FEDERAL RESERVE BANK *of* ATLANTA

A Gibbs Simulator for Restricted VAR Models

Daniel F. Waggoner and Tao Zha

Working Paper 2000-3
March 2000

Working Paper Series

A Gibbs Simulator for Restricted VAR Models

Daniel F. Waggoner and Tao Zha

Federal Reserve Bank of Atlanta
Working Paper 2000-3
March 2000

Abstract: Many economic applications call for simultaneous equations VAR modeling. We show that the existing importance sampler can be prohibitively inefficient for this type of models. We develop a Gibbs simulator that works for both simultaneous and recursive VAR models with a much broader range of linear restrictions than those in the existing literature. We show that the required computation is of an SUR type, and thus our method can be implemented cheaply even for large systems of multiple equations.

JEL classification: C15, C32, E50

Key words: simultaneous equations, recursive systems, independence, importance sampling, Gibbs sampler, posterior distributions

A Gibbs simulator for restricted VAR models

1. Introduction

Simultaneous equations time-series models such as structural vector autoregressions (VAR) have played an increasingly important role in modeling the actual economic behavior. Indeed, many economic theories call for models that take explicit account of non-recursive relationships among economic variables (Ingram and Whiteman 1994). As noted in Leeper, Sims, and Zha (1996), however, statistical inferences derived from a posterior simulator have remained a challenging task for simultaneous VAR modeling. The existing method used for simultaneous VAR models heavily relies on the importance sampling technique.[1] But this method can be prohibitively inefficient. When the shape of the posterior density for model parameters is highly non-Gaussian, as is often the case in non-recursive systems, the importance sampler tends to give unreasonably large weights for only a handful of simulated draws.

To remedy such inefficiency, we develop a method for Gibbs sampling in the simultaneous equations framework of VAR models. The method developed here is general enough to be applicable to linear restrictions on both contemporaneous and lagged relationships among endogenous and exogenous variables. In an earlier paper, Zha (1999) shows how to obtain finite-sample Bayesian inferences when certain lagged restrictions are in the block recursive form. We extend Zha's result and show that under general linear restrictions, Bayesian inferences of lagged and exogenous parameters can be obtained equation by equation. In theory, this result resembles the seemingly-unrelated-regressions (SUR) estimation (Zellner 1962; Highfield 1987); in practice, it enables one to obtain Bayesian inferences even for large systems of simultaneous equations.

---

[1] For examples of simultaneous VAR modeling, see Gordon and Leeper (1994); Leeper, Sims, and Zha (1996); Uhlig (1997); Zha (1999); Christiano, Eichenbaum, and Evans (1999); Kim and Roubini (1999).

For reduced-form VAR models, there exists a Gibbs sampler to approximate the joint distribution of the covariance matrix and reduced form parameters (Judge 1985, p.634). But this Gibbs sampler in general does not generate independent Monte Carlo (MC) draws and thus is less efficient than the widely used method of Doan (1992). We show that the Gibbs sampler developed in this paper produces independent MC draws for recursive (and thus reduced-form) VAR models, thus including the method of Doan as a special case. This theoretical result is of practical importance because most existing structural VAR models contain simultaneity only for two or three equations and all other equations are in the recursive form.

The remaining sections are organized as follows. The framework for simultaneous VAR models with general linear restrictions is laid out in Section 2. In Section 3 we prove that the posterior density function under such restrictions is of a SUR type. In Section 4 we develop the Gibbs sampler. An empirical example is discussed in Section 5.

2. Simultaneous equations framework with linear restrictions

The structural VAR models in the literature have the following simultaneous equations structure:

$$\mathbf{y}_t' \mathbf{A} = \sum_{\ell=1}^{p} \mathbf{y}_{t-\ell}' \mathbf{A}_\ell + \mathbf{z}' \mathbf{D} + \boldsymbol{\varepsilon}_t', \text{ for } t = 1,...,T. \tag{1}$$

where

- **A** is the $n \times n$ contemporaneous parameter matrix;

- $\mathbf{A}_\ell$ is an $n \times n$ lagged parameter matrix;

- **D** is the $h \times n$ exogenous parameter matrix;

- $\mathbf{y}_t$ is an $n \times 1$ column vector of endogenous variables at time t;

- $\mathbf{z}_t$ is an $h \times 1$ column vector of exogenous variables at time t;

- ε_t is an $n \times 1$ column vector of structural shocks at time t;

- p is the lag length and T is the sample size.

Following the VAR literature, the structural shocks are assumed to be Gaussian with

$$E\left(\varepsilon_t | \mathbf{y}_1, \cdots, \mathbf{y}_{t-1}, \mathbf{z}_1, \cdots, \mathbf{z}_T\right) = \underset{n \times 1}{\mathbf{0}} \text{ and } E\left(\varepsilon_t \varepsilon_t' | \mathbf{y}_1, \cdots, \mathbf{y}_{t-1}, \mathbf{z}_1, \cdots, \mathbf{z}_T\right) = \underset{n \times n}{\mathbf{I}} . \tag{2}$$

Note that the parameters of individual equations in (1) correspond to the columns of $\mathbf{A}$, $\mathbf{A}_\ell$, and $\mathbf{D}$. Let

$$\underset{1 \times k}{\mathbf{x}_t'} = \begin{bmatrix} \mathbf{y}_{t-1}' & \cdots & \mathbf{y}_{t-p}' & \mathbf{z}_t' \end{bmatrix} \text{ and } \underset{n \times k}{\mathbf{F}'} = \begin{bmatrix} \mathbf{A}_1' & \cdots & \mathbf{A}_p' & \mathbf{D}' \end{bmatrix},$$

where $k = np + h$. Then (1) can be compactly written as $\mathbf{y}_t' \mathbf{A} = \mathbf{x}_t' \mathbf{F} + \varepsilon_t'$. We will refer to $\mathbf{F}$ as lagged parameters even though $\mathbf{F}$ may also contain exogenous parameters.

Now consider linear restrictions on columns of both $\mathbf{A}$ and $\mathbf{F}$. For $1 \le i \le n$, let $\mathbf{a}_i$ be the ith column of $\mathbf{A}$, let $\mathbf{f}_i$ be the ith column of $\mathbf{F}$, let $\mathbf{Q}_i$ be an $n \times n$ matrix of rank q_i, and let $\mathbf{R}_i$ be a $k \times k$ matrix of rank r_i. We assume that $\mathbf{a}_i$ and $\mathbf{f}_i$ satisfy

$$\mathbf{Q}_i \mathbf{a}_i = \mathbf{0} ; \tag{3}$$

$$\mathbf{R}_i \mathbf{f}_i = \mathbf{0} . \tag{4}$$

The restrictions given by (3) are said to be non-degenerate if there exists at least one non-singular matrix $\mathbf{A}$ satisfying them. In this paper we assume that all restrictions are non-degenerate. In the existing work, only contemporaneous exclusion restrictions and lagged block recursion restrictions are employed. The types of linear restrictions represented by (3) and (4) give us a much broader range.

