Exact small-sample inference in stationary, fully regular, dynamic demand models

Philippe J. Deschamps*

Séminaire d’Économétrie, Université de Fribourg Suisse, Avenue de Beauregard 13, CH-1700 Fribourg, Switzerland

Received 1 June 1998; received in revised form 1 March 1999

Abstract

Asymptotics are known to be unreliable in multivariate models with cross-equation or non-linear restrictions, and the dimension of the problem makes bootstrapping impractical. In this paper, finite sample results are obtained by Markov chain Monte Carlo methods for a nearly non-stationary VAR, and for a differential dynamic demand model with homogeneity, Slutsky symmetry, and negativity. The full likelihood function is used in each case. Slutsky symmetry and negativity are tested using simulation estimates of partial Bayes factors. We argue that a diffuse prior on the long-run error covariance matrix helps to identify the equilibrium coefficients. © 2000 Elsevier Science S.A. All rights reserved.

JEL classification: C11; C15; C32; D12

Keywords: Vector autoregressions; Markov chain Monte Carlo; Exact likelihood; Truncated normal; Allais intensities

1. Introduction

Non-linear functions involving ratios of regression coefficients arise naturally in the context of stationary autoregressions, in the form of unconditional expectations. Similar functions arise in the context of dynamic demand models,
where the parameters of interest are the equilibrium multipliers. Cross-equation restrictions (such as Slutsky symmetry) and further non-linear restrictions (such as negativity) are of importance when demand systems are investigated.

However, even in simple cases, the maximum likelihood estimators of ratios of regression coefficients are known not to always possess finite moments in small samples; see, e.g., Zellner (1978). Furthermore, it is now well recognized that asymptotic approximations are usually misleading in the context of multivariate regression models, especially when cross-equation restrictions are present or when non-linear functions of the parameters are estimated.

The main purpose of this paper is to show that the Bayesian methodology, combined with the use of Markov chain Monte Carlo methods, provides a practical alternative to the usual (point and interval) frequentist estimation of demand models and may avoid these difficulties. The fact that these methods are relatively new may explain the relative rarity of their use in applied demand analysis; an exception is Terrell (1996), where the emphasis is on flexibility rather than global regularity. However, this rarity is unfortunate, since Percy (1992) shows that they are eminently suited to the estimation of multivariate regression models.

This paper discusses the estimation of two related, but somewhat different, dynamic demand models. The first one, in Sections 2–6, is a vector autoregression based on an assumption that the expenditure shares are logistic normal. Whereas this model is quite restrictive from a demand-theoretic standpoint, it is nevertheless useful for the dynamic prediction of budget shares on the unit simplex.

It will also serve as a convenient means of introducing, in a relatively simple context, most of the statistical methodology that will be used later on to investigate a much more sophisticated demand system. Estimation results based on two different diffuse priors will be presented, and the comparison used to justify the choice of prior in the more sophisticated model.

In addition, the pure vector autoregression is an interesting model in its own right, due to its popularity in the empirical literature. The dependent variables in our example appear to be stationary (implying the existence of unconditional expectations), but close to non-stationarity. This makes the estimation problem non-trivial; indeed, the risk exists that such a system is improperly classified as cointegrated, leading to improper reduced rank restrictions and dynamic misspecification.

The second model, discussed in Sections 7–9 and in Appendices C and D, is a differential dynamic demand system with all the regularity restrictions of economic theory (homogeneity, Slutsky symmetry, and negativity) imposed. Dynamics are allowed for by reparameterizing a VARX model in terms of the equilibrium coefficients; the resulting model is one of partial adjustment on the growth rates, and can be justified by minimizing the sum of an adjustment
cost and a disequilibrium cost (Bewley, 1986). The full and conditional maximum likelihood estimates of the differential system with Slutsky symmetry will be contrasted with the Bayesian small-sample estimates. In the fully regular model, the substitution relations between pairs of commodities will be analyzed, using the very different concepts of Hicks and Allais substitutability.

For reasons of practicality and convenience, the tests of dynamic specification in this paper will all be based on classical (non-Bayesian) methods. Indeed, all our models involve a large number of parameters, so that the Monte Carlo estimation of Bayes factors is very time-consuming; on the other hand, easily implemented small-sample versions of most classical dynamic specification tests are available. The situation is markedly different in the case of regularity tests. No small-sample corrections are available for the classical tests of Slutsky symmetry when dynamics are present. The same can be said of the Wald test of negativity proposed by Kodde and Palm (1987), even in the static case; moreover, its exact asymptotic critical values are difficult to determine, leading the authors to propose lower and upper bounds. For these reasons, Slutsky symmetry and negativity will be tested by means of simulation estimates of partial Bayes factors, using a methodology recently proposed by Chib and Greenberg (1998).

The practicality of the proposed method in arbitrary samples requires specialized simulation tools, which are presented in Appendices A and B. Appendix A discusses the generation of autoregressive coefficient matrices with truncated normal distributions on the stationarity region. Appendix B discusses the generation of truncated normal negative-definite matrices.

In this paper, $D_{a}$ denotes the $a^{2} \times a(a + 1)/2$ duplication matrix, and $D_{a}^{-}$ its Moore–Penrose inverse (Magnus and Neudecker, 1988, p. 49). $I_{a}$ denotes an identity matrix of order $a$, and $O_{a \times \beta}$ a null matrix with $a$ rows and $\beta$ columns.

$t_{a}$ is an $a \times 1$ vector of ones. $\rho(A)$ is the spectral radius of a square matrix $A$, i.e. the largest modulus of the eigenvalues of A. For a matrix $\Phi$ and a set $S$, $I_{S}(\Phi)$ is an indicator function which is unity if $\Phi \in S$, and 0 otherwise.

$f_{\text{NO}}(m, V)$ is the multivariate normal density function with expectation vector $m$ and covariance matrix $V$. $f_{\text{IW}}(T + n + 1, S)$ is the inverted Wishart density with $T + n + 1$ degrees of freedom and $n \times n$ symmetric scale matrix $S$; it is the density of the inverse of a matrix which has the Wishart distribution with $T$ degrees of freedom and scale matrix $S^{-1}$ (see Press, 1972, p. 110). $f_{\text{ST}}(m, V, v)$ is the multivariate Student density with location vector $m$, scale matrix $V$, and $v$ degrees of freedom; it is the density of $z/w$, where $z$ is $N(m, V)$ and where $w = \sqrt{\chi^{2}/v}$, with $\chi^{2}$ an independent Chi-square variate with $v$ degrees of freedom (see Muirhead, 1982, p. 33). Finally, we will adopt the convention that a matrix with a row or column dimension of zero is an empty matrix.
2. Stationary vector autoregressions

Consider a sample \((y_1, y_2, \ldots, y_{p+T})\) from the following stationary discrete-time \(n \times 1\) vector process:

\[
\Phi(L)y_t = \mu + \varepsilon_t, \quad (2.1)
\]

where \(\Phi(L) = I_n - \sum_{j=1}^{p} \Phi_j L^j\), \(L\) is the lag operator, \(\Phi_j\) is an \(n \times n\) matrix, and, conditionally on \((y_{t-1}, y_{t-2}, \ldots)\), \(\varepsilon_t\) is an \(n \times 1\) vector with non-degenerate normal distribution \(N(0, \Omega)\). The stationarity condition is conveniently imposed by modelling the sample observations as follows:

\[
y_t = \pi + u_t \quad \text{for } t = 1, \ldots, p, \quad (2.2)
\]

\[
\Phi(L)y_t = \Phi(1)\pi + \varepsilon_t \quad \text{for } t = p + 1, \ldots, p + T, \quad (2.3)
\]

where

\[
\pi = \Phi_{-1}(1)\mu, \quad (2.4)
\]

\[
u = \begin{bmatrix} u_1 \\ \vdots \\ u_p \end{bmatrix} \sim N(0, \Sigma), \quad (2.5)
\]

\[
\text{vec} \Sigma = [I_{n^2} - F(\Phi) \otimes F(\Phi)]^{-1} D_{np} \left( \begin{array}{c} O_{\ell \times m} \\ I_m \end{array} \right) D_n^+ \text{vec} \Omega, \quad (2.6)
\]

\[
F(\Phi) = \begin{pmatrix} O_{n(p-1) \times n} & I_{n(p-1)} \end{pmatrix}, \quad (2.7)
\]

\[
\Phi = (\Phi_p \ldots \Phi_1), \quad (2.8)
\]

with \(\ell = [np(np + 1) - n(n + 1)]/2\) and \(m = n(n + 1)/2\); see Deschamps (1997).

It is noteworthy that (2.2) and (2.3) may be rewritten as an Aitken regression equation with coefficient vector \(\pi\), and that (2.3) may be expressed as a classical multivariate regression model with coefficient matrix \(\Phi\). Indeed (2.2) and (2.3) imply that \(y_o = X_o \pi + v\) with \(v \sim N(0, V)\), where

\[
y_o = \begin{bmatrix} y_1 \\ \vdots \\ y_p \\ \Phi(L)y_{p+1} \\ \vdots \\ \Phi(L)y_{p+T} \end{bmatrix}, \quad (2.9)
\]
\[
X_n = \begin{pmatrix}
I_p \otimes I_n \\
I_T \otimes \Phi(1)
\end{pmatrix},
\]
\[\tag{2.10}\]

\[
v = \begin{pmatrix}
u_1 \\
\vdots \\
u_p \\
\varepsilon_{p+1} \\
\vdots \\
\varepsilon_{p+T}
\end{pmatrix},
\]
\[\tag{2.11}\]

\[
V = \left( \sum_{n=1}^{np} O_{np \times nT} \right)_{nT \times np} I_T \otimes \Omega.
\]
\[\tag{2.12}\]

On the other hand, upon noting that
\[
\Phi(L)y_t - \mu = (y_t - \pi) - [I_n - \Phi(L)](y_t - \pi) + \Phi(1)\pi - \mu
\]
\[
= (y_t - \pi) - \sum_{j=1}^{p} \Phi_j(y_{t-j} - \pi),
\]

Eq. (2.3) can be written as \( Y = \Phi X + E \), with
\[
Y = (y_{p+1} - \pi \ y_{p+2} - \pi \ \cdots \ y_{p+T} - \pi) \quad [n \times T],
\]
\[\tag{2.13}\]

\[
X = \begin{pmatrix}
y_1 - \pi \\
y_2 - \pi \\
\vdots \\
y_p - \pi \\
y_{p+1} - \pi \ y_{p+2} - \pi \ \cdots \ y_{p+T-1} - \pi
\end{pmatrix} \quad [np \times T],
\]
\[\tag{2.14}\]

\[
E = (\varepsilon_{p+1} \ \varepsilon_{p+2} \ \cdots \ \varepsilon_{p+T}) \quad [n \times T].
\]
\[\tag{2.15}\]

\[
\text{vec } E \sim \text{N}(0, I_T \otimes \Omega).
\]
\[\tag{2.16}\]

3. Likelihood, priors, and posteriors

Upon defining \( y_0 = (y'_1, \ldots, y'_p) \) and \( \Psi = (y'_1, \ldots, y'_{p+T}) \), we write the likelihood corresponding to (2.2) and (2.3) as
\[
\mathcal{L}(\Psi | \pi, \Phi, \Omega) \propto \mathcal{L}_1(\pi, \Phi, \Omega) \mathcal{L}_2(\pi, \Phi, \Omega)
\]
\[\tag{3.1}\]
with
\[ L_1(\pi, \Phi, \Omega) = (\det \Sigma)^{-1/2} \exp \left[ -\frac{1}{2} (y_0 - \mathbf{1}_p \otimes \pi)' \Sigma^{-1} (y_0 - \mathbf{1}_p \otimes \pi) \right], \tag{3.2} \]
\[ L_2(\pi, \Phi, \Omega) = (\det \Omega)^{-T/2} \]
\[ \exp \left[ -\frac{1}{2} \sum_{t=p+1}^{p+T} \left[ \Phi(L)y_t - \Phi(1)\pi \right]' \Omega^{-1} \left[ \Phi(L)y_t - \Phi(1)\pi \right] \right] \tag{3.3} \]
and with \( \Sigma \) as in (2.6)–(2.8).

In this paper, two prior densities on \((\pi, \Phi, \Omega)\) will be considered. The first choice assumes the independence of \(\pi, \Phi, \) and \(X\) and the usual diffuse priors on \(\pi\) and \(\Omega\). The prior on \(\Phi\) is an improper uniform density on the following (unbounded) set:
\[ S = \{ \Phi \mid \rho[F(\Phi)] < \varepsilon \}, \tag{3.4} \]
where \(\varepsilon\) is either unity, or a number slightly less than one. This type of prior has been used by Chib and Greenberg (1995b) in a different context. So, our first prior is
\[ p_1(\pi, \Phi, \Omega) \propto I_S(\Phi)(\det \Omega)^{-(n+1)/2} \tag{3.5} \]
implying as joint posterior:
\[ p_1(\pi, \Phi, \Omega \mid \mathcal{F}) \propto I_S(\Phi)(\det \Omega)^{-(n+1)/2} L_1(\pi, \Phi, \Omega)L_2(\pi, \Phi, \Omega). \tag{3.6} \]

The second choice does not assume the prior independence of \(\Phi\) and \(\Omega\); it is related to the minimum information prior density on \((\pi, \Phi, \Omega)\), which can easily be shown to be
\[ p_2(\pi, \Phi, \Omega) \propto I_S(\Phi)[\det \Theta(\Phi, \Omega)]^{-1/2} \tag{3.7} \]
with \(\Theta(\Phi, \Omega)\) equal to the (constant) \(n \times n\) diagonal blocks of \(\Sigma\) in (2.6). This is a straightforward multivariate extension of the argument in Zellner (1971, p. 220). In order to obtain the usual prior in the static case (where \(\Phi = \mathbf{O}\) and \(\Theta = \Omega\)), we modify \(p_2\) by increasing slightly the number of degrees of freedom. We thus take
\[ p_2(\pi, \Phi, \Omega) \propto I_S(\Phi)[\det \Theta(\Phi, \Omega)]^{-(n+1)/2} \tag{3.8} \]
so that the prior information on \(\Theta^{-1}\) can be characterized by a limit of proper prior densities. This second choice implies as joint posterior:
\[ p_2(\pi, \Phi, \Omega \mid \mathcal{F}) \propto I_S(\Phi)[\det \Theta(\Phi, \Omega)]^{-(n+1)/2} L_1(\pi, \Phi, \Omega)L_2(\pi, \Phi, \Omega). \tag{3.9} \]

4. Markov chain Monte Carlo

The method we will use in order to simulate (3.6) and (3.9) is a version of the Metropolis–Hastings algorithm, which has now been extensively described in
the literature; see, e.g., Tierney (1994) and Chib and Greenberg (1995a). In view of (3.1) and of our restatement of (2.2) and (2.3) as an Aitken regression equation, the full conditional posteriors of \( n \) have the form:

\[
p(n \mid D, U, X, Y) = f_{NO}[(X^*_V \mu^{-1} X^*_V)^{-1}(X^*_V \mu^{-1} y^*_V), (X^*_V \mu^{-1} X^*_V)^{-1}]. \tag{4.1}
\]

Similarly, our restatement of (2.3) as a multivariate regression equation implies:

\[
p_1(\Omega \mid \pi, \Phi, \Psi) \propto L_1(\pi, \Phi, \Omega)f_{IW}[T + n + 1, (Y - \Phi X)(Y - \Phi X)^T], \tag{4.2}
\]

\[
p_2(\Omega \mid \pi, \Phi, \Psi) \propto L_1(\pi, \Phi, \Omega)[\det \Theta(\Phi, \Omega)]^{-(n + 1)/2}f_{IW}[T, (Y - \Phi X)(Y - \Phi X)^T], \tag{4.3}
\]

\[
p_1(\text{vec} \Phi \mid \pi, \Omega, \Psi) \propto L_1(\pi, \Phi, \Omega)I_S(\Phi)\]

\[
f_{NO}[\text{vec} YX'(XX')^{-1}, (XX')^{-1} \otimes \Omega], \tag{4.4}
\]

\[
p_2(\text{vec} \Phi \mid \pi, \Omega, \Psi) \propto L_1(\pi, \Phi, \Omega)[\det \Theta(\Phi, \Omega)]^{-(n + 1)/2}I_S(\Phi)\]

\[
f_{NO}[\text{vec} YX'(XX')^{-1}, (XX')^{-1} \otimes \Omega]. \tag{4.5}
\]

Eq. (4.1) is a standard multinormal density. In (4.2) and (4.3), we have products of inverted Wisharts and partial kernels; this situation can be dealt with by generating candidates from \( f_{IW} \) in a Metropolis–Hastings rejection step (Chib and Greenberg, 1994 and 1995a). In (4.4) and (4.5), we have the additional difficulty that the candidate-generating density is truncated normal. It is well known that naive rejection sampling for truncated normal variates is not always practical, since the acceptance probability can become extremely small. As shown by Geweke (1991) and Robert (1995), Gibbs sampling is an attractive alternative. However, the fact that the truncation domain \( S \) cannot be described by a system of linear inequalities, and is not even a convex set, causes additional difficulties in our case. An answer to this problem is presented in Appendix A.

