{22})^2 \\
&Cov(e_{i3}, e_{i1}) = \delta_{31}\delta_{11}, Cov(e_{i3}, e_{i2}|e_{i1}) = \delta_{32}\delta_{22}, Var(e_{i3}|e_{i1}, e_{i2}) = (\delta_{33})^2 \\
&Cov(e_{i4}, e_{i1}) = \delta_{41}\delta_{11}, Cov(e_{i4}, e_{i2}|e_{i1}) = \delta_{42}\delta_{22}, \\
&Cov(e_{i4}, e_{i3}|e_{i1}, e_{i2}) = \delta_{43}\delta_{33}, Var(e_{i4}|e_{i1}, e_{i2}, e_{i3}) = (\delta_{44})^2 \\
&Cov(e_{iT}|e_{i1}) = \delta_{T1}\delta_{11}, Cov(e_{iT}, e_{i2}|e_{i1}) = \delta_{T2}\delta_{22}, \\
&Cov(e_{iT}, e_{i3}|e_{i1}, e_{i2}) = \delta_{T3}\delta_{33}, \ldots, Var\left(e_{iT}|e_{i1}, e_{i2}, e_{i3}, \ldots, e_{i(T-1)}\right) = (\delta_{TT})^2
\end{aligned}
$$

■

Hence, restriction (6) is equivalent to fixing the diagonal elements of the Cholesky decomposition equal to one. In addition, from (7) this restriction identifies the model, not imposing any unnecessary restrictions upon the parameters.

Let the prior for the elements out of the diagonal of Σ have a kernel equal to expression (5). The elements in the diagonal are fixed by normalisation (6), which ensures that Σ is positive definite. Note that since normalization (6) ensures that Σ is positive definite, the non-diagonal elements of Σ are not subject to any restrictions. From the Appendix, the prior variance of the free elements in Σ can be made arbitrarily large by choosing small values in the diagonal of K_0.

Let Σ_{jj} be the sub-matrix of Σ containing the first j rows and the first j columns of Σ and K_{jj} the corresponding sub-matrix of K. Let Σ_j be the vertical vector containing the first $(j-1)$ rows in the jth column and let K_j be the corresponding sub-matrix of K. Following Theorem 1, if Σ follows an unrestricted $IW_T(df, K)$, then

- Σ_T conditional on $\Sigma_{(T-1)(T-1)}$ and $\sigma_{TT\cdot 12\ldots(T-1)}$ follows a normal distribution
- $\sigma_{TT\cdot 12\ldots(T-1)}$ is independent of $\Sigma_{(T-1)(T-1)}$ and follows an inverted Wishart distribution.
- $\Sigma_{(T-1)(T-1)}$ follows an inverted Wishart distribution.

The third property holds because the marginal distribution of sub-matrices centered in the diagonal is also an inverted Wishart distribution (Press 1986, pp. 118-119). Since $\sigma_{TT\cdot 12\ldots(T-1)}$ is independent of $\Sigma_{(T-1)(T-1)}$, conditioning on $\sigma_{TT\cdot 12\ldots(T-1)} = 1$ does not change the marginal distribution of $\Sigma_{(T-1)(T-1)}$.

Consider now that Σ follows an $IW_T(df, K)$ with the restriction that $\sigma_{11} = \sigma_{22\cdot 1} = \ldots = \sigma_{TT\cdot 12\ldots(T-1)} = 1$. By applying the above argument recursively, the marginal distribution of Σ_2 given the restriction is a normal distribution. In addition, the distribution of Σ_n conditional on $\Sigma_{(n-1)(n-1)}$ and $\sigma_{nn\cdot 12\ldots(n-1)}$ is also a normal distribution, for $2 \leq n \leq T$. The appendix gives full detail of the distributions involved in this decomposition of the inverted Wishart density.