Suppose that $\mathbf{U}_i$ is an $n \times q_i$ matrix such that the columns of $\mathbf{U}_i$ form an orthonormal basis for the null space of $\mathbf{Q}_i$ and that $\mathbf{V}_i$ is a $k \times r_i$ matrix such that the columns of $\mathbf{V}_i$ form an orthonormal basis for the null space of $\mathbf{R}_i$. The columns $\mathbf{a}_i$ and $\mathbf{f}_i$ will satisfy the linear

restrictions (3) and (4) if and only if there exist a $q_i \times 1$ vector $\mathbf{b}_i$ and an $r_i \times 1$ vector $\mathbf{g}_i$ such that

$$\mathbf{a}_i = \mathbf{U}_i \mathbf{b}_i \tag{5}$$

$$\mathbf{f}_i = \mathbf{V}_i \mathbf{g}_i . \tag{6}$$

Thus the free parameter space is the set of all $\mathbf{b}_i$ and $\mathbf{g}_i$. For the rest of this paper, we shall work with the free parameter space with the understanding that the original parameters can be easily recovered via the linear transformations $\mathbf{U}_i$ and $\mathbf{V}_i$. The dimension of the free parameter space is often much smaller than the dimension of the full parameter space, especially when restrictions on lagged parameters are imposed. Thus, simulations will be more efficient even with the additional overhead of transforming the free parameters back into the original parameter space.

3. Posterior density functions

We consider the reference prior of Sims and Zha (1998), which is widely used for simultaneous equations VAR models. Specifically, the prior distribution of $\mathbf{a}_i$ and $\mathbf{f}_i$ takes the form

$$\mathbf{a}_i \sim \mathrm{N}\left(\mathbf{0}, \overline{\mathbf{S}}_i\right) \text{ and } \mathbf{f}_i \big| \mathbf{a}_i \sim \mathrm{N}\left(\overline{\mathbf{P}}_i \mathbf{a}_i, \overline{\mathbf{H}}_i\right),$$

where $\overline{\mathbf{S}}_i$ is an $n \times n$ symmetric, positive definite (SPD) matrix, $\overline{\mathbf{H}}_i$ is a $k \times k$ SPD matrix, and $\overline{\mathbf{P}}_i$ is a $k \times n$ matrix. According to (5) and (6), the distribution of $\mathbf{b}_i$ and $\mathbf{g}_i$ in the free parameter space is given by

$$\mathbf{b}_i \sim \mathrm{N}\left(\mathbf{0}, \widetilde{\mathbf{S}}_i\right) \text{ and } \mathbf{g}_i | \mathbf{b}_i \sim \mathrm{N}\left(\widetilde{\mathbf{P}}_i \mathbf{b}_i, \widetilde{\mathbf{H}}_i\right), \tag{7}$$

where

$$\widetilde{\mathbf{H}}_i = \left(\mathbf{V}_i' \overline{\mathbf{H}}_i^{-1} \mathbf{V}_i\right)^{-1};$$

4

$$\tilde{\mathbf{P}}_i = \tilde{\mathbf{H}}_i \mathbf{V}_i' \overline{\mathbf{H}}_i^{-1} \overline{\mathbf{P}}_i \mathbf{U}_i ;$$

$$\tilde{\mathbf{S}}_i = \left(\mathbf{U}_i' \overline{\mathbf{S}}_i^{-1} \mathbf{U}_i + \mathbf{U}_i' \overline{\mathbf{P}}_i' \overline{\mathbf{H}}_i^{-1} \overline{\mathbf{P}}_i \mathbf{U}_i - \tilde{\mathbf{P}}_i' \tilde{\mathbf{H}}_i^{-1} \tilde{\mathbf{P}}_i \right)^{-1}.$$

Clearly, $\tilde{\mathbf{S}}_i$ is a $q_i \times q_i$ SPD matrix, $\tilde{\mathbf{H}}_i$ is an $r_i \times r_i$ SPD matrix, and $\tilde{\mathbf{P}}_i$ is an $r_i \times q_i$ matrix. If there are no restrictions on the lagged parameters as in Sims and Zha (1998), we have $\tilde{\mathbf{H}}_i = \overline{\mathbf{H}}_i$, $\tilde{\mathbf{P}}_i = \overline{\mathbf{P}}_i \mathbf{U}_i$, and $\tilde{\mathbf{S}}_i = \left(\mathbf{U}_i' \overline{\mathbf{S}}_i^{-1} \mathbf{U}_i \right)^{-1}$.

The great bulk of structural VAR work has focused exclusively on linear restrictions on the contemporaneous parameters **A**. There are many instances, however, in which economics calls for restrictions on lagged relationships in multivariate time-series models. Zha (1999) shows how to derive exact Bayesian inferences when block recursions are imposed on the lagged structure in the VAR system. A block recursion is a special form of linear restrictions imposed on **F**. The following theorem generalizes the result of Zha (1999) under the linear restrictions given by (3) and (4)

Theorem 1. *For $i = 1, \cdots, n$ define*

$$\mathbf{H}_i = \left(\mathbf{V}_i' \mathbf{X}' \mathbf{X} \mathbf{V}_i + \tilde{\mathbf{H}}_i^{-1} \right)^{-1}$$

$$\mathbf{P}_i = \mathbf{H}_i \left(\mathbf{V}_i' \mathbf{X}' \mathbf{Y} \mathbf{U}_i + \tilde{\mathbf{H}}_i^{-1} \tilde{\mathbf{P}}_i \right)$$

$$\mathbf{S}_i = \left(\frac{1}{T} \left(\mathbf{U}_i' \mathbf{Y}' \mathbf{Y} \mathbf{U}_i + \tilde{\mathbf{S}}_i^{-1} + \tilde{\mathbf{P}}_i' \tilde{\mathbf{H}}_i^{-1} \tilde{\mathbf{P}}_i - \mathbf{P}_i' \mathbf{H}_i^{-1} \mathbf{P}_i \right) \right)^{-1}$$

Under the prior on $\mathbf{b}_i$ and $\mathbf{g}_i$ as specified in (7), the joint posterior pdf of $\mathbf{b}_i$ and $\mathbf{g}_i$ is

$$p(\mathbf{b}_i, \cdots, \mathbf{b}_n) \prod_{i=1}^{n} p(\mathbf{g}_i | \mathbf{b}_i)$$

where

$$p(\mathbf{b}_1, \cdots, \mathbf{b}_n) \propto \left| \det[\mathbf{U}_1 \mathbf{b}_1 | \cdots | \mathbf{U}_n \mathbf{b}_n] \right|^{T} \exp\left(-\frac{T}{2} \sum_{i=1}^{n} \mathbf{b}_i' \mathbf{S}_i^{-1} \mathbf{b}_i \right), \tag{8}$$

$$p(\mathbf{g}_i | \mathbf{b}_i) = N(\mathbf{P}_i \mathbf{b}_i, \mathbf{H}_i).\tag{9}$$