The Metropolis–Hastings rejection steps exacerbate the dependence of the Markov chain properties on the starting values \( (\Phi^{(0)}, \pi^{(0)}, \Omega^{(0)}) \), which have to be chosen carefully. Our proposal for this choice is presented in Section 5.

For now, we present the details of the proposed algorithm. In order to simulate (3.6) or (3.9), we generate \( N \) independent chains \( \{\Phi^{(i)}, \pi^{(i)}, \Omega^{(i)}, i = 1, \ldots, L_1\}^N_{i=1} \) and \( N \) additional independent chains \( \{\Phi^{(i)}, \pi^{(i)}, \Omega^{(i)}, i = 1, \ldots, L_2\}^{2N}_{j=N+1} \), with \( L_2 > L_1 \). The last passes of the 2\( N \) chains should yield an approximately i.i.d. sample from the joint posterior density. Convergence is checked by a robust Wald equality test on the expectation vectors of the two groups of generated variates, using the method described by Andrews and Monahan (1992) with a Parzen kernel; this is complemented by Wilcoxon rank-sum tests on all the univariate marginal distributions (Kendall and Stuart,
In the case of structural instability, $L_1$ and $L_2$ are increased and the $2N$ chains are restarted from their previous terminal passes. We define:

$$\mathcal{L}_*=\mathcal{L}_1(\pi, \Phi, \Omega)$$

if $p(\pi, \Phi, \Omega) = p_1(\pi, \Phi, \Omega)$,

$$= \mathcal{L}_1(\pi, \Phi, \Omega)[\det\Theta(\Phi, \Omega)]^{-(n+1)/2}$$

if $p(\pi, \Phi, \Omega) = p_2(\pi, \Phi, \Omega)$.

Each chain is generated as follows:

**Step 1.** Set $i = 1$.

**Step 2.** Sample $\pi^{(i)}$ from $p(\pi | \Phi^{(i-1)}, \Omega^{(i-1)}, \mathcal{Y})$ in (4.1).

**Step 3.** Compute $Y$ and $X$ according to (2.13) and (2.14) with $\pi = \pi^{(i)}$.

**Step 4.** Using naive rejection sampling (when practical) or the Gibbs sampling algorithm of Appendix A, draw a candidate $\Phi$ from:

$$f(\text{vec } \Phi) \propto f_{\text{NO}}[\text{vec } YX'(XX')^{-1}, (XX')^{-1} \otimes \Omega^{(i-1)}]I_S(\Phi),$$

set $\Phi^{(i)} = \Phi$ with probability:

$$\alpha = \min\left[\frac{\mathcal{L}_*(\pi^{(i)}, \Phi^{(i)}; \Omega^{(i-1)})}{\mathcal{L}_*(\pi^{(i)}, \Phi^{(i-1)}; \Omega^{(i-1)})}, 1\right]$$

and set $\Phi^{(i)} = \Phi^{(i-1)}$ with probability $1 - \alpha$.

**Step 5.** Draw a candidate $\Omega$ from

$$f_{iw}[v, (Y - \Phi^{(i)}X)(Y - \Phi^{(i)}X)],$$

where $v = T + n + 1$ or $v = T$ according to the choice of prior. Set $\Omega^{(i)} = \Omega$ with probability:

$$\beta = \min\left[\frac{\mathcal{L}_*(\pi^{(i)}, \Phi^{(i)}, \Omega)}{\mathcal{L}_*(\pi^{(i)}, \Phi^{(i-1)}, \Omega^{(i-1)})}, 1\right]$$

and set $\Omega^{(i)} = \Omega^{(i-1)}$ with probability $1 - \beta$.

**Step 6.** Set $i = i + 1$ and return to Step 2.

It will be seen in Appendix A that $S$ is arc-connected. Conditions similar to Proposition 2 of Chib and Greenberg (1994) will then be verified if $\mathcal{L}_*$ is bounded, or, equivalently, if $\Sigma$ in (2.6) is positive definite. This will be the case when $\varepsilon < 1$ in (3.4), since $\Omega$ is positive definite and since $\Phi \in S$. Indeed, it is well-known that $\Sigma$ is non-negative definite for stationary VARs (Lütkepohl, 1991, p. 25), and a singular $\Sigma$ would imply an exact linear relation between some elements of $(y_{t-p}', y_{t-p+1}', \ldots, y_t')$, contradicting the non-degeneracy of the disturbances in (2.1). The results in Tierney (1994) and Roberts and Smith (1994)
then imply that the Markov chains converge to the target distribution as $L_1 \to \infty$.

5. Choice of the starting values

The algorithm of Section 4 must be initialized at appropriate starting values. In this section, we propose to choose these starting values randomly, from a thick-tailed approximation to the joint posterior density of $(\pi, \Phi, \Omega)$. This approximation is constructed from the asymptotic distribution of the maximum likelihood estimates. A similar proposal was made by Koop et al. (1994).

Let $\hat{\Phi}(L)$ and $\hat{\mu}$ be the OLS estimates of $\Phi(L)$ and $\mu$ in (2.1), and let

\[
\hat{\pi} = \hat{\Phi}^{-1}(1)\hat{\mu},
\]
\[
\hat{e}_t = \hat{\Phi}(L)y_t - \hat{\mu},
\]
\[
\hat{\Omega} = \frac{1}{T - np - 1} \sum_{t=p+1}^{p+T} \hat{e}_t\hat{e}_t'.
\]

If $\hat{\Phi}$ is the OLS estimate of $\Phi$ in (2.8), standard results on maximum likelihood estimation give the following asymptotic approximation to the sampling distribution of $(\hat{\pi}, \hat{\Phi})$:

\[
f(\hat{\pi}, \text{vec} \hat{\Phi}) \approx f_{\text{NO}}\left(\left(\begin{array}{c} \pi \\ \text{vec} \Phi \end{array}\right), \hat{\mathcal{J}}^{-1}\right),
\]

where

\[
\hat{\mathcal{J}} = \sum_{t=p+1}^{p+T} \left(\hat{\Phi}(1)' Z_t \hat{\Omega}^{-1}(\hat{\Phi}(1) Z_t)\right)
\]

and

\[
Z_t = \begin{bmatrix} y_{t-p} - \hat{\pi} \\ \vdots \\ y_{t-1} - \hat{\pi} \\ y_t - \hat{\pi} \end{bmatrix} \otimes I_n.
\]

Furthermore, $\hat{\Omega}$ is asymptotically independent of $(\hat{\pi}, \hat{\Phi})$, and, from Zellner (1971, p. 227), an approximate marginal posterior distribution of $\Omega$ is Wishart inverse with $T - np + n$ degrees of freedom. The required starting values $\pi^{(0)}$, $\Phi^{(0)}$, and $\Omega^{(0)}$ are then random drawings from the following densities:

\[
p(\pi, \text{vec} \Phi) \propto f_{\text{ST}}\left(\left(\begin{array}{c} \hat{\pi} \\ \text{vec} \hat{\Phi} \end{array}\right), \hat{\mathcal{J}}^{-1}, 3\right) I_s(\Phi),
\]
\[
p(\Omega) = f_{\text{IW}}\left(T - np + n, \hat{\Omega}, \sum_{t=p+1}^{p+T} \hat{e}_t\hat{e}_t'\right).
\]
The truncated Student density in (5.5) is most conveniently sampled by a naive rejection procedure, since it is usually the case that \( \hat{\phi} \in S \).

6. A first comparative illustration

In this section, we will compare the Bayesian stationarity constrained and the OLS estimates in a four-equation VAR of order two, fitted to British data. Specifically, we consider Eq. (2.1) with \( p = 2 \), and the following endogenous variables:

\[
y_{it} = \ln \left( \frac{w_{it}}{w_{5t}} \right) \quad \text{for } i = 1, \ldots, 4 \quad \text{and} \quad t = 1, \ldots, 44,
\]

where \( w_{it} \) is the expenditure share for food \((i = 1)\), alcoholic drink and tobacco \((i = 2)\), clothing and footwear \((i = 3)\), energy products \((i = 4)\), and other goods excluding durables, housing services, and other services \((i = 5)\), out of the total expenditure for these five categories. The data are annual, cover the period 1952–1995, and are taken from the Economic Trends Annual Supplement (1996/1997 edition), published by the British Central Statistical Office.

Eq. (6.1) is of course the additive logistic transformation, which has been extensively studied by Aitchison (1986). The inverse transformation is given by

\[
w_{it} = \frac{\exp(y_{it})}{1 + \sum_{j=1}^{4} \exp(y_{jt})} \quad \text{for } i = 1, \ldots, 4,
\]

\[
w_{5t} = \frac{1}{1 + \sum_{j=1}^{4} \exp(y_{jt})}
\]

and it is straightforward to check that the polynomial \( \Phi(L) \) in (2.1), as well as the joint distribution of the shares in (6.2), are invariant with respect to the choice of normalization in (6.1).

It is also of some interest to note that the model resulting from (2.2) and (6.1) is consistent with utility theory, and in fact embodies a naive version of the Random Utility Hypothesis (see, e.g., Brown and Walker, 1989). Indeed, if the inverse logistic transformation (6.2) is applied to the \( y_{it} = \pi_{it} + u_{it} \) in Eq. (2.2), we may view the transforms as random, time-varying, and autocorrelated elasticities in a Cobb–Douglas utility function.

An estimation of the model by classical methods suggested that the data generating process is close to non-stationarity, but did not provide very strong evidence for trends in the joint process generating \( y_t \). Deterministic trends were not significant; an estimation of (2.1) by ordinary least squares gave a spectral radius of \( \rho[F(\hat{\phi})] = 0.956 \). Dickey–Fuller tests on the individual variables in
did not reject unit roots; however, the relevant issue here is one of joint stationarity, and it may be argued that cointegration tests have more power in this context. The Johansen trace test did not reject the hypothesis that the cointegrating rank \( r \) is strictly less than 4; however, both the Akaike information criterion and the Hannan–Quinn criterion favor \( r = 4 \), i.e. the absence of unit roots. Since little is known about the properties of the Johansen test in small samples, it seems therefore safe to consider the evidence as inconclusive and to investigate whether the unconditional expectation vector \( \pi \) can be reliably estimated by the Bayesian methodology of the previous sections. If \( r < 4 \), \( \pi \) does not exist, and the sampler cannot be expected to converge unless the prior parameter \( \varepsilon \) in (3.4) is strictly less than one (in this case, most of the posterior mass will be concentrated near \( \varepsilon \)).

Using the methodology of Section 4, 10,000 Markov chains of length 400 were generated with \( \varepsilon = 1 \) and the prior in (3.5). The length of 400 was chosen after convergence diagnostics with \( L_1 = 300 \) and \( L_2 = 400 \) passes gave conclusive results, the critical probabilities being all in excess of 0.05 and the robust Wald test yielding a critical probability of 0.85. Table 1 presents summaries of some estimated posterior distributions; for comparison purposes, the OLS estimates, some characteristics of their estimated asymptotic distributions (based on (5.4) in the case of the \( \hat{\pi}_i \)), and some characteristics of their estimated small-sample distributions (based on a parametric bootstrap) are also presented. In Table 1, \( m_i(\theta) \) denotes the \( i \)th sample moment of \( \theta \) and \( \theta_a \) denotes the \( a \)th percentage point of its distribution; NSE denotes the numerical standard error of the posterior expectation estimate \( m_1(\theta) \), computed by the method of Andrews and Monahan (1992) with a Parzen kernel and VAR(1) prewhitening; and RNE denotes its relative numerical efficiency, as defined by Geweke (1992).

The estimated Bayesian posterior distributions of the \( \mu_i \) appear to be closer to normality than the bootstrap distributions of the \( \hat{\mu}_i \), as indicated by the skewness and excess kurtosis statistics. On the other hand, the Bayesian, asymptotic, and bootstrap estimated standard errors of the \( \mu_i \) and \( \hat{\mu}_i \) are rather close.