The following algorithm describes how Σ can be sampled from its conditional posterior distribution:

Algorithm 1 *Step 0: Fix*

$$\sigma_{11} = \sigma_{22\cdot 1} = \sigma_{33\cdot 12} = \ldots = \sigma_{TT\cdot 12\ldots(T-1)} = 1$$

Step 1: Sample σ_{12} conditional on σ_{11} and $\sigma_{22\cdot 1}$ from a

$$N\left(\sigma_{11}K_{11}^{-1}K_2, (\sigma_{11})^2 K_{11}^{-1}\sigma_{22\cdot 1}\right)$$

Step 2: Fix $\sigma_{22} = \sigma_{22\cdot 1} + (\sigma_{12})^2 \sigma_{11}^{-1}$

Step 3: Sample Σ_3 conditional on Σ_{22} and $\sigma_{33\cdot 12}$ from a

$$N\left(\Sigma_{22}K_{22}^{-1}K_3, \sigma_{33\cdot 12}\Sigma_{22}K_{22}^{-1}\Sigma_{22}\right)$$

Step 4: Fix $\sigma_{33} = \sigma_{33\cdot 12} + \Sigma_3^T \Sigma_{22}^{-1}\Sigma_3$.

Step 2(n–1)-1: Sample Σ_n *conditional on* $\Sigma_{(n-1)(n-1)}$ *and* $\sigma_{nn\cdot 12\ldots(n-1)}$ *from a*

$N\left(\Sigma_{(n-1)(n-1)}K^{-1}_{(n-1)(n-1)}K_n, \sigma_{nn\cdot 12\ldots(n-1)}\Sigma_{(n-1)(n-1)}K^{-1}_{(n-1)(n-1)}\Sigma_{(n-1)(n-1)}\right)$

Step 2(n-1): Fix $\sigma_{nn} = \sigma_{TT\cdot 12\ldots(T-1)} + \Sigma_n^T\Sigma^{-1}_{(n-1)(n-1)}\Sigma_n$.

4 An alternative normalization.

This section describes how to transform the estimated values for (β, Σ) if another normalization is chosen. In particular, instead of normalization (6), one might be interested in choosing the more widely used normalization:

$$\sigma_{11} = \sigma_{22} = \ldots = \sigma_{TT} = 1 \tag{8}$$

A sample from the posterior of (β, Σ) given normalization (8) can be obtained by simply transforming the values sampled using Algorithm 1. Let C_1 be the diagonal matrix of dimension T with diagonal equal to $(1/\sqrt{\sigma_{11}}$, $1/\sqrt{\sigma_{22\cdot 1}}$, $1/\sqrt{\sigma_{33\cdot 12}}$,..., $1/\sqrt{\sigma_{TT\cdot 12\ldots(T-1)}})$ and let C_2 be the diagonal matrix of dimension T with diagonal equal to $(1/\sqrt{\sigma_{11}}$, $1/\sqrt{\sigma_{22}}$, $1/\sqrt{\sigma_{33}}$,..., $1/\sqrt{\sigma_{TT}})$. The parameters that are identified with normalization (6) are

$$\left(1/\sqrt{\sigma_{11}}\beta_1, \ldots, 1/\sqrt{\sigma_{TT\cdot 12\ldots(T-1)}}\beta_T, C_1\Sigma C_1\right) \tag{9}$$

When normalization (8) is used, the identified parameters are

$$(1/\sqrt{\sigma_{11}}\beta_1, \ldots, 1/\sqrt{\sigma_{TT}}\beta_T, C_2\Sigma C_2) \tag{10}$$

Let $\left(\beta^k, \Sigma^k\right)_1$ be the kth value in the chain when normalization (6) is used. And let $\left(\beta^k, \Sigma^k\right)_2$ be the kth value in a chain in which normalization (8) is chosen. To obtain a sample from the posterior when normalization (8) is used, transform $\left(\beta^k, \Sigma^k\right)_1$ in the following way:

- Construct C_2^k as the diagonal matrix with diagonal equal to

$$\left(1/\sqrt{\sigma_{11}^k}, 1/\sqrt{\sigma_{22}^k}, 1/\sqrt{\sigma_{33}^k}, \ldots, 1/\sqrt{\sigma_{TT}^k}\right)$$

- Fix $\left(\beta^k, \Sigma^k\right)_2 = \left(1/\sqrt{\sigma_{11}^k}\beta_1, \ldots, 1/\sqrt{\sigma_{TT}^k}\beta_T, , C_2^k\Sigma^k C_2^k\right)$

5 Conclusion

This paper has described a simple method (Algorithm 1) to sample Σ directly from its posterior conditional density. This method makes it possible to use a standard Gibbs Sampling algorithm (e.g. Gelfand and Smith 1990) and hence avoids the use of a Metropolis step.