Proof. It follows from (2), (5), and (6) that the likelihood function for $\mathbf{b}_1, \cdots, \mathbf{b}_n$ and $\mathbf{g}_1, \cdots, \mathbf{g}_n$ is proportional to

$$\left| \det[\mathbf{U}_1 \mathbf{b}_1 | \cdots | \mathbf{U}_n \mathbf{b}_n] \right|^T \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (\mathbf{b}_i' \mathbf{U}_i' \mathbf{Y}' \mathbf{Y} \mathbf{U}_i \mathbf{b}_i - 2\mathbf{g}_i' \mathbf{V}_i \mathbf{X}' \mathbf{Y} \mathbf{U}_i \mathbf{b}_i + \mathbf{g}_i' \mathbf{V}_i \mathbf{X}' \mathbf{X} \mathbf{V}_i \mathbf{g}_i) \right).\tag{10}$$

Multiplying the likelihood by the prior given by (7) leads to the following posterior density kernel:

$$\left| \det[\mathbf{U}_1 \mathbf{b}_1 | \cdots | \mathbf{U}_n \mathbf{b}_n] \right|^T \exp\left(-\frac{T}{2} \sum_{i=1}^{n} \mathbf{b}_i' \mathbf{S}_i^{-1} \mathbf{b}_i \right) \times \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (\mathbf{g}_i - \mathbf{P}_i \mathbf{b}_i)' \mathbf{H}_i^{-1} (\mathbf{g}_i - \mathbf{P}_i \mathbf{b}_i) \right),$$

which is equivalent to (8) and (9). QED.

It is clear from (10) in the proof of Theorem 1 that if the prior on $\mathbf{A}$ and $\mathbf{F}$ is flat, then the posterior density function (i.e., the likelihood function) can be decomposed to (8) and (9). Thus, Theorem 1 includes Theorem 2 of Zha (1999) as a special case in which the marginal posterior density function (8) can be further decomposed when $\mathbf{A}$ is block recursive. It leads to the feasible calculation of exact finite-sample inferences when identifying restrictions on the lag structure are called for. By feasible we mean that the simulation of the distribution for $\mathbf{F}$ can be done equation by equation. Although this feature is typical in seemly unrelated regressions, it is not typical in many Bayesian vector autoregressions (Sims and Zha 1998). Theorem 1, therefore, provides a powerful result.

When the inferences of lagged and exogenous parameters cannot be decomposed into several manageable calculations, the researcher is likely to suffer the curse of dimensionality. For example, in the 13-lag and 20-variable simultaneous VAR model of Leeper, Sims, and Zha (1996), the dimension of $\mathbf{F}$ is of order 5220. Repeatedly solving least-square problems of this order over thousands of iterations is prohibitively expensive in widely available workstations.

A large system of multiple equations is desirable if one wishes to understand the details of transmission mechanisms through many different sectors in the actual economy. The larger the system becomes, however, the less reliable the estimation and inference are. In such a case, economic restrictions on lagged and exogenous relationships become necessary (Zellner 1985). Theorem 1 generalizes the results in Sims and Zha (1998) and Zha (1999) to allow a much broader set of linear restrictions that may be called for by economic problems. Since the distribution for $\mathbf{F}$ conditional on $\mathbf{A}$ can be calculated one equation at time, Theorem 1 can be readily applied to fairly large systems of multiple equations.

To obtain exact inferences it is necessary to simulate the joint posterior distribution of $\mathbf{A}$ and $\mathbf{F}$. There are two steps. First, simulate draws of $\mathbf{A}$ from the marginal posterior density (8). Second, given each draw of $\mathbf{A}$, simulate draws of $\mathbf{F}$ from the conditional posterior density (9). The second step is straightforward because it requires draws only from multivariate normal distributions. The first step, however, has been a challenging task. The difficulty rises with the simultaneity inherent in the system. To overcome such difficulty, we develop a method for Gibbs sampling of $\mathbf{A}$ in the next section.

4.   Gibbs sampler

4.1.   Theoretic foundation

While the desirable properties for the convergence and efficiency of Gibbs sampling are well documented in Chib and Greenberg (1995) and Geweke (1995), its derivation for a particular problem can be difficult. The following theorem provides a theoretical foundation for our Gibbs sampler. The central result states that drawing from the distribution of $\mathbf{b}_i$ conditional on $\mathbf{b}_1, \cdots, \mathbf{b}_{i-1}, \mathbf{b}_{i+1}, \cdots, \mathbf{b}_n$ is equivalent to drawing independently from a multivariate Gaussian distribution and a special univariate distribution.

Theorem 2. *The random vector* $\mathbf{b}_i$ *conditional on* $\mathbf{b}_1, \cdots, \mathbf{b}_{i-1}, \mathbf{b}_{i+1}, \cdots, \mathbf{b}_n$ *is a linear function of*

q_i *independent random variables* β_j *such that*

(a) the density function of β_1 *is proportional to* $|\beta_1|^T \exp\left(-T\beta_1^2/2\right)$;

(b) for $2 \le j \le q_i$, β_j *is normally distributed with mean zero and variance* $1/T$.

Proof: Since the restrictions are assumed to be non-degenerate, with probability one, the set of

vectors $\left\{\mathbf{U}_j \mathbf{b}_j \mid j \ne i\right\}$ will be linearly independent and if $\mathbf{w}$ is any non-zero vector that is

perpendicular to each vector in $\left\{\mathbf{U}_j \mathbf{b}_j \mid j \ne i\right\}$, then $\mathbf{U}_i' \mathbf{w}$ will also be non-zero. Let $\mathbf{T}_i$ be an

$q_i \times q_i$ matrix such that $\mathbf{T}_i \mathbf{T}_i' = \mathbf{S}_i$. Define

$$\mathbf{w}_1 = \mathbf{T}_i' \mathbf{U}_i' \mathbf{w} \big/ \left\| \mathbf{T}_i' \mathbf{U}_i' \mathbf{w} \right\|. \tag{11}$$

Choose $\mathbf{w}_2, \cdots, \mathbf{w}_{q_i}$ so that $\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_{q_i}$ forms an orthonormal basis for $\mathbf{R}^{q_i}$. By construction,