By contrast, the last four columns of Table 1 reveal large differences between the three methods in the case of the unconditional expectation estimates. The posterior distributions of the \( \pi_i \) are very far from normality; kernel estimation reveals densities that are close to zero over a wide interval, except for a sharp central peak. The situation is similar, but much more extreme, when the bootstrap densities of the OLS estimates \( \hat{\pi}_i \) are considered: the bootstrap confidence intervals are too wide to be useful, and the magnitude of the higher-order sample moments (reported for comparison purposes only) indeed casts doubts on the existence of the corresponding population moments. The asymptotic approximations turn out to be very poor. The Bayesian point estimates \( m_1(\pi_i) \) are quite different from the OLS estimates \( \hat{\pi}_i \); they are, however, closer than the OLS estimates to the sample means \( \bar{y}_1 = 0.587, \bar{y}_2 = 0.098, \bar{y}_3 = -0.262, \) and \( \bar{y}_4 = -0.427 \).
Table 1  
Bayesian (prior 3.5) and OLS estimation of logistic normal VAR

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$\mu_1$</th>
<th>$\mu_2$</th>
<th>$\mu_3$</th>
<th>$\mu_4$</th>
<th>$\pi_1$</th>
<th>$\pi_2$</th>
<th>$\pi_3$</th>
<th>$\pi_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_1(\theta)$</td>
<td>0.065</td>
<td>0.072</td>
<td>-0.213</td>
<td>-0.052</td>
<td>0.39</td>
<td>-0.01</td>
<td>-0.36</td>
<td>-0.72</td>
</tr>
<tr>
<td>$m_1^{1/2}(\theta)$</td>
<td>0.091</td>
<td>0.107</td>
<td>0.084</td>
<td>0.184</td>
<td>0.67</td>
<td>0.50</td>
<td>0.53</td>
<td>0.36</td>
</tr>
<tr>
<td>$m_3(\theta)$</td>
<td>0.016</td>
<td>0.132</td>
<td>-0.049</td>
<td>0.017</td>
<td>-9.56</td>
<td>-9.38</td>
<td>-9.27</td>
<td>-3.07</td>
</tr>
<tr>
<td>$m_3^{1/2}(\theta)$</td>
<td>0.209</td>
<td>0.291</td>
<td>0.084</td>
<td>0.239</td>
<td>368.11</td>
<td>384.11</td>
<td>355.09</td>
<td>217.04</td>
</tr>
<tr>
<td>$m_4^3(\theta)$</td>
<td>0.052</td>
<td>0.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m_4^{1/2}(\theta)$</td>
<td>0.36</td>
<td>0.72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m_4^{3/2}(\theta)$</td>
<td>0.016</td>
<td>0.132</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Estimated posterior distributions

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$F_{1/2}(\theta)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.01$</td>
<td>-0.154</td>
</tr>
<tr>
<td>$0.05$</td>
<td>0.064</td>
</tr>
<tr>
<td>$0.99$</td>
<td>0.280</td>
</tr>
<tr>
<td>$m_{\text{min}}$</td>
<td>-0.368</td>
</tr>
<tr>
<td>$m_{\text{max}}$</td>
<td>0.446</td>
</tr>
<tr>
<td>$\text{NSE}$</td>
<td>0.001</td>
</tr>
<tr>
<td>$\text{RNE}$</td>
<td>1.005</td>
</tr>
</tbody>
</table>

OLS estimates and estimated asymptotic distributions

<table>
<thead>
<tr>
<th>$\hat{\theta}$</th>
<th>$\mathcal{F}_{1/2}(\hat{\theta})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.01$</td>
<td>-0.134</td>
</tr>
<tr>
<td>$0.05$</td>
<td>0.242</td>
</tr>
<tr>
<td>$0.99$</td>
<td>0.054</td>
</tr>
</tbody>
</table>

Estimated small-sample distributions (OLS)

<table>
<thead>
<tr>
<th>$\hat{\theta}$</th>
<th>$F_{1/2}(\hat{\theta})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.01$</td>
<td>-0.188</td>
</tr>
<tr>
<td>$0.05$</td>
<td>0.049</td>
</tr>
<tr>
<td>$0.99$</td>
<td>0.298</td>
</tr>
<tr>
<td>$\text{min}$</td>
<td>-0.525</td>
</tr>
<tr>
<td>$\text{max}$</td>
<td>0.498</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\hat{\theta}$</th>
<th>$\mathcal{F}_{1/2}(\hat{\theta})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.01$</td>
<td>-0.313</td>
</tr>
<tr>
<td>$0.05$</td>
<td>-0.108</td>
</tr>
<tr>
<td>$0.99$</td>
<td>0.218</td>
</tr>
<tr>
<td>$\text{min}$</td>
<td>-0.119</td>
</tr>
<tr>
<td>$\text{max}$</td>
<td>-0.293</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\hat{\theta}$</th>
<th>$\mathcal{F}_{1/2}(\hat{\theta})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.01$</td>
<td>-0.531</td>
</tr>
<tr>
<td>$0.05$</td>
<td>0.032</td>
</tr>
<tr>
<td>$0.99$</td>
<td>0.119</td>
</tr>
<tr>
<td>$\text{min}$</td>
<td>0.235</td>
</tr>
<tr>
<td>$\text{max}$</td>
<td>0.094</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\hat{\theta}$</th>
<th>$\mathcal{F}_{1/2}(\hat{\theta})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.01$</td>
<td>-0.365</td>
</tr>
<tr>
<td>$0.05$</td>
<td>0.086</td>
</tr>
<tr>
<td>$0.99$</td>
<td>0.213</td>
</tr>
<tr>
<td>$\text{min}$</td>
<td>0.171</td>
</tr>
<tr>
<td>$\text{max}$</td>
<td>0.140</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\hat{\theta}$</th>
<th>$\mathcal{F}_{1/2}(\hat{\theta})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.01$</td>
<td>-0.665</td>
</tr>
<tr>
<td>$0.05$</td>
<td>0.140</td>
</tr>
<tr>
<td>$0.99$</td>
<td>0.635</td>
</tr>
<tr>
<td>$\text{min}$</td>
<td>-1.119</td>
</tr>
<tr>
<td>$\text{max}$</td>
<td>-0.372</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\hat{\theta}$</th>
<th>$\mathcal{F}_{1/2}(\hat{\theta})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.01$</td>
<td>-0.768</td>
</tr>
<tr>
<td>$0.05$</td>
<td>0.728</td>
</tr>
<tr>
<td>$0.99$</td>
<td>0.772</td>
</tr>
<tr>
<td>$\text{min}$</td>
<td>-0.728</td>
</tr>
<tr>
<td>$\text{max}$</td>
<td>0.728</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\hat{\theta}$</th>
<th>$\mathcal{F}_{1/2}(\hat{\theta})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.01$</td>
<td>-0.365</td>
</tr>
<tr>
<td>$0.05$</td>
<td>0.140</td>
</tr>
<tr>
<td>$0.99$</td>
<td>0.635</td>
</tr>
<tr>
<td>$\text{min}$</td>
<td>0.171</td>
</tr>
<tr>
<td>$\text{max}$</td>
<td>0.140</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\hat{\theta}$</th>
<th>$\mathcal{F}_{1/2}(\hat{\theta})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.01$</td>
<td>-0.665</td>
</tr>
<tr>
<td>$0.05$</td>
<td>0.140</td>
</tr>
<tr>
<td>$0.99$</td>
<td>0.635</td>
</tr>
<tr>
<td>$\text{min}$</td>
<td>-1.119</td>
</tr>
<tr>
<td>$\text{max}$</td>
<td>-0.372</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\hat{\theta}$</th>
<th>$\mathcal{F}_{1/2}(\hat{\theta})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.01$</td>
<td>-0.768</td>
</tr>
<tr>
<td>$0.05$</td>
<td>0.728</td>
</tr>
<tr>
<td>$0.99$</td>
<td>0.772</td>
</tr>
<tr>
<td>$\text{min}$</td>
<td>-0.728</td>
</tr>
<tr>
<td>$\text{max}$</td>
<td>0.728</td>
</tr>
</tbody>
</table>
In Table 2, we present the results of a Bayesian estimation of the same model with $\varepsilon = 1$ and the prior in (3.8), based on 10,000 replications with chain lengths of 400. The differences between Table 2 and the first rows of Table 1 are striking. Even though the point estimates are very similar, the posterior distributions of the $\pi_i$ are much closer to normality when prior (3.8) is used. Outliers have been eliminated, and the posterior confidence intervals have become much narrower. An examination of the posterior densities of the spectral radius $\rho[F(\Phi)]$ provides an explanation. When prior (3.5) is used, the mode of the spectral radius is slightly less than, but very close to, unity; with prior (3.8), the mode takes the smaller value of approximately 0.98. This reduction in the spectral radius is not unexpected; indeed, the minimum information prior is known to put zero weight on unit roots in the simple case where $n = 1$ and $p = 1$ (see Zellner, 1971). It is interesting to note, however, that this translates into radically different posterior distributions for the $\pi_i$ (even though the posterior distributions of the $\mu_i$ remain quite similar, and quite close to normality).

This simple experiment leads to two conclusions. Firstly, prior (3.8) is more appropriate than prior (3.5) when the model is known to be stationary. Secondly, the Bayesian method is here the only one which allows useful inference on the expectations $\pi_i$. Asymptotics are misleading, the bootstrap reveals that conditional maximum likelihood has very poor small-sample properties, and the full maximum likelihood method of Deschamps (1997) proved impractical for a process so close to non-stationarity. On the other hand, the Bayesian confidence intervals are reasonably narrow, and do not suggest improper posterior distributions.

### Table 2
Bayesian estimation of logistic normal VAR (prior 3.8)

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$\mu_1$</th>
<th>$\mu_2$</th>
<th>$\mu_3$</th>
<th>$\mu_4$</th>
<th>$\pi_1$</th>
<th>$\pi_2$</th>
<th>$\pi_3$</th>
<th>$\pi_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_1(\theta)$</td>
<td>0.075</td>
<td>0.050</td>
<td>-0.244</td>
<td>-0.064</td>
<td>0.461</td>
<td>0.027</td>
<td>-0.323</td>
<td>-0.649</td>
</tr>
<tr>
<td>$m_2(\theta)$</td>
<td>0.091</td>
<td>0.104</td>
<td>0.078</td>
<td>0.165</td>
<td>0.115</td>
<td>0.079</td>
<td>0.086</td>
<td>0.097</td>
</tr>
<tr>
<td>$m_3(\theta)$</td>
<td>0.007</td>
<td>-0.089</td>
<td>-0.134</td>
<td>-0.034</td>
<td>-0.366</td>
<td>-0.447</td>
<td>-0.344</td>
<td>-0.949</td>
</tr>
<tr>
<td>$m_4(\theta)$</td>
<td>0.226</td>
<td>0.149</td>
<td>0.111</td>
<td>0.266</td>
<td>2.894</td>
<td>2.949</td>
<td>3.169</td>
<td>2.202</td>
</tr>
<tr>
<td>$\theta_{0.01}$</td>
<td>-0.139</td>
<td>-0.199</td>
<td>-0.440</td>
<td>-0.463</td>
<td>0.154</td>
<td>-0.186</td>
<td>-0.555</td>
<td>-0.947</td>
</tr>
<tr>
<td>$\theta_{0.5}$</td>
<td>0.076</td>
<td>0.052</td>
<td>-0.243</td>
<td>-0.065</td>
<td>0.464</td>
<td>0.031</td>
<td>-0.321</td>
<td>-0.637</td>
</tr>
<tr>
<td>$\theta_{0.99}$</td>
<td>0.287</td>
<td>0.285</td>
<td>-0.068</td>
<td>0.323</td>
<td>0.732</td>
<td>0.211</td>
<td>-0.119</td>
<td>-0.469</td>
</tr>
<tr>
<td>min</td>
<td>-0.318</td>
<td>-0.376</td>
<td>-0.595</td>
<td>-0.866</td>
<td>-0.435</td>
<td>-0.604</td>
<td>-0.992</td>
<td>-1.251</td>
</tr>
<tr>
<td>max</td>
<td>0.534</td>
<td>0.436</td>
<td>0.026</td>
<td>0.584</td>
<td>1.109</td>
<td>0.485</td>
<td>0.215</td>
<td>-0.354</td>
</tr>
<tr>
<td>NSE</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.002</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>RNE</td>
<td>1.017</td>
<td>1.012</td>
<td>0.985</td>
<td>1.002</td>
<td>0.994</td>
<td>1.006</td>
<td>0.993</td>
<td>1.002</td>
</tr>
</tbody>
</table>
7. Fully regular dynamic demand models

It is straightforward to extend the arguments in Section 2 to autoregressive distributed lag (AD) models. A general treatment is presented in Deschamps (1997). In the empirical applications of this paper, the degree of the distributed lag polynomial may be plausibly restricted to zero. We will therefore economize on notation by taking the vector AD($p$, 0) model as maintained hypothesis. This model reads as

$$\Phi(L)y_t = \Gamma_0 x_t + \varepsilon_t \quad \text{for } t = p + 1, \ldots, p + T.$$  \hfill (7.1)

If both $y_t$ and $x_t$ are jointly stationary,\(^1\) we may write counterparts of Eqs. (2.2) and (2.3) as

$$y_t = \Pi x_t + u_t \quad \text{for } t = 1, \ldots, p,$$  \hfill (7.2)

$$\Phi(L)y_t = \Phi(1)\Pi x_t + \varepsilon_t \quad \text{for } t = p + 1, \ldots, p + T$$  \hfill (7.3)

with $\Pi = \Phi^{-1}(1)\Gamma_0$.

As first noted by Anderson and Blundell (1982), Eq. (7.2) may be interpreted as an equilibrium demand relation if $y_t$ and $x_t$ contain the dependent and explanatory variables in a demand model. If $y_t$ and $x_t$ contain observations on the levels of the variables, however, the plausibility of the stationarity assumption depends crucially on the definition of the commodity groups, as well as on the chosen sample period. For reasons given in Section 9, we will ensure stationarity by formulating an equilibrium demand relation in first differences. The model we use is the differential version of the price-homogeneous CBS model (Keller and Van Driel, 1985; Barten, 1989 and 1993), which reads

$$\bar{w}_{it} A \ln \left( \frac{q_{it}}{Q_t} \right) = \pi_i + \beta_i A \ln Q_t + \sum_{j=1}^{n} \gamma_{ij} A \ln \left( \frac{p_{ij}}{p_{n+1,t}} \right) + u_{it}$$

for $i = 1, \ldots, n + 1,$ \hfill (7.4)

where $\bar{w}_{it} = (w_{it} + w_{i,t-1})/2$ with $w_{it}$ the expenditure share on commodity $i$, where $q_{it}$ is the real expenditure per capita on commodity $i$, where

$$A \ln Q_t = \sum_{i=1}^{n+1} \bar{w}_{it} A \ln q_{it},$$

\(^1\) It may be argued that the stationarity assumption on $x_t$ is not necessary for Bayesian inference, since such inference is conditional on $x_t$. It is, however, necessary when a comparison with frequentist estimation is intended; such a comparison will be made in Section 9.
and where $p_j$ is the price deflator for commodity $j$. Barten (1993) finds that this model dominates both the Rotterdam system and a differential version of the Almost Ideal system of Deaton and Muellbauer (1980).