Appendix: Decomposition of an inverted Wishart

Let $f_N(x;\mu,\Phi)$ denote the density function of a $N(\mu,\Phi)$, evaluated at x. And let $f_{IW}(x;p,df,K)$ denote the density function of an inverted Wishart of dimension p, degrees of freedom df, and expected value $K\frac{1}{df-2p-2}$.

Proposition 1 *If Σ follows an unrestricted inverted Wishart $IW_T(df,K)$, then the density function of $\left(\sigma_{11},\sigma_{22\cdot 1},\sigma_{12},\sigma_{33\cdot 12},\Sigma_3,...,\sigma_{TT\cdot 12...(T-1)},\Sigma_T\right)$ can be expressed as:*

$$\begin{aligned}
&f_{IW}(\sigma_{11};1,df-2T+2,k_{11})\times\\
&f_{IW}(\sigma_{22\cdot 1};1,df-2T+3,k_{22\cdot 1})\times\\
&f_N\left(\sigma_{12};\sigma_{11}(k_{11})^{-1}K_2,\sigma_{11}(k_{11})^{-1}\sigma_{11}\sigma_{22\cdot 1}\right)\times\\
&f_{IW}(\sigma_{33\cdot 12};1,df-2T+4,k_{33\cdot 12})\times\\
&f_N\left(\Sigma_3;\Sigma_{22}(K_{22})^{-1}K_3,\Sigma_{22}(K_{22})^{-1}\Sigma_{22}\sigma_{33\cdot 12}\right)\times\\
&...\times\\
&...\times\\
&f_{IW}\left(\sigma_{TT\cdot 12...(T-1)};1,df-2T+(T+1),k_{TT\cdot 12...(T-1)}\right)\times\\
&f_N(\Sigma_T;\Sigma_{(T-1)(T-1)}\left(K_{(T-1)(T-1)}\right)^{-1}K_T,\Sigma_{(T-1)(T-1)}\left(K_{(T-1)(T-1)}\right)^{-1}\\
&\Sigma_{(T-1)(T-1)}\sigma_{TT\cdot 12...(T-1)})
\end{aligned}$$

Proof. This decomposition results from applying recursively Theorem 1, in the way described in Section 3. ■

References

Bauwens, L., Lubrano, M. and Richard, J. (1999) *Bayesian Inference in Dynamic Econometrics Models.* Oxford University Press, Oxford.

Chib., S and Greenberg, E. (1998) "Analysis of multivariate probit models" *Biometrika* 85:347-361

Cowles, K., Carlin, B. P., and Connett, J. E. (1996) "Bayesian Tobit Modellig of Longitudinal Ordinal Clinical Trial Compliance Data with Nonignorable Missingness" *Journal of the American Statistical Association*, 91, 86-98.

Gelfand, A. E., and Smith, A. F. M. (1990) "Sampling Based Approaches to Calculating Marginal Densities" *Journal of the American Statistical Association*, 85, 398-409.

Liu, C., (2001) "The Art of Data Augmentation: Discussion" *Journal of Computational and Graphical Statistics* 10, 75-81

McCulloch, R. E., Polson, N. G. and Rossi, P. E. (2000) "A Bayesian Analysis of the Multinomial Probit Model with Fully Identified Parameters" *Journal of Econometrics* 99, 173-193

Metropolis, N., Rosenbluth, A. W. , Rosenbluth, M. N. , Teller, A. H. and Teller, E. (1953) "Equation of state calculations by fast compuing machine" *Journal of Chemical Physics* 21, 1087-91

Nobile, A. (2000) "Comment: Bayesian Multinomial Probit Models with a Normalization Constraint" *Journal of Econometrics* 99, 2, 335-45

Press, S. J. (1982) *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference*, Florida: Krieger.