$\mathbf{U}_i \mathbf{T}_i \mathbf{w}_j$ is perpendicular to $\mathbf{w}$ and so is a linear combination of $\left\{\mathbf{U}_j \mathbf{b}_j \mid j \ne i\right\}$. If $\beta_1, \cdots, \beta_{q_i}$ are

independent random variables whose distributions are given by (a) and (b), define the random

variable $\mathbf{b}_i$ by

$$\mathbf{b}_i = \mathbf{T}_i \sum_{j=1}^{q_i} \beta_j \mathbf{w}_j.$$

The results follow from

$$\left| \det\left[\mathbf{U}_1 \mathbf{b}_1 | \cdots | \mathbf{U}_n \mathbf{b}_n \right] \right|^T \exp\left(-\frac{T}{2} \sum_{i=1}^{n} \mathbf{b}_i' \mathbf{S}_i^{-1} \mathbf{b}_i \right)$$

$$\propto \left| \det\left[\mathbf{U}_1 \mathbf{b}_1 | \cdots | \mathbf{U}_i \mathbf{T}_i \sum_{j=1}^{q_i} \beta_j \mathbf{w}_j | \cdots | \mathbf{U}_n \mathbf{b}_n \right] \right|^T \times \exp\left(-\frac{T}{2} \left(\sum_{j=1}^{q_i} \beta_j \mathbf{w}_j \right)' \left(\sum_{j=1}^{q_i} \beta_j \mathbf{w}_j \right) \right)$$

$$\propto |\beta_1|^T \exp\left(-\frac{T}{2} \beta_1^2 \right) \times \prod_{j=2}^{n} \exp\left(-\frac{T}{2} \beta_j^2 \right).$$

QED.

    The next algorithm lays out the steps for implementing our Gibbs sampler.

Algorithm 1. *The Gibbs sampler involves the following steps.*

(a) *Choose an arbitrary matrix* $\mathbf{A}^{(0)}$ *satisfying the restrictions given by (3) (typically, the estimate at the peak of the posterior density function if available).*

(b) *Given* $\mathbf{A}^{(i-1)} = \left[\mathbf{a}_1^{(i-1)} | \cdots | \mathbf{a}_n^{(i-1)} \right]$, *simulate* $\mathbf{A}^{(i)} = \left[\mathbf{a}_1^{(i)} | \cdots | \mathbf{a}_n^{(i)} \right]$ *by*

    (b.1) simulating $\mathbf{a}_1^{(i)}$ *from the distribution of* $\mathbf{a}_1 | \mathbf{a}_2^{(i-1)}, \cdots, \mathbf{a}_n^{(i-1)}$;

    (b.2) simulating $\mathbf{a}_2^{(i)}$ *from* $\mathbf{a}_2 | \mathbf{a}_1^{(i)}, \mathbf{a}_3^{(i-1)}, \cdots, \mathbf{a}_n^{(i-1)}$;

    $\vdots$

    (b.n) simulating $\mathbf{a}_n^{(i)}$ *from* $\mathbf{a}_n | \mathbf{a}_1^{(i)}, \cdots, \mathbf{a}_{n-1}^{(i)}$.

(c) *Collect the sequence* $\left\{ \mathbf{A}^{(0)}, \cdots, \mathbf{A}^{(N_1+N_2)} \right\}$ *and keep only the last* N_2 *values of the sequence.*

Step (c) of Algorithm 1 concerns a choice of N_1 and N_2. If the initial matrix $\mathbf{A}^{(0)}$ is random but not drawn from the target distribution, the first N_1 draws are usually discarded to protect against an unlikely initial draw. Our experiments indicate that Algorithm 1 quickly (often in 2 or 3 iterations) returns to the nontrivial probability region even if the initial draw was of extremely low probability. The intuition is that the mode of $\mathbf{A}^{(i)}$ given $\mathbf{A}^{(i-1)}$, which can be easily computed via Theorem 2 by setting $\beta_1 = 1$ and $\beta_j = 0$ for $2 \leq j \leq q_i$, quickly converges to the ML estimate. Thus, setting N_1 to 100 is more than adequate for most VAR models.

The choice of sample size (N_2), however, depends on the serial correlation of draws from our Gibbs sampler. In other words, one would like to know the number of independent (effective) draws obtained in the N_2 actual draws. In Section 4.4 we give conditions under which all draws produced by Algorithm 1 are independent. In section 5 we discuss a measure of the effective sample size for the situation where draws may be serially correlated.

9

In step (b) of Algorithm 1, all simulations are carried out according to Theorem 2. Specifically, the random vector $\mathbf{a}_i$ conditional on $\mathbf{a}_1, \cdots, \mathbf{a}_{i-1}, \mathbf{a}_{i+1}, \cdots, \mathbf{a}_n$ is represented as

$$\mathbf{a}_i = \beta_1 \mathbf{U}_i' \mathbf{T}_i^{-1} \mathbf{w}_1 + \sum_{j=2}^{q_i-1} \beta_j \mathbf{U}_i' \mathbf{T}_i^{-1} \mathbf{w}_j . \tag{12}$$

The random variable β_j is normal for $2 \le j \le q_i$, so the second term in (12) is straightforward to simulate. The only computational complication involves the simulation from the less standard distribution of β_1 and the construction of the orthonormal basis $\mathbf{w}_1, \cdots, \mathbf{w}_{q_k}$, whose techniques are now described in the next two sections.

4.2. *Simulating the special univariate distribution*

The probability density function of β_1 discussed in Section 4.1 belongs to the following family of distributions:

$$p(x) = \frac{T^{(k+1)/2}}{2^{(k-1)/2} \Gamma((k+1)/2)} \left(x^2\right)^{k/2} \exp\left(-\frac{T}{2} x^2\right), \text{ for } k, T > 0, \tag{13}$$

where $\Gamma(\cdot)$ is the standard gamma function. Note that when $k = T$, $p(x)$ is exactly the pdf of β_1. To draw x from $p(x)$, let $r = x^2$. The Jacobian of transforming x to r is $1/(2\sqrt{r})$. It follows from (13), that the pdf of r is proportional to

$$r^{(k+1)/2-1} \exp\left(-\frac{Tr}{2}\right). \tag{14}$$

If k is an integer, (14) is the univariate Wishart (UW) density (Zellner 1974, p.389-394). Denote this univariate Wishart by

$$\text{UW}\left(T^{-1}, k+1\right). \tag{15}$$

Note that (14) or (15) is also a Gamma density function with the two parameters being $(k+1)/2$ and $2/T$. The random variable x from the distribution (13) is a square root of the UW variable.

Now, simulating random draws from (13) becomes straightforward. First, draw a vector $z = (z_1, \cdots, z_{k+1})'$ identically and independently from the normal distribution with mean zero and variance $1/T$. Then, form $r = z'z$. The random draw r so obtained is from $\mathrm{UW}(T^{-1}, k+1)$. For each draw r, assign $x = \sqrt{r}$ or $x = -\sqrt{r}$, each with probability one-half. The draw of x so computed is from the probability distribution (13).