Since the sum over $i$ of the dependent variables in (7.4) is identically zero, this is a singular demand system, so that only the first $n$ equations are relevant for the estimation procedure. The vectors $y_t$ and $x_t$ and the matrix $\Pi$ in (7.2) thus take the form:

$$y_t = \left[ \tilde{w}_{1t} A \ln \left( \frac{q_{1t}}{Q_t} \right), \ldots, \tilde{w}_{mt} A \ln \left( \frac{q_{mt}}{Q_t} \right) \right],$$

\hspace{1cm} (7.5)

$$x_t = \left[ 1, A \ln Q_t, A \ln \left( \frac{p_{1t}}{p_{n+1,t}} \right), \ldots, A \ln \left( \frac{p_{nt}}{p_{n+1,t}} \right) \right],$$

\hspace{1cm} (7.6)

$$\Pi = \begin{pmatrix} \alpha_1 & \beta_1 & \gamma_{11} & \cdots & \gamma_{1n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_n & \beta_n & \gamma_{n1} & \cdots & \gamma_{nn} \end{pmatrix} = (\Pi_0, \Gamma),$$

\hspace{1cm} (7.7)

where $\Gamma$ denotes the matrix consisting of the last $n$ columns of $\Pi$. Full regularity implies that $\Gamma$ be symmetric and negative definite; the parameters of the complete equilibrium model can of course be recovered from the adding-up and homogeneity restrictions.

We define $\Sigma$ as in (2.6)–(2.8), and $z_t = \Pi x_t$. The likelihood is then a straightforward extension of (3.1), and may be written as

$$L(\mathcal{Y} | \Pi, \Phi, \Omega, \mathcal{X}) \propto L_1(\Pi, \Phi, \Omega) L_2(\Pi, \Phi, \Omega)$$

\hspace{1cm} (7.8)

with

$$L_1(\Pi, \Phi, \Omega) = (\det \Sigma)^{-1/2} \exp \left[ -\frac{1}{2} \left[ y_0 - (I_p \otimes \Pi) x_0 \right] \Sigma^{-1} \right] \times \left[ y_0 - (I_p \otimes \Pi) x_0 \right],$$

\hspace{1cm} (7.9)

$$L_2(\Pi, \Phi, \Omega) = (\det \Omega)^{-\tau/2} \exp \left[ -\frac{1}{2} \sum_{t=p+1}^{p+T} \left[ \Phi(L)y_t - \Phi(1)z_t \right] \Omega^{-1} \right] \times \left[ \Phi(L)y_t - \Phi(1)z_t \right],$$

\hspace{1cm} (7.10)

where $y_0 = (y_1', \ldots, y_p')', x_0 = (x_1', \ldots, x_p')', \mathcal{Y} = (y_1', \ldots, y_{p+T}')$, and $\mathcal{X} = (x_1', \ldots, x_{p+T}')$. 
As in Section 2, the model corresponding to (7.8) may be written as an Aitken regression equation $y_\gamma = X_\gamma \gamma + v$ with the coefficient vector $\gamma$ containing the unconstrained elements of vec II, and Eq. (7.3) may be written as a classical multivariate regression $Y = \Phi X + E$, with the coefficient matrix $\Phi$ defined in (2.8). The counterparts of Eqs. (2.9) and (2.10) are

$$y_\gamma = \begin{pmatrix} y_1 \\ \vdots \\ y_p \\ \Phi(L)y_{p+1} \\ \vdots \\ \Phi(L)y_{p+T} \end{pmatrix}, \quad (7.11)$$

$$X_\gamma = \begin{pmatrix} x_1 \otimes I_n \\ \vdots \\ x_p \otimes I_n \\ x_{p+1} \otimes \Phi(1) \\ \vdots \\ x_{p+T} \otimes \Phi(1) \end{pmatrix} D_S, \quad (7.12)$$

where $D_S$ is a matrix of zeros and ones such that vec II = $D_S \gamma$. This conveniently imposes the symmetry constraint. Also, upon noting that (7.3) implies

$$y_t - z_t = \sum_{j=1}^p \Phi_j(y_{t-j} - z_t) + \epsilon_t,$$

the counterparts of Eqs. (2.13) and (2.14) are

$$Y = \begin{pmatrix} y_{p+1} - z_{p+1} & y_{p+2} - z_{p+2} & \cdots & y_{p+T} - z_{p+T} \\ \vdots & \vdots & \ddots & \vdots \\ y_p - z_{p+1} & y_{p+1} - z_{p+2} & \cdots & y_{p+T-1} - z_{p+T} \end{pmatrix}, \quad (7.13)$$

$$X = \begin{pmatrix} y_1 - z_{p+1} & y_2 - z_{p+2} & \cdots & y_T - z_{p+T} \\ \vdots & \vdots & \ddots & \vdots \\ y_{p+1} - z_{p+1} & y_{p+2} - z_{p+2} & \cdots & y_{p+T-1} - z_{p+T} \end{pmatrix}. \quad (7.14)$$

As before, we define $\Theta(\Phi, \Omega)$ as the $n \times n$ matrix given by the diagonal blocks of $\Sigma$, and extend our second prior in Section 3 to impose full regularity on the equilibrium demand relation. Upon recalling that $\Pi = (\Pi_0, \Gamma)$ and denoting by $N$ the set of all negative definite $n \times n$ matrices, our joint prior becomes

$$p(\Pi_0, \Gamma, \Phi, \Omega) \propto I_s(\Phi)I_N(\Gamma)[\det \Theta(\Phi, \Omega)]^{-(n+1)/2} \quad (7.15)$$
implying as joint posterior
\[
p(\Pi_0, \Gamma, \Phi, \Omega | \mathcal{Y}, \mathcal{X}) \propto I_S(\Phi) I_N(\Gamma) [\det \Theta(\Phi, \Omega)]^{-(n+1)/2} \mathcal{L}_1(\Pi, \Phi, \Omega) \mathcal{L}_2(\Pi, \Phi, \Omega).
\] (7.16)

The arguments of Section 4 are now readily extended to the present context. We let \( V \) be as in (2.12), \( X_{**} \) consist of the first 2n columns of \( X_\ast \), and \( X_\Gamma \) consist of the remaining columns (indexed from 2n+1 to 2n+n(n+1)/2) of \( X_\ast \). We also define \( v(\Gamma) \) as the column vector containing the lower triangle of \( \Gamma \), and

\[
y_\Gamma = y_\ast - X_{**} \text{vec} \Pi_0, \tag{7.17}
\]
\[
y_{**} = y_\ast - X_\Gamma v(\Gamma). \tag{7.18}
\]

The full conditional posteriors are
\[
p(\text{vec} \Pi_0 | \Phi, \Gamma, \Omega, \mathcal{Y}, \mathcal{X}) = f_{\text{NO}}[(X_{**} V^{-1} X_{**})^{-1} X' V^{-1} y_{**}, (X_{**} V^{-1} X_{**})^{-1}], \tag{7.19}
\]
\[
p(v(\Gamma) | \Pi_0, \Phi, \Omega, \mathcal{Y}, \mathcal{X}) \propto f_{\text{NO}}[(X_\Gamma V^{-1} X_\Gamma)^{-1} X_\Gamma V^{-1} y_\Gamma, (X_\Gamma V^{-1} X_\Gamma)^{-1}] I_N(\Gamma), \tag{7.20}
\]
\[
p(\Omega | \Pi, \Phi, \mathcal{Y}, \mathcal{X}) \propto f_{\text{IW}}[T, (Y - \Phi X)(Y - \Phi X')] \mathcal{L}_\#(\Pi, \Phi, \Omega), \tag{7.21}
\]
\[
p(\text{vec} \Phi | \Pi, \Omega, \mathcal{Y}, \mathcal{X}) \propto f_{\text{NO}}[\text{vec} \ Y' X'(X' X')^{-1}, \ (X' X')^{-1} \otimes \Omega] I_S(\Phi) \mathcal{L}_\#(\Pi, \Phi, \Omega) \tag{7.22}
\]

with
\[
\mathcal{L}_\#(\Pi, \Phi, \Omega) = \mathcal{L}_1(\Pi, \Phi, \Omega)[\det \Theta(\Phi, \Omega)]^{-(n+1)/2}. \tag{7.23}
\]

If negativity is not imposed, \( I_N(\Gamma) \) in (7.15) and (7.16) is replaced by unity; if symmetry is not imposed, \( D_\# \) in (7.12) is replaced by an identity matrix. In both cases (7.19) and (7.20) are replaced by a single multinormal density.

A full description of the Metropolis–Hastings algorithm for simulating (7.16) is found in Appendix C.

8. Bayes factors

We now discuss Bayesian Slutsky symmetry and negativity tests. Given two models \( M_1 \) and \( M_2 \) with parameter vector \( \theta \) and observed data \( \mathcal{Y} \), the Bayes factor for \( M_1 \) against \( M_2 \) is defined by
\[
B_{12} = \frac{P_1(\mathcal{Y})}{P_2(\mathcal{Y})} \tag{8.1}
\]
with
\[
P_k(\mathcal{Y}) = \int \mathcal{L}_k(\mathcal{Y} \mid \theta)p_k(\theta) \, d\theta \quad \text{for } k = 1, 2, \tag{8.2}
\]

where \( \mathcal{L}_k(\mathcal{Y} \mid \theta) \) is the likelihood for model \( M_k \) and \( p_k(\theta) \) is the prior density of \( \theta \) under model \( M_k \), assumed to be proper. Eq. (8.2) defines the marginal likelihood of model \( M_k \).

In our context, proper prior densities can be obtained by using part of our data as a training sample; this use of ‘partial Bayes factors’ is advocated by authors mentioned in the survey paper by Kass and Raftery (1995). So, the CBS models with homogeneity \((k = H)\), symmetry \((k = S)\), and negativity \((k = N)\) are re-estimated using the first \( T_x + p \) observations, where \( T_x \) is the lowest number which ensures the non-singularity of the OLS estimate of \( \Omega \) in the homogeneous model. The resulting posteriors
\[
p_k(\Pi, \Phi, \Omega \mid \mathcal{Y}_x, \mathcal{X}_x) \quad \text{for } k = H, S, N \tag{8.3}
\]

are then viewed, for the purpose of estimating (8.2), as proper priors in a second inference involving the remaining \( T_B = T - T_x \) observations. The likelihood for these observations is
\[
\mathcal{L}_k(\mathcal{Y}_B \mid \Pi, \Phi, \Omega, \mathcal{X}_B) = (2\pi)^{-(nT_B)/2} |\Omega|^{T_B/2} \exp \left[ -\frac{1}{2} \sum_{t=T_x+1}^{T+p} \left[ \Phi(L)y_t - \Phi(1)z_t \right] \Omega^{-1} \left[ \Phi(L)y_t - \Phi(1)z_t \right] \right]. \tag{8.4}
\]

It is unfortunately unfeasible in practice to estimate (8.2) by a prior-weighted average of likelihood values, due to the enormous range of the likelihood function. However, a methodology described in Chib (1995) and Chib and Greenberg (1998) may be used to estimate the marginal likelihood from the simulation output, as the ratio of the prior times the likelihood divided by the posterior, with all normalizing constants included and where all functions are evaluated at a point \((\Pi^*_k, \Phi^*_k, \Omega^*_k)\) of high posterior mass. Specifically, the partial marginal likelihood for model \( M_k \) is estimated from:
\[
P_k(\mathcal{Y} \mid \mathcal{Y}_x, \mathcal{X}) = \mathcal{L}_k(\mathcal{Y} \mid \Pi^*_k, \Phi^*_k, \Omega^*_k, \mathcal{X}_B)p_k(\Pi^*_k, \Phi^*_k, \Omega^*_k \mid \mathcal{Y}_x, \mathcal{X}_x) \bigg/ p_k(\Pi^*_k, \Phi^*_k, \Omega^*_k \mid \mathcal{Y}, \mathcal{X}). \tag{8.5}
\]

In view of (8.5), the problem of estimating marginal likelihoods for the homogeneous, symmetric, and fully regular models reduces to estimating posterior density ordinates. Following Chib and Greenberg (1998), this is done by suitably factorizing the density and using a combination of kernel and other techniques. The details of this exercise are given in Appendix D.
9. Estimation results for the dynamic CBS model

In this section, we discuss the estimation of the model of Section 7, using the commodities, data source, and sample period described in the first paragraph of Section 6. Dickey–Fuller tests on all the variables in (7.5)–(7.6) led, in every instance, to a strong rejection of the unit root hypothesis; and, in this case, the Johansen trace test and various information criteria provided unanimous evidence against unit roots in the joint processes describing $y_t$ and $x_t$. It also appears that the differential formulation is, in this case, appropriate: cointegration tests led us to strongly reject the hypothesis that the exogenous variables in a levels version of the CBS model are jointly stationary, whereas the joint stationarity of the endogenous variables was not rejected. So, the cointegration approach used by Ng (1995) and Attfield (1997) does not appear appropriate for this model and data set. An attempt to estimate a levels $AD(1,1)$ version provided further evidence in favor of the differential model, with most of the posterior mass concentrated near $\Phi(1) = O$.2

A preliminary specification search on model (7.1) with the variables in (7.5)–(7.6), based on the Akaike information criterion, led to the choice of $p = 1$; additional lags of $x_t$ in that equation were not jointly significant. It is also interesting to note that a static version of the model, with $p = 0$ in (7.1) and symmetry not imposed, exhibits only marginal dynamic misspecification: a multivariate Breusch–Godfrey statistic with exact simulation-based critical values (Deschamps, 1993, 1996) is significant at the 7%, but not at the 6% level. Imposing Slutsky symmetry on the dynamic model led to a very good agreement with the data, the likelihood ratio statistic being 3.83; the probability that this value be exceeded by a Chi-square variate with six degrees of freedom is 0.70.

Our first endeavour was the comparison of the Bayesian and classical estimates when symmetry (but not negativity) is imposed. This comparison was intended as an empirical validation of our simulation methodology, as a confirmation that our choice of prior does not introduce unwanted information on the parameters of main interest, and as a means of assessing the quality of the asymptotic approximations generally used in the context of classical estimation.

Two frequentist methods are available for the estimation of a stationary dynamic allocation model with Slutsky symmetry. The first (and more commonly used) one is conditional maximum likelihood, based on the maximization of (7.10); this can be obtained by iterating generalized least squares on (C.1) and (C.2) in Appendix C, with $\hat{\Omega} = T^{-1} \sum \hat{e}_i \hat{e}_i'$. The second one is full maximum

---

2 It should be stressed, however, that this does not imply the general invalidity of a stationary formulation in levels: different conclusions are obtained when Energy and Other Goods are aggregated, as in Deschamps (1997).
likelihood, based on the maximization of (7.8). This method is fully described in Deschamps (1997).