4.3. Constructing the orthonormal basis

In this section we provide a stable and computationally efficient method for constructing the orthonormal basis $\mathbf{w}_1, \cdots, \mathbf{w}_{q_i}$ required in Theorem 2 and Algorithm 1. The vector $\mathbf{w}$ is the key to the construction of the $\mathbf{w}_j$. This vector can be found by using Gaussian elimination or the LU decomposition to solve the system $\mathbf{w}'\mathbf{X} = \mathbf{0}$, where

$$\mathbf{X} = \left[\mathbf{a}_1 | \cdots | \mathbf{a}_{i-1} | \mathbf{0} | \mathbf{a}_{i+1} | \cdots | \mathbf{a}_n \right].$$

The vector $\mathbf{w}_1$ can be computed via equation (11). If $\mathbf{w}_1' = \left[w_1, \cdots, w_{q_i} \right]$, then define

$$\mathbf{w}_j' = \left[w_1 w_j, \cdots, w_{j-1} w_j, -c_{j-1}, 0, \cdots, 0 \right] / \sqrt{c_{j-1} c_j},$$

where $c_j = \sum_{k=1}^{j} w_k^2$. A direct computation verifies that $\mathbf{w}_1, \cdots, \mathbf{w}_{q_i}$ is an orthonomal basis. This computation will be stable if it is implicitly arranged so that w_1 is the largest of $w_1, \cdots, w_{q_i}$.

4.4. Independent Decompositions

In this section we establish a general result implying that our Gibbs sampler will produce independent draws of $\mathbf{A}$ if the model is recursive. Clearly, the draws of $\mathbf{A}$ are independent if and only if the distribution of $\mathbf{b}_i$ is independent of $\mathbf{b}_1, \cdots, \mathbf{b}_{i-1}, \mathbf{b}_{i+1}, \cdots, \mathbf{b}_n$, for $1 \le i \le n$. From the proof of Theorem 2, it is easy to see that the distribution of $\mathbf{b}_i$ will be independent of $\mathbf{b}_1, \cdots, \mathbf{b}_{i-1}, \mathbf{b}_{i+1}, \cdots, \mathbf{b}_n$ if and only if the vector $\mathbf{w}_1$ does not depend on $\mathbf{b}_1, \cdots, \mathbf{b}_{i-1}, \mathbf{b}_{i+1}, \cdots, \mathbf{b}_n$. The following theorem gives necessary and sufficient conditions for this result.

Theorem 3. *The distribution of* $\mathbf{b}_i$ *will be independent of* $\mathbf{b}_1,\cdots,\mathbf{b}_{i-1},\mathbf{b}_{i+1},\cdots,\mathbf{b}_n$, *for* $1\le i\le n$, *if there exists a permutation matrix* Π *and an orthogonal matrix* Γ *such that the matrix* $\Gamma\mathbf{A}\Pi$ *is upper triangular for all matrices* $\mathbf{A}$ *satisfying the restrictions given by (3).*

Proof: Assume there exists a permutation matrix Π and an orthogonal matrix Γ such that the matrix $\Gamma\mathbf{A}\Pi$ is upper triangular for all matrices $\mathbf{A}$ satisfying the restrictions given by (3). Denote the jth column of the $n\times n$ identity matrix by $\mathbf{e}_j$. Let $\mathbf{A}$ be any matrix satisfying the restrictions given by (3) and, given i, let $\mathbf{w}$ be a vector such that $\mathbf{w}'\mathbf{A}\mathbf{e}_j =0$ for $j\ne i$. The proof will follow if we show that up to a scalar multiple, the vector $\mathbf{U}_i\mathbf{w}$ does not depend on $\mathbf{A}$. Let π be the permutation associated with Π^2. Since

$$\mathbf{w}'\Gamma'\Gamma\mathbf{A}\Pi\mathbf{e}_{\pi(j)} = \mathbf{w}'\mathbf{A}\mathbf{e}_j = 0, \text{ for } j\ne i,$$

and $\Gamma\mathbf{A}\Pi$ is upper triangular, it follows that the jth coordinate of $\Gamma\mathbf{w}$ is zero for $j<\pi(i)$. Since $\Gamma\mathbf{A}\Pi$ is upper triangular for all matrices $\mathbf{A}$ satisfying the restrictions given by (3), it follows that the column space of $\mathbf{U}_i$ is contained in the span of the first $\pi(i)$ columns of Γ'. Since Γ is orthogonal, $\mathbf{U}_i'\Gamma'\mathbf{e}_j = 0$ for $j>\pi(i)$. Thus $\mathbf{U}_i'\mathbf{w} = \mathbf{U}_i'\Gamma'\Gamma\mathbf{w}$ must be a scalar multiple of $\mathbf{U}_i'\Gamma'\mathbf{e}_{\pi(i)}$, which does not depend on $\mathbf{A}$. QED.

Theorem 3 can be interpreted as saying that if, after an appropriate reordering of the equations and linear transformation of the variables, the restricted matrix $\mathbf{A}$ is upper triangular, then the Gibbs sampler will produce independent draws. When the restrictions are of the exclusion type, Theorem 3 has the following more intuitive interpretation.

---

[2] The permutation π is associated with the permutation matrix Π if, for any matrix $\mathbf{A}$, column i of $\mathbf{A}$ is equal to column $\pi(i)$ of $\mathbf{A}\Pi$.

Corollary 1. *If the restrictions given by (3) are exclusion restrictions and there exists a reordering of the equations and variables such that the resulting system is upper triangular, then* $\mathbf{b}_i$ *will be independent of* $\mathbf{b}_1, \cdots, \mathbf{b}_{i-1}, \mathbf{b}_{i+1}, \cdots, \mathbf{b}_n$, *for* $1 \le i \le n$.

Proof: Note that a reordering of the equations is equivalent to right multiplication by a permutation matrix and a reordering of the variables is equivalent to left multiplication by a permutation matrix. Since permutation matrices are orthogonal, the corollary follows directly from Theorem 3. QED.

Both Theorem 3 and Corollary 1 give sufficient conditions for the independence of draws from the Gibbs sampler. Though we do not prove it, these conditions are also necessary.

5. An empirical example

In this section we apply the Gibbs sampler to an empirical example and compare the result with that produced by the existing importance sampler. The example is Sims (1986) second simultaneous equations model fitted to quarterly data over the sample period 1948:1-1989:3. The six variables are the 3-month Treasure Bill rate (R), M1, real GNP (y), GNP deflator (P), the unemployment rate (U), and gross domestic business investment (I). All variables are in logarithm except the interest rate and the unemployment rate, which are expressed as percentages. The prior we use is in the form of (7), whose specification is detailed in Sims and Zha (1998).[3] For the reader's convenience, the linear restrictions on **A** are presented in Table 1.

---

[3] See also Robertson and Tallman (1999).

Table 1. Sims (1986) Second Identification

	MP	*MD*	*Output*	*Price*	*Unemp*	*ID*
R	X	X	X	X	X	
M1	X	X		X		
Y		X	X	X	X	
P		X		X	X	
U					X	
I		X	X		X	X

The restrictions described in Table 1 are of an exclusion type. Each column represents a behavioral equation labeled at the top of the column. The label "MP" stands for monetary policy, "MD" stands for money demand, "Unemp" unemployment, and "ID" investment demand. A cell filled with "X" implies that the corresponding variable labeled on the left enters the corresponding equation labeled at the top. Empty cells imply zero restrictions. For example, the first column of Table 1 describes the monetary policy equation in which the Fed responds contemporaneously to the interest rate (R) and money (M1), but not to other variables. The justification for other behavioral equations can be found in Sims's original paper. The total number of free parameters in **A** (X's in Table 1) is 20.

To check efficiency of Algorithm 1, we computed for each of the 20 parameters the ratio *B/W* suggested by Gelman et al (1995, pp. 331-333) where *B* and *W* are the "between sequence" and "within sequence" estimates of the variance of a parameter. The results are reported in Table 2. The computation took 1,000 sequences and for each sequence 10,000 MC draws.[4] Increasing the number of sequences or the number of MC draws did not significantly change the estimate of *B/W*, which is evidence that the reported estimates of *B/W* are accurate. A conservative estimate of the effective MC sample size is the number of MC draws divided by

---

[4] The initial 1,000 draws are generated from a normal distribution and the value of N_1 is set to 100.

this ratio (Sims and Zha 1999). The *B/W* value for individual parameters in the Unemp and ID equations is about 1, implying that all MC draws are independent and thus these two equations are recursive to the rest of the system. The worst value for all parameters is about 107, implying that we have at least 1 effective draw from every 107 MC draws. Thus the Gibbs sampler is quite efficient for this model.[5]

Table 2. B/W

	MP	*MD*	*Output*	*Price*	*Unemp*	*ID*
R	74.55	68.80	58.15	100.45	1.01	-
M1	31.78	62.32	-	73.74	-	-
Y	-	38.12	106.77	89.03	0.99	-
P	-	91.56	-	76.20	1.00	-
U	-	-	-	-	0.97	-
I	-	29.03	86.63	-	0.99	0.97

To illustrate the effect of simultaneity on the posterior shape, we focus on the two free parameters in the MP equation (the column headed by "MP" in Table 1), denoted by a_{11} and a_{21}. The joint marginal posterior density function of a_{11} and a_{21} can be examined visually. Figure 1 plots this density function. The shape of the density is quite non-Gaussian, showing a thin, curved ridge in the (a_{11}, a_{21}) plane.[6] The curved ridge can be better seen from the scatter plot displayed in Figure 2. The existing importance sampler approximates such a non-Gaussian shape of the posterior density function by first drawing from the asymptotic Gaussian (or Student *t*) distribution approximated by the second-order Taylor expansion of the logarithm of (8) at the

---

[5] By comparison, the Metropolis sampler developed in an earlier paper of Waggoner and Zha (1997) had *B/W* values 40 to 80 times larger than those reported in Table 2. Furthermore, the efficiency in the Metropolis sampler is sensitive to the choice of jumping kernels and rejection ratios.

[6] Figure 1 was produced using a sequence of 1.2 million draws obtained via Algorithm 1. The initial draw was the estimate of **A** at the peak of the posterior function and the initial 100 draws were discarded. The simulation, coded in C, took approximately 6 minutes on a Pentium 400 MHz desktop.

peak of the posterior density function. Each draw of **A** is then weighted by the ratio of the value of (8) to the value of the approximating pdf. Because of this narrow curved ridge, almost all draws produced from importance sampling using a Gaussian (or *t*-distributional) approximation will lie in the very low probability region and will receive negligible weights. The simulation will produce a few draws that lie on the curved thin ridge with dominant weights. As a result, the importance sampler will produce a long sequence of MC draws that are dominated by only a few draws.

Figure 3a displays the probability density of a_{11} and a_{21}, produced from 1.2 million MC draws from the importance sampling. The importance function used here is a multivariate Student-*t* distribution with three degrees of freedom to take account of possible fat-tails in the distribution.[7] Compared to Figure 1, it can be seen that the approximation is dominated by about 10 draws. We then tripled the number of MC draws to 3.6 million draws and the resulting posterior density is plotted in Figure 3b. Clearly, the approximation is now dominated by a single draw.

When the posterior density function is so non-Gaussian, importance sampling based on any practical number of draws will become highly inaccurate. For our example of 1.2 million draws, the inverse of the largest weight is about 28; for 3.6 million draws, this small number is further reduced to 6. These numbers are so small relative to the corresponding total numbers of draws that the sampler is essentially of no use.

6. Conclusion

A great many economic applications call for simultaneous equations modeling. Exact finite sample inferences, however, become difficult when the model is non-recursive. Indeed, we have shown that the existing importance sampler can be prohibitively inefficient for the simultaneous

---

[7] We have examined a wide range of different degrees of freedom and of different covariance matrices but the result does not improve.

equations type of VAR models. We have developed a simulation method to obtain exact Bayesian inferences for this type of models. The method works for models with both contemporaneous and lagged linear restrictions and generates independent MC draws for recursive models. Because the most expensive part of simulation can be accomplished equation by equation, our method readily applies to estimation and inference in fairly large systems of simultaneous equations.

References

Chib, Siddhartha and Edward Greenberg, 1995. "Understanding the Metropolis-Hastings Algorithm," *The American Statistician 49 (4)*, 327-335.

Doan, Thomas A., 1992. *RATS User's Manual Version 4*. Estima.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin, 1995. *Bayesian Data Analysis*, London: Cahpman & Hall.

Geweke, John, 1995. "Monte Carlo Simulation and Numerical Integration," in H. Amman, D. Kendrick and J. Rust (eds.), *Handbook of Computational Economics*, Amsterdam: North-Holland.

Gordon, David B. and Eric M. Leeper, 1994. "The Dynamic Impacts of Monetary Policy: An Exercise in Tentative Identification," *Journal of Political Economy 102*, 1228-1247.

Christiano, Lawrence J., Martin Eichenbaum, and Charles Evans, 1999. "Monetary Policy Shocks: What Have We Learned and To What End?" in the *Handbook of Macroeconomics* (eds. John Taylor and Michael Woodford), forthcoming.

Highfield, Richard A., 1987. "Forecasting with Bayesian State Space Models," *Ph.D. dissertation*, Graduate School of Business, University of Chicago.

Ingram, Beth F. and Charles H. Whiteman, 1994. "Supplanting the 'Minnesota' prior forecasting macroeconomic time series using real business cycle model priors," *Journal of Monetary Economics 34*, 497-510.

Kim, Soyong and N. Roubini, 1999. "Exchange Rate Anomalies in the Industrial Countries: A Solution with a Structural VAR Approach," *Journal of Monetary Economics*, forthcoming.

Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lutkepohl, Tsoung-Chao Lee, 1985. *The Theory and Practice of Econometrics*, 2nd ed., New York: Wiley.

Leeper, Eric M., Christopher A. Sims, and Tao Zha, 1996. "What Does Monetary Policy Do?" *Brookings Papers On Economic Activity 2*, 1-63.

Robertson, John C. and Ellis W. Tallman, 1999. "Vector Autoregressions: Forecasting and Reality," Federal Reserve Bank of Atlanta *Economic Review* (Second Quarter).

Sims, Christopher A., 1986. "Are Forecasting Models Usable for Policy Analysis," *Quarterly Review* of the Federal Reserve Bank of Minneapolis, Winter.

Sims, Christopher A. and Tao Zha, 1998. "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review 39 (4)*, 949-968.

_____, 1999. "Error Bands for Impulse Responses," *Econometrica 67(5)*, 1113-55.

Uhlig, Harald, 1997. "Bayesian Vector Autoregressions with Stochastic Volatility," *Econometrica 65 (1)*, 59-74.

Waggoner, Daniel F. and Tao Zha, 1997. "Normalization, Probability Distribution, and Impulse Responses," Federal Reserve Bank of Atlanta *Working Paper 97-11*.

Zellner, Arnold, 1962. "An Efficient Method of Estimating seemingly unrelated regressions and tests of aggregation bias," *Journal of the American Statistical Association 57*, 348-368.

_____, 1971. *An Introduction to Bayesian Inference in Econometrics*, New York: John Wiley & Sons, Inc.

_____, 1985. "Bayesian Econometrics," *Econometrica 53*, 253-269.

Zha, Tao, 1999. "Block Recursion and Structural Vector Autoregressions," *Journal of Econometrics 90*, 291-316.
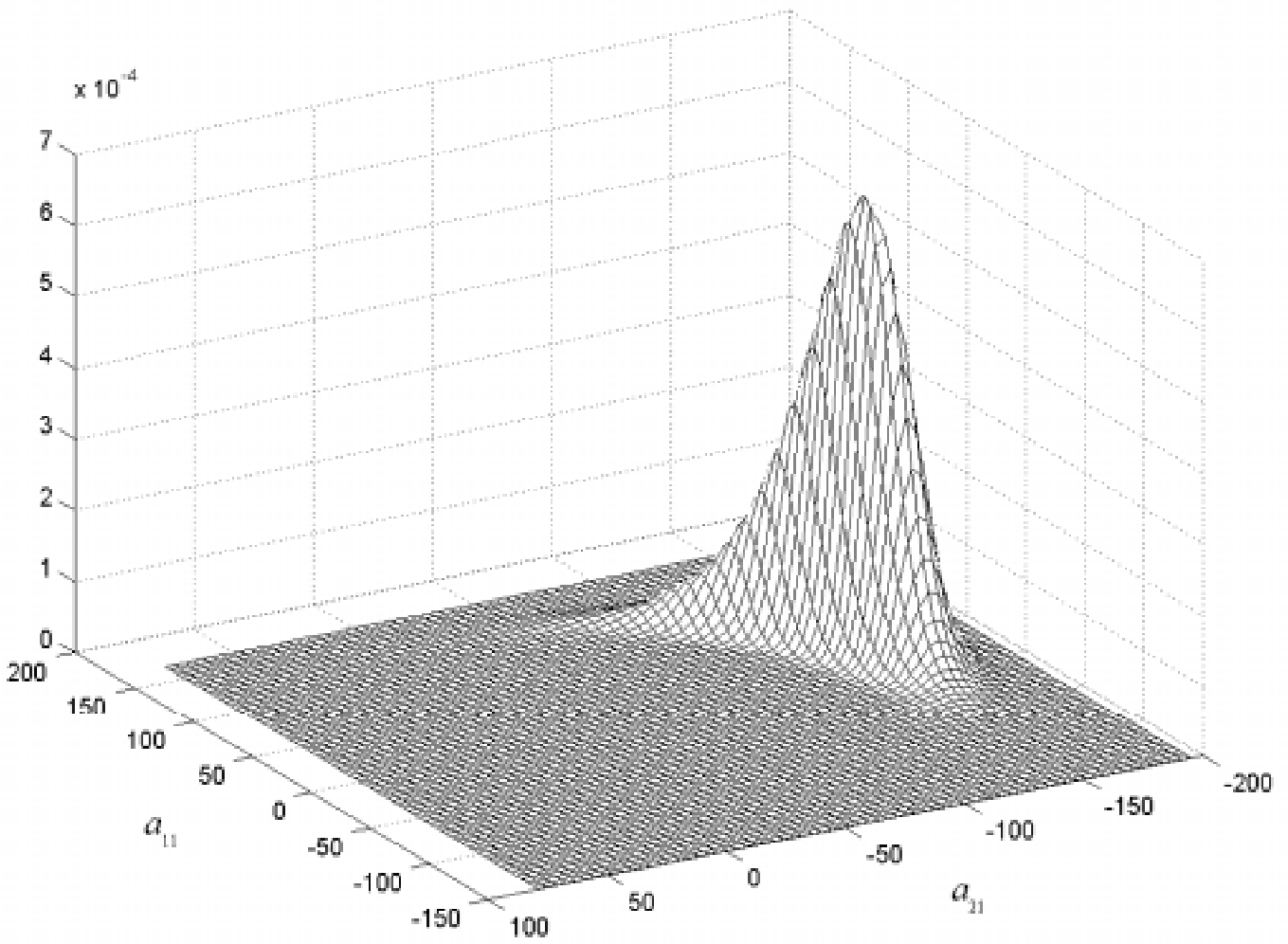
Figure 1. The density simulated from Gibbs sampling