Table 3 presents the estimation results for both methods, with \( \hat{\theta}_{0.01} \) and \( \hat{\theta}_{0.99} \) the bounds of a 98% confidence interval based on the normal asymptotic distribution of \( \hat{\theta} \). These bounds are provided in order to facilitate the

<table>
<thead>
<tr>
<th>( \theta )</th>
<th>Full ML</th>
<th>Conditional ML</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{\theta} )</td>
<td>( \hat{\theta}^{1/2}(\hat{\theta}) )</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-0.113</td>
<td>0.046</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>-0.035</td>
<td>0.053</td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td>0.090</td>
<td>0.034</td>
</tr>
<tr>
<td>( \beta_4 )</td>
<td>0.000</td>
<td>0.043</td>
</tr>
<tr>
<td>( \beta_5 )</td>
<td>0.059</td>
<td>0.051</td>
</tr>
<tr>
<td>( \gamma_{11} )</td>
<td>0.003</td>
<td>0.022</td>
</tr>
<tr>
<td>( \gamma_{21} )</td>
<td>0.048</td>
<td>0.018</td>
</tr>
<tr>
<td>( \gamma_{22} )</td>
<td>-0.116</td>
<td>0.021</td>
</tr>
<tr>
<td>( \gamma_{31} )</td>
<td>-0.041</td>
<td>0.017</td>
</tr>
<tr>
<td>( \gamma_{32} )</td>
<td>0.022</td>
<td>0.013</td>
</tr>
<tr>
<td>( \gamma_{33} )</td>
<td>-0.048</td>
<td>0.021</td>
</tr>
<tr>
<td>( \gamma_{41} )</td>
<td>0.011</td>
<td>0.012</td>
</tr>
<tr>
<td>( \gamma_{42} )</td>
<td>0.013</td>
<td>0.013</td>
</tr>
<tr>
<td>( \gamma_{43} )</td>
<td>-0.003</td>
<td>0.010</td>
</tr>
<tr>
<td>( \gamma_{44} )</td>
<td>-0.007</td>
<td>0.013</td>
</tr>
<tr>
<td>( \gamma_{51} )</td>
<td>-0.021</td>
<td>0.018</td>
</tr>
<tr>
<td>( \gamma_{52} )</td>
<td>0.033</td>
<td>0.019</td>
</tr>
<tr>
<td>( \gamma_{53} )</td>
<td>0.071</td>
<td>0.017</td>
</tr>
<tr>
<td>( \gamma_{54} )</td>
<td>-0.015</td>
<td>0.015</td>
</tr>
<tr>
<td>( \gamma_{55} )</td>
<td>-0.068</td>
<td>0.027</td>
</tr>
<tr>
<td>( \phi_{11} )</td>
<td>-0.165</td>
<td>0.147</td>
</tr>
<tr>
<td>( \phi_{12} )</td>
<td>-0.396</td>
<td>0.166</td>
</tr>
<tr>
<td>( \phi_{13} )</td>
<td>0.115</td>
<td>0.203</td>
</tr>
<tr>
<td>( \phi_{14} )</td>
<td>0.030</td>
<td>0.180</td>
</tr>
<tr>
<td>( \phi_{21} )</td>
<td>-0.074</td>
<td>0.128</td>
</tr>
<tr>
<td>( \phi_{22} )</td>
<td>0.232</td>
<td>0.144</td>
</tr>
<tr>
<td>( \phi_{23} )</td>
<td>0.178</td>
<td>0.175</td>
</tr>
<tr>
<td>( \phi_{24} )</td>
<td>0.296</td>
<td>0.156</td>
</tr>
<tr>
<td>( \phi_{31} )</td>
<td>0.172</td>
<td>0.115</td>
</tr>
<tr>
<td>( \phi_{32} )</td>
<td>0.032</td>
<td>0.131</td>
</tr>
<tr>
<td>( \phi_{33} )</td>
<td>0.048</td>
<td>0.157</td>
</tr>
<tr>
<td>( \phi_{34} )</td>
<td>-0.083</td>
<td>0.138</td>
</tr>
<tr>
<td>( \phi_{41} )</td>
<td>0.178</td>
<td>0.125</td>
</tr>
<tr>
<td>( \phi_{42} )</td>
<td>0.478</td>
<td>0.140</td>
</tr>
<tr>
<td>( \phi_{43} )</td>
<td>0.010</td>
<td>0.170</td>
</tr>
<tr>
<td>( \phi_{44} )</td>
<td>0.130</td>
<td>0.151</td>
</tr>
</tbody>
</table>
comparison with the Bayesian small-sample results that will follow. The
differences between the point estimates of the \( \beta_i \) obtained by the two methods can be
substantial, even though the estimated model does not have a strong dynamic
component. On the other hand, the estimates of the \( \gamma_{ij} \) and \( \phi_{ij} \) are more similar
across methods. Two autoregression coefficients \( \hat{\phi}_{12} \) and \( \hat{\phi}_{42} \) are clearly
significant. Note that the matrix of autoregression coefficients in the complete
system is not identifiable (Anderson and Blundell, 1982).

Table 4 presents the results of the Bayesian estimation of (7.2)–(7.3) with
symmetry (but not negativity) imposed. They are based on 10,000 replications
and chain lengths of 300. A value of \( \varepsilon = 0.99 \) in (3.4) is appropriate for differ-
cenced data, and was chosen in order to avoid occasional numerical instability.
As before, \( m_i(\theta) \) denotes the \( i \)th sample moment, \( \theta_{a} \) denotes the \( a \)th sample
quantile, and NSE is the numerical standard error.

A comparison of Tables 3 and 4 reveals the remarkable closeness of the point
estimates (\( \theta \) in column 1 of Table 3; \( m_1(\theta) \) and \( \theta_{0.5} \) in Table 4) of the budget
and price coefficients \( \beta_i \) and \( \gamma_{ij} \) in the full ML and Bayesian procedures.
However, the asymptotic approximation used in Table 3 appears to generally
overestimate the actual small-sample precision, indicated by the lengths of the
confidence intervals in Table 4. (Incidentally, this overestimation is even larger
when the ML estimates in the slightly misspecified static model are compared
with the results in Table 4). As in Section 6, the skewness and excess kurtosis
statistics indicate very substantial deviations from normality for the equilibrium
coefficients \( \beta_i \) and \( \gamma_{ij} \), whereas the normal approximations are more appropriate
in the case of the autoregression coefficients \( \phi_{ij} \).

The Bayesian estimation results for the fully regular model are presented in
Table 5. They are based on 10,000 replications, chain lengths of 300, and
\( \varepsilon = 0.99 \). A comparison of Tables 4 and 5 does not reveal large differences in the
point estimates of the \( \beta_i \) and \( \gamma_{ij} \), except for \( \gamma_{11}, \gamma_{33}, \gamma_{55} \), and \( \gamma_{31} \). However, in
Table 5, the marginal posterior distributions of the equilibrium coefficients
appear to be even further distant from normality than in Table 4, and the
posterior confidence intervals are generally larger.

Figs. 1 and 2 reveal that food can be unambiguously classified as a necessity,
clothing/footwear and (to a lesser extent) other goods as luxuries, the other
commodities being intermediate cases with approximately equivalent rankings.
Figs. 3 and 4 reveal that food and energy are the most price-inelastic commodi-
ties and drink/tobacco the most price-elastic, with clothing/footwear and other
goods having intermediate own-price Slutsky elasticities. The elasticities were
evaluated at the average budget shares.

The partial marginal likelihoods and partial Bayes factors are presented in
Table 6; the prior densities used in defining the marginal likelihoods were based
on a training sample involving the first \( T_z + 1 = 15 \) observations. The corre-
sponding classical procedure would be a test of predictive failure, with the last
two-thirds of the sample set aside for that purpose. The symmetric model clearly
Table 4  
Bayesian estimation of the symmetric dynamic CBS model

<table>
<thead>
<tr>
<th>θ</th>
<th>$m_1(θ)$</th>
<th>$\frac{m_2(θ)}{m_3^2(θ)}$</th>
<th>$\frac{m_3(θ)}{m_4(θ)}$</th>
<th>$\frac{m_4(θ)}{m_5(θ)}$</th>
<th>$\theta_{0.01}$</th>
<th>$\theta_{0.5}$</th>
<th>$\theta_{0.99}$</th>
<th>NSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>β1</td>
<td>−0.120</td>
<td>0.073</td>
<td>−0.872</td>
<td>13.46</td>
<td>−0.325</td>
<td>−0.118</td>
<td>0.041</td>
<td>0.0007</td>
</tr>
<tr>
<td>β2</td>
<td>−0.027</td>
<td>0.084</td>
<td>1.323</td>
<td>24.64</td>
<td>−0.220</td>
<td>−0.031</td>
<td>0.208</td>
<td>0.0008</td>
</tr>
<tr>
<td>β3</td>
<td>0.093</td>
<td>0.055</td>
<td>−0.345</td>
<td>6.80</td>
<td>−0.048</td>
<td>0.093</td>
<td>0.230</td>
<td>0.0005</td>
</tr>
<tr>
<td>β4</td>
<td>−0.001</td>
<td>0.067</td>
<td>0.821</td>
<td>16.35</td>
<td>−0.152</td>
<td>−0.003</td>
<td>0.181</td>
<td>0.0007</td>
</tr>
<tr>
<td>β5</td>
<td>0.055</td>
<td>0.080</td>
<td>−0.665</td>
<td>20.54</td>
<td>−0.159</td>
<td>0.057</td>
<td>0.234</td>
<td>0.0008</td>
</tr>
<tr>
<td>γ11</td>
<td>0.006</td>
<td>0.036</td>
<td>−0.247</td>
<td>4.07</td>
<td>−0.087</td>
<td>0.006</td>
<td>0.090</td>
<td>0.0004</td>
</tr>
<tr>
<td>γ21</td>
<td>0.051</td>
<td>0.030</td>
<td>1.145</td>
<td>16.48</td>
<td>−0.017</td>
<td>0.049</td>
<td>0.135</td>
<td>0.0003</td>
</tr>
<tr>
<td>γ22</td>
<td>−0.122</td>
<td>0.038</td>
<td>−2.207</td>
<td>32.10</td>
<td>−0.228</td>
<td>−0.120</td>
<td>−0.040</td>
<td>0.0004</td>
</tr>
<tr>
<td>γ31</td>
<td>−0.044</td>
<td>0.026</td>
<td>−0.209</td>
<td>3.74</td>
<td>−0.109</td>
<td>−0.044</td>
<td>0.016</td>
<td>0.0003</td>
</tr>
<tr>
<td>γ32</td>
<td>0.023</td>
<td>0.021</td>
<td>0.749</td>
<td>15.03</td>
<td>−0.029</td>
<td>0.023</td>
<td>0.073</td>
<td>0.0002</td>
</tr>
<tr>
<td>γ33</td>
<td>−0.054</td>
<td>0.033</td>
<td>−0.821</td>
<td>4.59</td>
<td>−0.148</td>
<td>−0.052</td>
<td>0.020</td>
<td>0.0003</td>
</tr>
<tr>
<td>γ41</td>
<td>0.012</td>
<td>0.021</td>
<td>1.220</td>
<td>20.79</td>
<td>−0.038</td>
<td>0.011</td>
<td>0.067</td>
<td>0.0002</td>
</tr>
<tr>
<td>γ42</td>
<td>0.013</td>
<td>0.023</td>
<td>−2.139</td>
<td>45.40</td>
<td>−0.047</td>
<td>0.013</td>
<td>0.066</td>
<td>0.0002</td>
</tr>
<tr>
<td>γ43</td>
<td>−0.001</td>
<td>0.016</td>
<td>1.621</td>
<td>28.43</td>
<td>−0.036</td>
<td>−0.002</td>
<td>0.043</td>
<td>0.0002</td>
</tr>
<tr>
<td>γ44</td>
<td>−0.008</td>
<td>0.021</td>
<td>−1.015</td>
<td>29.87</td>
<td>−0.064</td>
<td>−0.008</td>
<td>0.043</td>
<td>0.0002</td>
</tr>
<tr>
<td>γ51</td>
<td>−0.024</td>
<td>0.027</td>
<td>−1.247</td>
<td>23.03</td>
<td>−0.092</td>
<td>−0.023</td>
<td>0.037</td>
<td>0.0003</td>
</tr>
<tr>
<td>γ52</td>
<td>0.036</td>
<td>0.032</td>
<td>2.425</td>
<td>43.10</td>
<td>−0.029</td>
<td>0.034</td>
<td>0.120</td>
<td>0.0003</td>
</tr>
<tr>
<td>γ53</td>
<td>0.076</td>
<td>0.026</td>
<td>0.745</td>
<td>4.30</td>
<td>0.023</td>
<td>0.074</td>
<td>0.152</td>
<td>0.0003</td>
</tr>
<tr>
<td>γ54</td>
<td>−0.016</td>
<td>0.022</td>
<td>1.724</td>
<td>55.62</td>
<td>−0.070</td>
<td>−0.016</td>
<td>0.036</td>
<td>0.0002</td>
</tr>
<tr>
<td>γ55</td>
<td>−0.072</td>
<td>0.039</td>
<td>−1.632</td>
<td>29.51</td>
<td>−0.173</td>
<td>−0.071</td>
<td>0.012</td>
<td>0.0004</td>
</tr>
<tr>
<td>φ11</td>
<td>−0.084</td>
<td>0.181</td>
<td>0.053</td>
<td>0.14</td>
<td>−0.501</td>
<td>−0.085</td>
<td>0.346</td>
<td>0.0018</td>
</tr>
<tr>
<td>φ12</td>
<td>−0.374</td>
<td>0.201</td>
<td>0.022</td>
<td>0.14</td>
<td>−0.845</td>
<td>−0.374</td>
<td>0.100</td>
<td>0.0020</td>
</tr>
<tr>
<td>φ13</td>
<td>0.042</td>
<td>0.253</td>
<td>−0.025</td>
<td>0.09</td>
<td>0.553</td>
<td>0.042</td>
<td>0.625</td>
<td>0.0025</td>
</tr>
<tr>
<td>φ14</td>
<td>−0.039</td>
<td>0.225</td>
<td>−0.006</td>
<td>0.17</td>
<td>−0.559</td>
<td>−0.037</td>
<td>0.495</td>
<td>0.0022</td>
</tr>
<tr>
<td>φ21</td>
<td>−0.055</td>
<td>0.145</td>
<td>0.027</td>
<td>0.15</td>
<td>−0.397</td>
<td>−0.056</td>
<td>0.290</td>
<td>0.0015</td>
</tr>
<tr>
<td>φ22</td>
<td>0.301</td>
<td>0.166</td>
<td>−0.011</td>
<td>0.14</td>
<td>−0.102</td>
<td>0.302</td>
<td>0.694</td>
<td>0.0017</td>
</tr>
<tr>
<td>φ23</td>
<td>0.159</td>
<td>0.200</td>
<td>−0.020</td>
<td>0.04</td>
<td>−0.316</td>
<td>0.160</td>
<td>0.623</td>
<td>0.0020</td>
</tr>
<tr>
<td>φ24</td>
<td>0.307</td>
<td>0.181</td>
<td>−0.032</td>
<td>0.22</td>
<td>−0.121</td>
<td>0.309</td>
<td>0.729</td>
<td>0.0018</td>
</tr>
<tr>
<td>φ31</td>
<td>0.167</td>
<td>0.136</td>
<td>0.047</td>
<td>0.25</td>
<td>−0.154</td>
<td>0.168</td>
<td>0.493</td>
<td>0.0014</td>
</tr>
<tr>
<td>φ32</td>
<td>0.042</td>
<td>0.152</td>
<td>0.047</td>
<td>0.14</td>
<td>−0.309</td>
<td>0.040</td>
<td>0.404</td>
<td>0.0015</td>
</tr>
<tr>
<td>φ33</td>
<td>0.196</td>
<td>0.195</td>
<td>0.115</td>
<td>0.12</td>
<td>−0.249</td>
<td>0.193</td>
<td>0.673</td>
<td>0.0019</td>
</tr>
<tr>
<td>φ34</td>
<td>−0.048</td>
<td>0.161</td>
<td>−0.023</td>
<td>0.29</td>
<td>−0.434</td>
<td>−0.048</td>
<td>0.338</td>
<td>0.0016</td>
</tr>
<tr>
<td>φ41</td>
<td>0.145</td>
<td>0.141</td>
<td>0.032</td>
<td>0.19</td>
<td>−0.184</td>
<td>0.144</td>
<td>0.482</td>
<td>0.0014</td>
</tr>
<tr>
<td>φ42</td>
<td>0.432</td>
<td>0.157</td>
<td>0.091</td>
<td>0.16</td>
<td>0.076</td>
<td>0.428</td>
<td>0.809</td>
<td>0.0016</td>
</tr>
<tr>
<td>φ43</td>
<td>0.004</td>
<td>0.188</td>
<td>0.025</td>
<td>0.13</td>
<td>−0.443</td>
<td>0.003</td>
<td>0.456</td>
<td>0.0019</td>
</tr>
<tr>
<td>φ44</td>
<td>0.198</td>
<td>0.172</td>
<td>0.018</td>
<td>0.06</td>
<td>−0.204</td>
<td>0.198</td>
<td>0.611</td>
<td>0.0017</td>
</tr>
</tbody>
</table>

dominates; using the classification given by Jeffreys (1961, Appendix B), the evidence against the other models can be characterized as decisive. This confirms our previous asymptotic LR test of symmetry, and the evidence against the fully regular model reflects the fact that most of the posterior mass is concentrated on matrices Φ that are not negative definite. It is also interesting to note
Table 5
Bayesian estimation of the regular dynamic CBS model

<table>
<thead>
<tr>
<th>( \theta )</th>
<th>( m_3(\theta) )</th>
<th>( m_4(\theta) )</th>
<th>( \frac{1}{m_3^2(\theta)} )</th>
<th>( \frac{m_4(\theta)}{m_3^2(\theta)} - 3 )</th>
<th>( \theta_{0.01} )</th>
<th>( \theta_{0.5} )</th>
<th>( \theta_{0.99} )</th>
<th>NSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_1 )</td>
<td>-0.118</td>
<td>0.087</td>
<td>-3.70</td>
<td>118.3</td>
<td>-0.339</td>
<td>-0.116</td>
<td>0.082</td>
<td>0.0009</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>-0.040</td>
<td>0.097</td>
<td>2.92</td>
<td>52.2</td>
<td>-0.258</td>
<td>-0.043</td>
<td>0.225</td>
<td>0.0010</td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td>0.106</td>
<td>0.060</td>
<td>-0.03</td>
<td>59.7</td>
<td>-0.032</td>
<td>0.105</td>
<td>0.243</td>
<td>0.0006</td>
</tr>
<tr>
<td>( \beta_4 )</td>
<td>-0.014</td>
<td>0.095</td>
<td>10.40</td>
<td>316.8</td>
<td>-0.184</td>
<td>-0.019</td>
<td>0.211</td>
<td>0.0010</td>
</tr>
<tr>
<td>( \beta_5 )</td>
<td>0.067</td>
<td>0.103</td>
<td>-6.55</td>
<td>134.4</td>
<td>-0.203</td>
<td>0.072</td>
<td>0.261</td>
<td>0.0010</td>
</tr>
<tr>
<td>( \gamma_{11} )</td>
<td>-0.057</td>
<td>0.045</td>
<td>-12.22</td>
<td>366.7</td>
<td>-0.189</td>
<td>-0.050</td>
<td>0.011</td>
<td>0.0005</td>
</tr>
<tr>
<td>( \gamma_{21} )</td>
<td>0.063</td>
<td>0.042</td>
<td>7.93</td>
<td>164.0</td>
<td>0.009</td>
<td>0.057</td>
<td>0.194</td>
<td>0.0004</td>
</tr>
<tr>
<td>( \gamma_{22} )</td>
<td>-0.140</td>
<td>0.049</td>
<td>-5.45</td>
<td>64.9</td>
<td>-0.301</td>
<td>-0.132</td>
<td>0.073</td>
<td>0.0005</td>
</tr>
<tr>
<td>( \gamma_{31} )</td>
<td>-0.006</td>
<td>0.027</td>
<td>6.76</td>
<td>195.6</td>
<td>-0.056</td>
<td>-0.008</td>
<td>0.063</td>
<td>0.0003</td>
</tr>
<tr>
<td>( \gamma_{32} )</td>
<td>0.025</td>
<td>0.026</td>
<td>0.20</td>
<td>131.0</td>
<td>-0.036</td>
<td>0.025</td>
<td>0.083</td>
<td>0.0003</td>
</tr>
<tr>
<td>( \gamma_{33} )</td>
<td>0.090</td>
<td>0.034</td>
<td>-3.35</td>
<td>37.6</td>
<td>0.000</td>
<td>0.087</td>
<td>0.001</td>
<td>0.0003</td>
</tr>
<tr>
<td>( \gamma_{41} )</td>
<td>0.018</td>
<td>0.028</td>
<td>10.16</td>
<td>229.1</td>
<td>-0.012</td>
<td>0.013</td>
<td>0.110</td>
<td>0.0003</td>
</tr>
<tr>
<td>( \gamma_{42} )</td>
<td>0.002</td>
<td>0.029</td>
<td>-7.33</td>
<td>122.2</td>
<td>-0.095</td>
<td>0.006</td>
<td>0.036</td>
<td>0.0003</td>
</tr>
<tr>
<td>( \gamma_{43} )</td>
<td>0.005</td>
<td>0.018</td>
<td>1.75</td>
<td>143.3</td>
<td>-0.025</td>
<td>0.004</td>
<td>0.054</td>
<td>0.0002</td>
</tr>
<tr>
<td>( \gamma_{44} )</td>
<td>0.029</td>
<td>0.029</td>
<td>13.18</td>
<td>301.7</td>
<td>-0.011</td>
<td>-0.023</td>
<td>0.004</td>
<td>0.0003</td>
</tr>
<tr>
<td>( \gamma_{51} )</td>
<td>-0.018</td>
<td>0.036</td>
<td>-9.92</td>
<td>277.1</td>
<td>0.019</td>
<td>0.014</td>
<td>0.037</td>
<td>0.0004</td>
</tr>
<tr>
<td>( \gamma_{52} )</td>
<td>0.050</td>
<td>0.040</td>
<td>6.61</td>
<td>111.3</td>
<td>0.002</td>
<td>0.044</td>
<td>0.173</td>
<td>0.0004</td>
</tr>
<tr>
<td>( \gamma_{53} )</td>
<td>0.065</td>
<td>0.029</td>
<td>4.61</td>
<td>93.0</td>
<td>0.014</td>
<td>0.063</td>
<td>0.143</td>
<td>0.0003</td>
</tr>
<tr>
<td>( \gamma_{54} )</td>
<td>0.003</td>
<td>0.027</td>
<td>7.99</td>
<td>133.7</td>
<td>0.000</td>
<td>0.000</td>
<td>0.088</td>
<td>0.0003</td>
</tr>
<tr>
<td>( \gamma_{55} )</td>
<td>-0.102</td>
<td>0.046</td>
<td>-6.77</td>
<td>137.9</td>
<td>-0.242</td>
<td>-0.095</td>
<td>0.036</td>
<td>0.0005</td>
</tr>
<tr>
<td>( \phi_{11} )</td>
<td>-0.115</td>
<td>0.196</td>
<td>0.13</td>
<td>0.2</td>
<td>-0.558</td>
<td>-0.119</td>
<td>0.362</td>
<td>0.0020</td>
</tr>
<tr>
<td>( \phi_{12} )</td>
<td>-0.434</td>
<td>0.221</td>
<td>0.10</td>
<td>0.3</td>
<td>-0.934</td>
<td>-0.438</td>
<td>0.111</td>
<td>0.0022</td>
</tr>
<tr>
<td>( \phi_{13} )</td>
<td>0.001</td>
<td>0.276</td>
<td>-0.02</td>
<td>0.1</td>
<td>0.066</td>
<td>0.002</td>
<td>0.646</td>
<td>0.0028</td>
</tr>
<tr>
<td>( \phi_{14} )</td>
<td>-0.106</td>
<td>0.242</td>
<td>-0.01</td>
<td>0.3</td>
<td>-0.678</td>
<td>-0.106</td>
<td>0.460</td>
<td>0.0024</td>
</tr>
<tr>
<td>( \phi_{21} )</td>
<td>-0.039</td>
<td>0.147</td>
<td>-0.06</td>
<td>0.1</td>
<td>-0.393</td>
<td>-0.038</td>
<td>0.296</td>
<td>0.0015</td>
</tr>
<tr>
<td>( \phi_{22} )</td>
<td>0.320</td>
<td>0.170</td>
<td>-0.06</td>
<td>0.2</td>
<td>-0.093</td>
<td>0.323</td>
<td>0.715</td>
<td>0.0017</td>
</tr>
<tr>
<td>( \phi_{23} )</td>
<td>0.161</td>
<td>0.197</td>
<td>-0.01</td>
<td>0.2</td>
<td>-0.298</td>
<td>0.163</td>
<td>0.627</td>
<td>0.0020</td>
</tr>
<tr>
<td>( \phi_{24} )</td>
<td>0.323</td>
<td>0.179</td>
<td>0.01</td>
<td>0.2</td>
<td>-0.096</td>
<td>0.323</td>
<td>0.749</td>
<td>0.0018</td>
</tr>
<tr>
<td>( \phi_{31} )</td>
<td>0.194</td>
<td>0.135</td>
<td>0.03</td>
<td>0.4</td>
<td>-0.122</td>
<td>0.193</td>
<td>0.519</td>
<td>0.0013</td>
</tr>
<tr>
<td>( \phi_{32} )</td>
<td>0.083</td>
<td>0.157</td>
<td>0.02</td>
<td>0.2</td>
<td>-0.290</td>
<td>0.083</td>
<td>0.460</td>
<td>0.0016</td>
</tr>
<tr>
<td>( \phi_{33} )</td>
<td>0.206</td>
<td>0.201</td>
<td>0.21</td>
<td>0.4</td>
<td>-0.246</td>
<td>0.201</td>
<td>0.710</td>
<td>0.0020</td>
</tr>
<tr>
<td>( \phi_{34} )</td>
<td>-0.009</td>
<td>0.166</td>
<td>0.06</td>
<td>0.6</td>
<td>-0.413</td>
<td>-0.012</td>
<td>0.389</td>
<td>0.0017</td>
</tr>
<tr>
<td>( \phi_{41} )</td>
<td>0.127</td>
<td>0.145</td>
<td>0.00</td>
<td>0.2</td>
<td>-0.220</td>
<td>0.127</td>
<td>0.472</td>
<td>0.0014</td>
</tr>
<tr>
<td>( \phi_{42} )</td>
<td>0.428</td>
<td>0.165</td>
<td>0.01</td>
<td>0.2</td>
<td>0.038</td>
<td>0.430</td>
<td>0.835</td>
<td>0.0017</td>
</tr>
<tr>
<td>( \phi_{43} )</td>
<td>-0.003</td>
<td>0.196</td>
<td>-0.04</td>
<td>0.3</td>
<td>-0.471</td>
<td>-0.001</td>
<td>0.452</td>
<td>0.0020</td>
</tr>
<tr>
<td>( \phi_{44} )</td>
<td>0.192</td>
<td>0.180</td>
<td>0.10</td>
<td>0.3</td>
<td>-0.230</td>
<td>0.189</td>
<td>0.612</td>
<td>0.0018</td>
</tr>
</tbody>
</table>

that there is approximately as much evidence for the homogeneous model against the fully regular model as for the symmetric against the homogeneous model.

However, an interesting aspect of the fully regular model is the investigation of complementarity and substitution relations between commodity pairs.
Hicksian substitutes are defined by $\gamma_{ij} > 0$, Hicksian complements by $\gamma_{ij} < 0$, whereas $\gamma_{ij} = 0$ indicates Hicks neutrality. This characterization has been criticized by Barten (1990) and by other authors as counter-intuitive, partly on the grounds that a good is its own perfect substitute, whereas full regularity implies that $\gamma_{ii} \leq 0$. Barten (1990) shows that a more natural concept, that of Allais substitutability, can be easily formulated in the context of differential demand systems. The Allais coefficients are defined by

$$a_{ij} = \sum_{h} u_{h} q_{h} \left( \frac{u_{ij}}{u_{i} u_{j}} - \frac{u_{rs}}{u_{r} u_{s}} \right),$$

where $q_{i}$ is the demand for commodity $i$, $u_{i}$ its marginal utility, and $u_{ij}$ is the $j$th partial derivative of $u_{i}$. The Allais coefficients are invariant with respect to the choice of the utility index and are defined in terms of a difference from a standard pair $(r, s)$ with neutral interaction. They can be interpreted as a weighted rate of change of $\ln(u_{i})$ with respect to $\ln(q_{j})$, for a constant utility level. $a_{ij} < 0$ indicates substitution, $a_{ij} > 0$ complementarity, and $a_{ij} = 0$
neutrality. If \(a_{ii}a_{jj} > 0\), the Allais intensity \(a_{ij}^a\) is

\[
a_{ij}^a = \frac{a_{ij}}{\sqrt{a_{ii}a_{jj}}}.
\] (9.2)

The frequentist estimation of Allais intensities is fraught with two difficulties. Firstly, the concept may not make sense when negativity is not imposed; but negativity constrained coefficients are awkward to estimate by maximum likelihood, at least when dynamics are present. Secondly, the Allais coefficients and intensities are themselves highly non-linear functions of the equilibrium coefficients. Given these facts and the limited number of observations usually available, the reliability of asymptotics is an uncertain proposition.

These difficulties do not arise in a Bayesian context. From replications of the \(\beta_i\) and \(\gamma_{ij}\), one can easily simulate the \((n + 2) \times (n + 2)\) matrix:

\[
M = \begin{pmatrix}
\Gamma^* & (\beta^* + \bar{w}) \\
(\beta^* + \bar{w})^t & 0
\end{pmatrix}^{-1},
\] (9.3)
where $\vec{w}$ is the vector of the $n + 1$ average budget shares and $\Gamma^*, \beta^*$ contain the price and budget coefficients in the complete CBS system. From Barten (1990, Eq. (4.15) and 1993, Eq. (52)), the Allais coefficients can be evaluated by $a_{ij} = M_{ij}^\beta - M_{rs}^\beta$, where $M^\beta$ is the leading $(n + 1) \times (n + 1)$ diagonal block of $M$ with elements $M_{ij}^\beta$. In order to compare the concepts of Hicks and Allais substitutability, it makes sense to choose the standard pair $(r, s)$ as Hicks neutral. An examination of Table 5 reveals that the estimated posterior median of $\gamma_{54}$ vanishes to three decimal places, leading us to choose energy and other goods as having neutral Allais interaction.