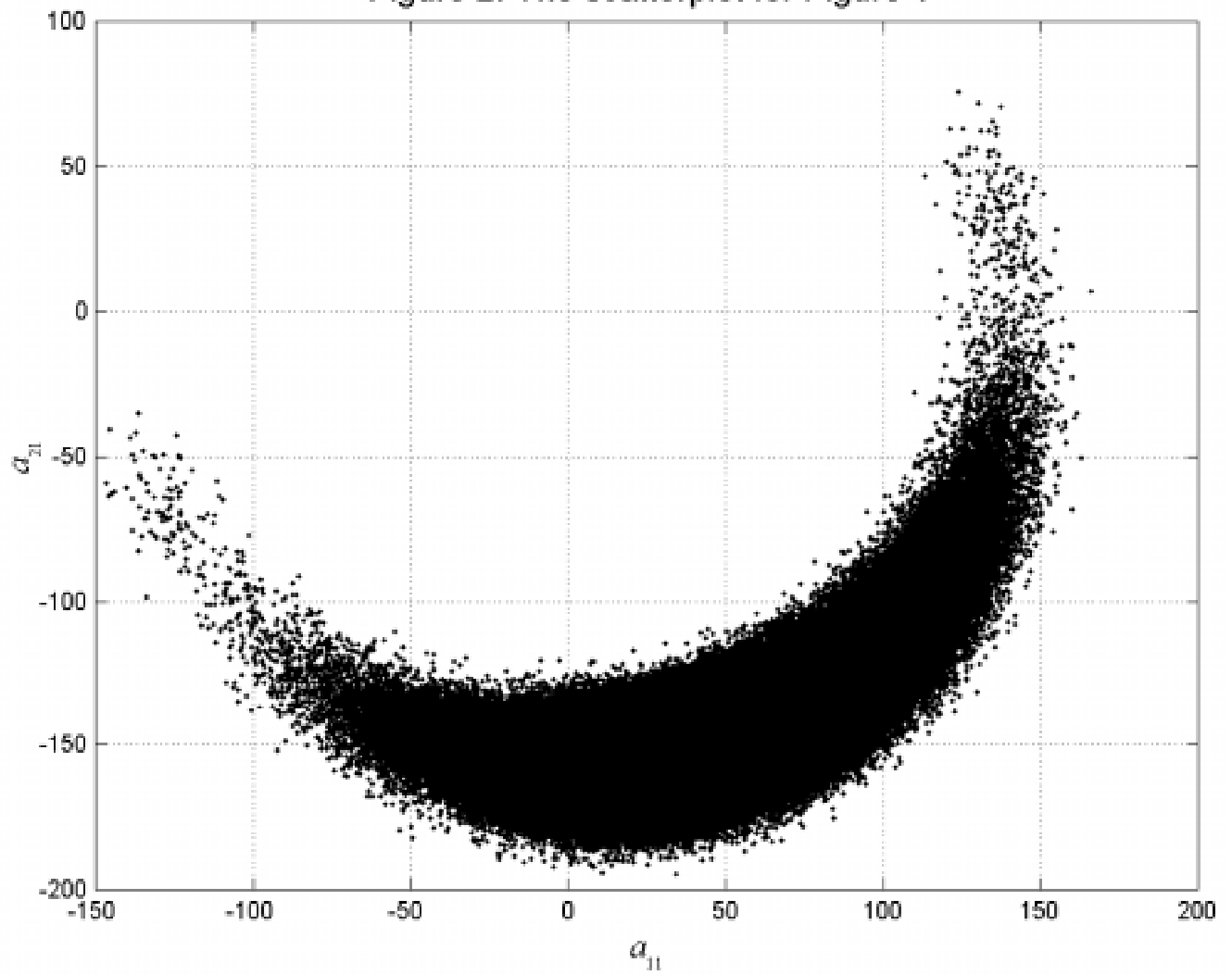Figure 2. The scatterplot for Figure 1
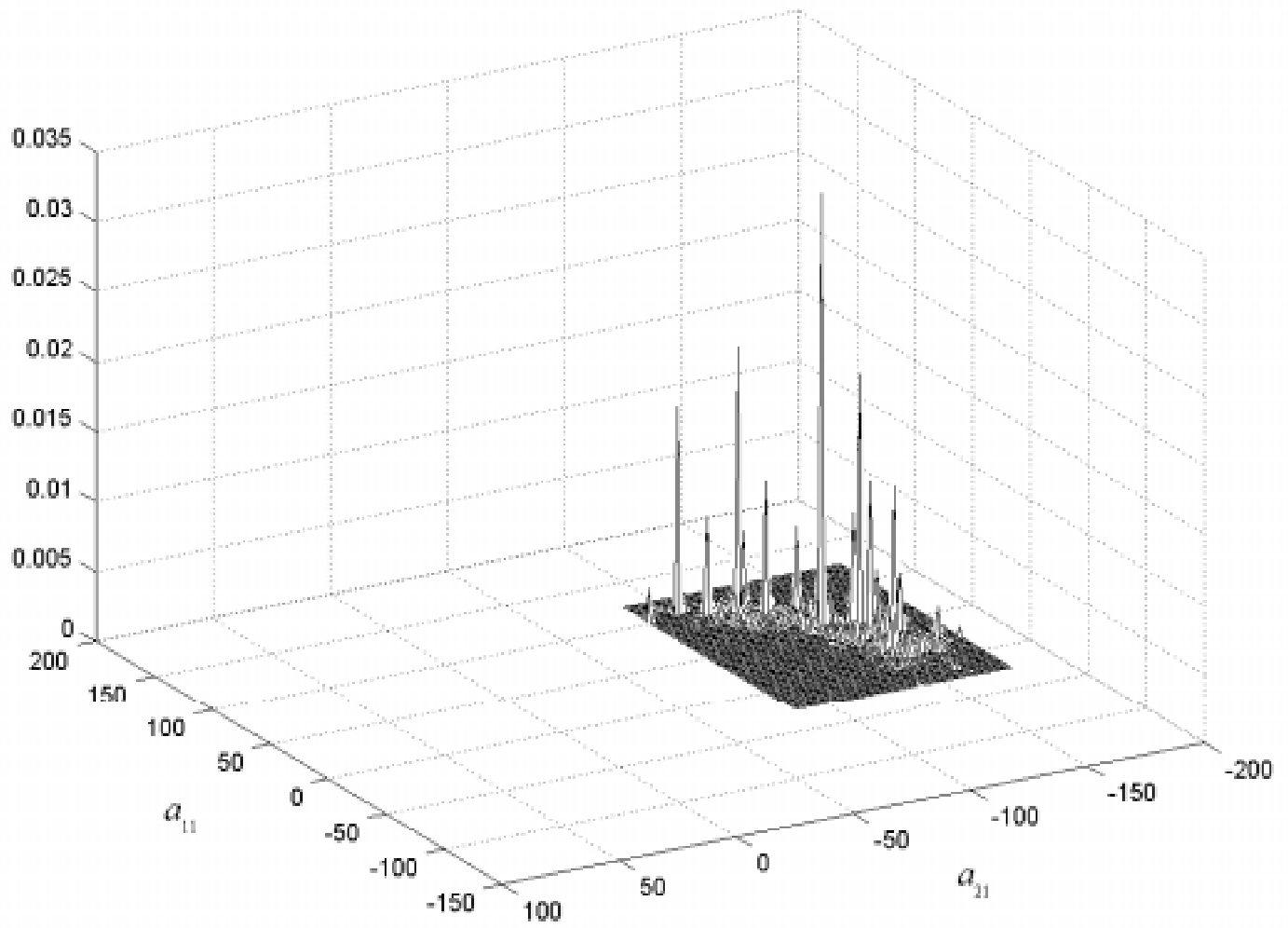
Figure 3a. The density simulated from importance sampling

Figure 3b. The density simulated from importance sampling