Fig. 5 presents the posterior densities of all the cross-price Slutsky coefficients except $\gamma_{54}$ in the fully regular model. Fig. 6 presents the posterior densities of the corresponding Allais intensities. Contrary to our own prior expectations, the agreement between the two types of classification criteria is very close. (Food, drink), (drink, clothing), (food, energy), (drink, other goods), and (clothing, other goods) are both Hicks and Allais substitutes. (Clothing, energy) are both Hicks and Allais neutral. (Food, clothing) and (food, other goods) are slight Hicks complements but Allais neutral. The most noticeable difference is for Drink and
Energy, which are Hicks neutral but Allais substitutes. There are no Allais complements, which shows that the Hicks concept does not relatively favor substitutability in this example. This contradicts prior expectations based on the fact that \( \sum_{j=1}^{n} \gamma_{ij} = 0 \) with \( \gamma_{ii} \leq 0 \).
Due to the definition of the Allais intensities in (9.2), those replications for which $a_{ii}a_{jj} \leq 0$ for some pair $(i, j)$ were eliminated from the sample before estimating the densities in Fig. 6. This led to the deletion of 254 (out of 10,000) observations. By contrast, the same exercise in the model with symmetry (but not negativity) led to eliminating about half the sample, and the resulting estimated posterior densities were uninterpretable (and sometimes bimodal).

10. Conclusions

This paper has attempted to show the practicality and usefulness of the Bayesian approach, used in conjunction with Markov chain Monte Carlo methods, for the estimation of nearly non-stationary vector autoregressions and fully regular dynamic demand models. Two different diffuse priors were investigated. The first one assumes only the existence and invertibility of the conditional error covariance matrix, whereas the other assumes the existence and invertibility of the long-run error covariance matrix. It is the position of this author that the second prior is more appropriate when the model is known to be stationary.
One major advantage of the present approach over frequentist estimation techniques is its general applicability, since it does not rely on the convergence of local optimization algorithms. Another major advantage of the method is its validity in small samples. The availability of small-sample results was found to be of crucial importance when concepts requiring the imposition of non-linear restrictions (such as Allais intensities) are investigated. Also, we found that problems created by unit roots are revealed very quickly (after a few hundred replications) by the behavior of the generated Markov chains. Whereas this does not obviate the need for formal unit root or cointegration tests, it does provide useful information in the case where the evidence provided by such tests is inconclusive.

The estimation of the Bayes factors necessitated the Markov chain Monte Carlo estimation of 44 different posterior densities, some of which were based on a sample size just sufficient to identify the model. The algorithm was found to work well in all cases, although 700 passes per replication were needed to ensure proper convergence for the fully regular model based on just 15 observations. In this case only, a relatively high percentage of the chains had to be discarded due to numerical problems (such as the failure of a matrix inversion or eigenvalue routine) occurring at the beginning of the chain. This amounts to a rejection step on the initial candidate density described in Appendix C, an admissible procedure since this density is heuristically chosen.
The decisiveness of the evidence against the fully regular model (odds of more than a million to one) came as a surprise, and such a clear-cut result could not have been obtained by means of known classical tests. It implies that a theoretically consistent analysis of complementarity and substitution relations can only be achieved, in the context of this model and data, at the cost of a dramatic loss in predictive performance. The justification for imposing full regularity thus depends ultimately on an investigator’s objectives, and this imposition would not seem warranted if the sole interest were in, say, budget elasticities. Indeed, a comparison of Figs. 1 and 2 shows very little difference between the symmetric and fully regular models.

The main disadvantage of our Bayesian approach to regularity testing is computational cost. Even though the procedure in Appendix D could, in theory, be scaled up to larger models, an application involving (say) 7 equations and 200 quarterly observations would be impractical at present. However, the lack of classical small-sample alternatives makes this an interesting topic for further research.

Acknowledgements

The author wishes to acknowledge helpful comments from two anonymous referees, and from participants at the econometrics seminar of the University of Limburg at Maastricht, particularly F. Palm, P. Schotman, and J.P. Urbain. He claims full responsibility for any error or omission.

Appendix A. Truncating multinormal matrices on the stationarity region

In this appendix, we consider the problem of generating \( n \times np \) matrices \( \Phi \) distributed according to

\[
f(\text{vec} \Phi) \propto f_{\text{NO}}(\text{vec} M_{\Phi}, V_{\Phi})I_{\mathcal{S}}(\Phi), \tag{A.1}
\]

where \( \mathcal{S} \) is the set in (3.4). \( \mathcal{S} \) is not convex in general: indeed, the two matrices:

\[
A = \begin{pmatrix} -0.625 & 0.125 \\ 0.25 & 0.625 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & 25 \\ 0 & 0 \end{pmatrix}
\]

have \( \rho(A) \approx 0.65 \) and \( \rho(B) = 0 \), whereas \( C = 0.8A + 0.2B \) has \( \rho(C) \approx 1.13 \). \( \mathcal{S} \) is unbounded if \( n > 1 \) since any triangular matrix with null diagonal elements has a spectral radius of zero. However, it is open since the eigenvalues of \( F(\Phi) \) are continuous functions of the elements of \( \Phi \); and it is easy to show that it is
arc-connected. Indeed, if

$$G(x, \Phi) = \Phi \begin{bmatrix} x^p & 0 & \ldots & 0 \\ 0 & x^{p-1} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & x \end{bmatrix} I_n,$$

the same arguments as in Theorem 2 of Deschamps (1997) can be used to show that $G(x, \Phi) \in S$ if $\Phi \in S$ and if $0 \leq x \leq 1$. If $\Phi_0 \in S$, $\Phi_1 \in S$, and

$$H(\delta) = G[\max(0, \delta), \Phi_0] + G[\max(0, -\delta), \Phi_1],$$

we have that $H(-1) = \Phi_1$, $H(1) = \Phi_0$, and $H(\delta) \in S$ for $-1 < \delta < 1$, so that $H(\delta)$ forms a continuous admissible path from $\Phi_1$ to $\Phi_0$ when $\delta$ ranges from $-1$ to $+1$.

The continuity of the multinormal density and the openness of $S$ imply that the target density (A.1) is lower semicontinuous at zero. Since all its marginal densities are bounded, the conditions of Theorem 2 in Roberts and Smith (1994) have been verified, and a Gibbs sampler defined on the $n^2p$ blocks consisting of the elements of $\Phi$ has the required convergence properties.

Robert (1995) describes a Gibbs sampler for the generation of multinormal vectors truncated on a convex set. The non-convexity of $S$ mandates a modification of his algorithm in our case. Let $\Phi_{-ij}$ be the $(n^2p-1)$-vector consisting of all the elements of vec $\Phi$ except $\Phi_{ij}$. It will be shown that for given $\Phi_{-ij}$, the invertibility condition on $\Phi(L)$ requires $\Phi_{ij}$ to lie in a union of at most $2np + 1$ possibly disjoint candidate intervals, whose bounds can be computed using Proposition 1 below. On each interval, the sign of $\rho[F(\Phi)] - \varepsilon$ does not change. Each interval is tested in turn, and those intervals for which $\rho[F(\Phi)] < \varepsilon$ are retained. Let $R$ be the number of these intervals, and let $l_k, u_k$ denote the lower and upper bounds of the $k$th retained interval. Let $\mu_{ij}$ and $\sigma_{ij}$ denote the expectation and standard error of the conditional normal density of $\Phi_{ij}$ given $\Phi_{-ij}$, as derived from $f_{NO}(\text{vec} M \Phi, V \Phi)$. Robert’s algorithm for the truncated univariate normal is first applied to draw candidates $z_k$ with densities proportional to $f_{NO}(\mu_{ij}, \sigma_{ij}^2)I_{[u_k, u_k]}(z_k)$, for $k = 1, \ldots, R$. Let $\Psi(\cdot)$ denote the standard normal integral, and define:

$$p_k = \Psi\left(\frac{u_k - \mu_{ij}}{\sigma_{ij}}\right) - \Psi\left(\frac{l_k - \mu_{ij}}{\sigma_{ij}}\right),$$

We construct the following distribution function:

$$F(\ell) = \frac{\sum_{k=1}^{R} p_k}{\sum_{k=1}^{R} p_k} \quad \text{for } \ell = 1, \ldots, R,$$
draw a random number \( u \) distributed uniformly on \([0, 1]\), and choose candidate \( z \), if \( F(r - 1) \leq u \leq F(r) \) with \( F(0) = 0 \). This generates the appropriate truncated normal draw which is required at each step of the Gibbs sampler. A suitable random starting point for the Markov chain can be readily found, using the normalization described in Deschamps (1997, Theorem 2).

By continuity of the spectral radius, the boundary of \( S \) is characterized by \( \rho[F(\Phi)] = \varepsilon \). The following proposition implies that any finite truncation bound on an element of \( \Phi \), given all the others, can be identified from the roots of a polynomial; it also substantiates our claim that the support of a full conditional is the union of a small number of possibly disjoint intervals.

**Proposition 1.** Let \( F(\Phi) \) be as in (2.7) with \( \Phi \in S \). For \( i = n(p - 1) + 1, \ldots, np \) and for \( j = 1, \ldots, np \), let

\[
F_{ij}(x) = F(\Phi) + xe_i e'_j,
\]

where \( e_i \) and \( e_j \) denote the \( i \)th and \( j \)th \( np \times 1 \) unit vectors. Let, furthermore

\[
A_{ij}(x) = \det[\varepsilon^2 I_{n^2p^2} - F_{ij}(x) \otimes F_{ij}(x)].
\]

Then

1. \( A_{ij}(x) \) is a polynomial of degree at most equal to \( 2np \).
2. The set \( \{x|x \in \mathbb{R}, \rho[F_{ij}(x)] < \varepsilon \} \) is non-empty.
3. If \( \rho[F_{ij}(x)] = \varepsilon \), then \( A_{ij}(x) = 0 \).
4. The sign of \( \rho[F_{ij}(x)] - \varepsilon \) is well-defined on each of the \( 2np + 1 \) following intervals of values of \( x \): \( (-\infty, a_1], [a_1, a_2], [a_2, a_3], \ldots, [a_{2np}, + \infty) \), where \( a_1, \ldots, a_{2np} \) are the real parts of the roots of \( A_{ij}(x) = 0 \), arranged in increasing order.

**Proof.** From Birkhoff and MacLane (1965, p. 281), the determinant of \( F_{ij}(x) \) is a linear function of \( x \). Claim 1 then follows from the fact that

\[
\det[F_{ij}(x) \otimes F_{ij}(x)] = [\det F_{ij}(x)]^{2np},
\]

since \( \det(\varepsilon^2 I - A) \) is a polynomial of the same degree as \( \det(A) \). The degree may be less than \( 2np \) since cancellation of terms may occur. Claim 2 follows from the fact that \( \Phi \in S \) and from the continuity of the spectral radius. In order to prove Claim 3, we note that the eigenvalues of \( F_{ij} \otimes F_{ij} \) are the squares and cross-products of the eigenvalues of \( F_{ij} \). If \( \mu \) is an eigenvalue of \( F_{ij} \) with \( |\mu| = \varepsilon \) and with conjugate \( \mu^* \), we then have that \( \lambda = \mu \mu^* = |\mu|^2 = \varepsilon^2 \) is an eigenvalue of \( F_{ij} \otimes F_{ij} \). But the characteristic equation of \( F_{ij} \otimes F_{ij} \) may be written as

\[
\det[\varepsilon^2 I - F_{ij} \otimes F_{ij} - (\varepsilon^2 - \lambda)I] = 0,
\]

showing that \( \varepsilon^2 - |\mu|^2 = 0 \) is an eigenvalue of \( \varepsilon^2 I - F_{ij} \otimes F_{ij} \). For the proof of Claim 4, we first consider an interval \([a_k, a_{k+1}]\) with \( a_k \neq a_{k+1} \), and we let
\[ x_0 = \frac{(a_k + a_{k+1})}{2}; \quad \rho[F_{ij}(x_0)] = \varepsilon \] would contradict Claim 3, since \( x_0 \) is not a root of \( A_{ij}(x) = 0 \). Assume that \( \rho[F_{ij}(x_0)] < \varepsilon \) and that \( \rho[F_{ij}(x)] > \varepsilon \) for some \( x \in [a_k, a_{k+1}] \). The continuity of the spectral radius would then imply \( \rho[F_{ij}(x)] = \varepsilon \) for some \( x \in (a_k, a_{k+1}) \), which would again contradict Claim 3. If \( \rho[F_{ij}(x_0)] > \varepsilon \), the same argument implies that \( \rho[F_{ij}(x)] \geq \varepsilon \) for all \( x \in [a_k, a_{k+1}] \). When \( a_k = a_{k+1} \), the sign is obviously well-determined since the interval collapses to a single point. For an interval such as \( (-\infty, a_1] \) or \( [a_{2np}, +\infty) \), the same argument as above will apply upon redefining \( x_0 \) as \( a_1 - 1 \) or \( a_{2np} + 1 \). □

The roots \( a_k \) can be evaluated numerically from the coefficients of \( A_{ij}(x) = 0 \), and these coefficients can be evaluated by polynomial regression. It is well-known, however, that naive polynomial regression is numerically very unstable; for this reason, \( A_{ij}(x) = 0 \) is first represented in terms of the first \( 2np + 1 \) Chebyshev polynomials \( T_k(x) = \cos[k \arccos(x)] \) for \( k = 0, \ldots, 2np \); see, e.g., Stiefel (1963, p. 233). These polynomials, and the determinant in (A.3), are evaluated at \( x = 0 \), and at the \( 2np \) roots of \( T_{2np}(x) \), which are contained in the interval \( (-1, +1) \). The \( 2np + 1 \) coefficients of \( A_{ij}(x) = 0 \) are then computed by least squares, followed by a change of basis. For moderate values of \( np \), this procedure is numerically well-behaved due to the orthogonality of the Chebyshev polynomials.

The GAUSS function POLYROOT can be used to compute the roots of \( A_{ij}(x) = 0 \), but was not found numerically precise in some instances. Therefore, the previously described procedure was first applied to locate intervals containing roots of \( \rho[F_{ij}(x)] = \varepsilon \). A secant method was then applied in order to compute these roots with a high degree of precision.

The algorithm of this appendix was tested on problems involving randomly chosen \( M_\phi \) and \( V_\phi \); whenever possible, the code was checked by comparison with a naive rejection procedure. Independent chains with randomly chosen starting points appeared to give the best results, and satisfactory outcomes were obtained with chain lengths as low as 50.

Appendix B. Simulating negative-definite truncated normal matrices

Consider the generation of an \( n \times n \) matrix \( \Gamma \) with density:

\[ f[\gamma(\Gamma)] \propto f_{\text{SO}}(m, V) I_N(\Gamma), \]

where \( N \) is the set of all negative-definite matrices. Since \( N \) is convex and open, the Gibbs sampling algorithm described in Robert (1995) is immediately applicable, provided that one-dimensional ‘cuts’ of \( N \) are available. The purpose of this appendix is therefore the determination of bounds \( l_{ij} \) and \( u_{ij} \), dependent on all the elements of a negative-definite \( \Gamma \) save \( \gamma_{ij} \), such that \( \Gamma \) stays negative.
definite on the interval \( l_{ij} < \gamma_{ij} < u_{ij} \). An explicit solution is given by the following proposition.

**Proposition 2.** Let \( \Gamma \) be negative definite with elements \( \gamma_{ij} \), for \( i = 1, \ldots, j \) and \( j = 1, \ldots, n \) with \( n > 1 \). For any square matrix \( A \), let \( A_{-j,-j} \) be the submatrix of \( A \) obtained by deleting its \( j \)th row and column; let \( A_{-i,j} \) be the column vector obtained by deleting the \( i \)th element from the \( j \)th column of \( A \); and let \( A_{-i,j} \) be the column vector obtained by deleting elements \( i \) and \( j \) from the \( j \)th column of \( A \). If any of these selections is empty, replace it by zero. Let \( V^j = (\Gamma_{-j,-j})^{-1} \) and let \( V^j_{ij} \) be its \( i \)th diagonal element. If all the elements of \( \Gamma \) but \( \gamma_{jj} \) remain fixed, then \( \Gamma \) stays negative definite whenever

\[
\gamma_{jj} < (\Gamma_{-j,j})^j V^j (\Gamma_{-j,j}).
\] (B.1)

If all the elements of \( \Gamma \) but \( \gamma_{ij} \) (for \( i \neq j \)) remain fixed, then \( \Gamma \) stays negative definite over the interval

\[
l_{ij} < \gamma_{ij} < u_{ij},
\]

where \( l_{ij} \) and \( u_{ij} \) are the lower and upper roots of the following quadratic equation:

\[
\gamma_{ij}^2 V^j_{ii} + 2\gamma_{ij}(V^j_{-i,i})\Gamma_{-i,j} + [(\Gamma_{-ij,j})^j V^j_{ii} - (\Gamma_{-ij,j})^j \gamma_{jj}] = 0.
\] (B.2)

**Proof.** Let \( \gamma_{ij} \) (for \( i \leq j \)) be the element of \( \Gamma \) for which conditional truncation bounds must be computed. Let \( M \) be the matrix obtained by permutation of rows \( j \) and \( n \) and columns \( j \) and \( n \) of \( \Gamma \). Observe that \( M = P \Gamma P' \) for a suitable permutation matrix \( P \), so that \( M \) is negative definite if and only if \( \Gamma \) is negative definite. Also, observe that since \( \Gamma \) and \( M \) are symmetric, \( \gamma_{ij} \) becomes the element in row \( n \) and column \( i \), row \( i \) and column \( n \) of \( M \) when \( i \neq j \), and becomes the \( n \)th diagonal element of \( M \) when \( i = j \). The first \( n - 1 \) principal minors of \( M \) do not depend on \( \gamma_{ij} \), so that the negative definiteness of \( \Gamma \) is equivalent to \(( -1)^n \det M > 0 \). Let

\[
M = \begin{pmatrix} A & b \\ b' & a \end{pmatrix}
\]

with \( A \) the leading principal submatrix of order \( n - 1 \). It is well known that

\[
\det M = (\det A)(a - b'A^{-1}b).
\]

We note that \( A \) does not depend on \( \gamma_{ij} \) and is negative definite by assumption, so that \(( -1)^{n-1} \det A > 0 \). Hence, the sign of \(( -1)^n \det A \) is negative, and the negative definiteness of \( \Gamma \) is equivalent to \( a < b'A^{-1}b \). When \( i = j \), we have that
\[ a = \gamma_{jj}, \quad A^{-1} = V_j, \quad \text{and} \quad b = \Gamma_{-j,j}; \] inequality (B.1) follows. When \( i < j \), we first note that

\[
b'A^{-1}b = b^2\alpha_{ss} + 2b_s \sum_{k \neq s} a_{ks}b_k + \sum_{k \neq s, \ell \neq s} a_{k\ell}b_kb_\ell,
\]

where the superscripts denote elements of \( A^{-1} \). If \( b_s = \gamma_{ij} \), then \( \alpha_{ss} = V_{ij} \), the \( b_k \) (for \( k \neq s \)) are the elements of \( \Gamma_{-ij,j} \), the \( a_{ks} \) (for \( k \neq s \)) are the elements of \( V_{j-i} \), and the \( a_{k\ell} \) (for \( k, \ell \neq s \)) are the elements of \( V_{j-i} \). Eq. (B.2) follows since \( a = \gamma_{jj} \). For the left-hand side of (B.2) to be positive, \( \gamma_{ij} \) must lie between the roots of the quadratic, and Proposition 2 is proved. \( \square \)

**Appendix C. Metropolis–Hastings estimation of regular dynamic demand systems**

In this appendix, we extend the algorithm of Section 4 to the simulation of (7.16). Convergence follows from arguments similar to those in the last paragraph of Section 4, upon noting that \( \mathbf{N} \) is convex and open. Each chain is generated as follows:

**Step 1.** Set \( i = 1 \).

**Step 2.** Let \( \Phi = \Phi^{(i-1)} \), \( \Omega = \Omega^{(i-1)} \), \( \Gamma = \Gamma^{(i-1)} \); Compute \( V \) from (2.12), \( X_{**} \) and \( X_f \) by selection from (7.12), \( y_* \) as in (7.11), and \( y_{**} \) as in (7.18); draw \( \Pi_0^{(i)} \) from the normal density in (7.19).

**Step 3.** Let \( y_*, X_{**}, X_f \), and \( V \) be as in Step 2; compute \( y_f \) from (7.17) with \( \Pi_0 = \Pi_0^{(i)} \); using naive rejection sampling (when practical) or the Gibbs sampling algorithm of Appendix B, draw \( \Gamma^{(i)} \) from the truncated normal density in (7.20);

**Step 4.** Form the matrix:

\[
\Pi^{(i)} = (\Pi_0^{(i)} \Gamma^{(i)})
\]

and compute \( Y \) and \( X \) from (7.13) and (7.14) with

\[
z_t = \Pi^{(i)} x_t \quad \text{for} \quad t = p + 1, \ldots, p + T.
\]

**Step 5.** Using naive rejection sampling (when practical) or the Gibbs sampling algorithm of Appendix A, draw a candidate \( \Phi \) from

\[
f(\text{vec } \Phi) \propto f_{NO}[\text{vec } YX'(XX')^{-1}, (XX')^{-1} \otimes \Omega^{(i-1)}]I_S(\Phi),
\]

set \( \Phi^{(i)} = \Phi \) with probability

\[
\alpha = \min \left[ \frac{L_*(\Pi^{(i)}, \Phi, \Omega^{(i-1)})}{L_*(\Pi^{(i)}, \Phi^{(i-1)}, \Omega^{(i-1)})}, 1 \right]
\]

and set \( \Phi^{(i)} = \Phi^{(i-1)} \) with probability \( 1 - \alpha \).
Step 6. Draw a candidate \( \Omega \) from
\[
f_{w}[T, (Y - \Phi^{(i)}X)(Y - \Phi^{(i)}X)^\prime],
\]
set \( \Omega^{(i)} = \Omega \) with probability
\[
\beta = \min\left[ \frac{\mathcal{L}_*(\Pi^{(i)}, \Phi^{(i)}, \Omega)}{\mathcal{L}_*(\Pi^{(i)}, \Phi^{(i)}, \Omega^{(i-1)})}, 1 \right]
\]
and set \( \Omega^{(i)} = \Omega^{(i-1)} \) with probability \( 1 - \beta \).

Step 7. Set \( i = i + 1 \) and return to Step 2.

Suitable random starting values are, as in Section 5, drawn from a thick-tailed approximation to the joint posterior density. Specifically, we iterate on the parameterizations:
\[
y_t - \Pi X_t = \sum_{j=1}^{p} \Phi_j(y_{t-j} - \Pi X_t) + \varepsilon_t,
\]
\[
\Phi(L)y_t = [x_t \otimes \Phi(1)]D_S\left( \begin{array}{c} \text{vec } \Pi_0 \\ \text{vec } \Gamma \end{array} \right) + \varepsilon_t
\]
(C.1)

(C.2)
to compute generalized least-squares estimates \( \hat{\Phi}, \hat{\Pi}_0, \hat{\Gamma} \). We note that \( \hat{\Gamma} \) is symmetric, and let
\[
\hat{\Pi} = (\hat{\Pi}_0 \hat{\Gamma}),
\]
\[
\hat{\varepsilon}_t = \hat{\Phi}(L)y_t - \hat{\Phi}(1)(\hat{\Pi} X_t),
\]
\[
\hat{Q} = \frac{1}{T - n(p + 1) - 2} \sum_{t=p+1}^{p+T} \hat{\varepsilon}_t \hat{\varepsilon}_t^\prime.
\]
(C.4)
(C.5)
The counterpart of Eq. (5.4) becomes
\[
f(\text{vec } \hat{\Pi}_0, \text{vec } \hat{\Gamma}, \text{vec } \hat{\Phi}) \approx f_{NO}\left( \begin{array}{c} \text{vec } \Pi_0 \\ \text{vec } \Gamma \\ \text{vec } \Phi \end{array} \right),
\]
(C.6)
where as can be shown (see Deschamps, 1997, Section 8):
\[
D = \begin{pmatrix} D_S & O \\ O & I_{n^2p} \end{pmatrix},
\]
\[
\hat{\mathcal{J}} = \sum_{t=p+1}^{p+T} \begin{pmatrix} Z_{1t}^\prime \\ Z_{2t}^\prime \end{pmatrix} \hat{Q}^{-1}(Z_{1t}^\prime, Z_{2t}^\prime),
\]
\[ Z_{1t} = x_t \otimes \hat{\Phi}(1), \]
\[ Z_{2t} = \begin{bmatrix} y_{t-p} - \hat{H}x_t \\ \vdots \\ y_{t-1} - \hat{H}x_t \end{bmatrix} \otimes I_n. \]

It proved impractical to sample a truncated studentized version of (C.6) by naive rejection. We therefore chose our starting values \( \Pi_0^{(0)}, \Gamma^{(0)}, \Phi^{(0)}, \) and \( \Omega^{(0)} \) as random drawings from the following densities:

\[
p(\Omega) = f_{IW}(T - np - 1, \sum_{t=p+1}^{p+T} \hat{e}_t \hat{e}_t^\top), \tag{C.7}
\]
\[
p(\text{vec} \Pi_0, \text{vec} \Phi) \propto f_{ST}\left(\begin{bmatrix} \text{vec} \hat{H}_0 \\ \text{vec} \hat{\Phi} \end{bmatrix}, V_1, 3 \right) I_s(\Phi), \tag{C.8}
\]
\[
p(\text{vec} \Gamma) \propto f_{ST}[\text{vec}(\hat{\Gamma}), V_2, 3] I_N(\Gamma) \tag{C.9}
\]

with \( V_1 \) and \( V_2 \) appropriate scale matrices, derived from the marginal distributions implied by (C.6). We use naive rejection to simulate (C.8). The truncated Student density in (C.9) may be sampled by adapting the method proposed by Geweke (1991) and using the results in Appendix B. When negativity is not imposed, we may replace Steps 2 and 3 by:

**Step 2 + 3.** Set \( \Phi = \Phi^{(i-1)} \) and \( \Omega = \Omega^{(i-1)} \). Compute \( V \) from (2.12), \( y_* \) and \( X_* \) from (7.11) and (7.12). Draw \( \Pi^{(i)} \) from

\[ f_{NO}[(X_* V^{-1} X_*)^{-1} X_* V^{-1} y_*, (X_* V^{-1} X_*)^{-1}]. \]

**Appendix D. Estimating a posterior density ordinate**

Since the technique is very similar for the three versions (homogeneous, symmetric, and regular) of the differential CBS model, we will only discuss the fully regular case. We let \( H \equiv \Omega^{-1} \), and denote by \( (\Pi_0^*, \Gamma^*, \Phi^*, H^*) \) the matrices containing the means of the posterior replications (based on all \( T + p \) observations) of \( (\Pi_0, \Gamma, \Phi, \Omega^{-1}) \). For notational convenience, let \( p(\Pi_0, \Gamma, \Phi, \Omega) \equiv p_N(\Pi_0, \Gamma, \Phi, \Omega \mid \mathcal{Y}, \mathcal{X}) \) or \( p_N(\Pi_0, \Gamma, \Phi, \Omega \mid \mathcal{Y}_*, \mathcal{X}_*) \), as the case may be. We may write

\[
p(\Pi_0^*, \Gamma^*, \Phi^*, \Omega^*) = p(H^*) p(\Phi^* \mid H^*) p(\Pi_0^* \mid \Phi^*, H^*)
\]
\[
p(\Gamma^* \mid \Pi_0^*, \Phi^*, H^*)(\det H^*)^{n+1}, \tag{D.1}
\]

where \( (\det H^*)^{n+1} \) is the Jacobian of the transformation.

In view of (7.20), the last density in this product is truncated normal, and its normalizing constant can be precisely estimated by naive rejection.
The third density can be written
\begin{equation}
p(P^*_0 | \Phi^*, H^*) = \int p(P^* | \Gamma, \Phi^*, H^*) p(\Gamma | \Phi^*, H^*) \, d\Gamma. \quad (D.2)
\end{equation}

In view of (7.19), (D.2) can be estimated as an average of normal density ordinates over \( G \) replications \((P^*_0, \Gamma^j)\) obtained by reduced Markov chains conditional on \( \Phi^* \) and \( H^* \):
\begin{equation}
\hat{p}(P^*_0 | \Phi^*, H^*) = \frac{1}{G} \sum_{j=1}^{G} p(P^*_0 | \Gamma^j, \Phi^*, H^*). \quad (D.3)
\end{equation}

A similar technique is unavailable for the first two terms in (D.1), which must therefore be estimated by a multivariate kernel method; see Chib and Greenberg (1998) for details. Since kernel methods are known to be inaccurate for problems involving more than 4 or 5 dimensions, we had to further decompose the first two terms in (D.1). Upon denoting by \( U^* \) the \( k \)th column of \( U \), we may write
\begin{equation}
p(U^* | H^*) = p(U^*_1 | \Phi^*_2, \Phi^*_3, \Phi^*_4, H^*) p(U^*_2 | \Phi^*_3, \Phi^*_4, H^*) p(U^*_3 | \Phi^*_4, H^*) p(U^*_4 | \Phi^*_1, H^*) \quad (D.4)
\end{equation}

and note that for \( k = 1, 2, 3 \), the conditional density of \( U^*_k \) can be estimated by a multivariate kernel method from the simulation output of reduced Markov chains on \((P^*_0, \Gamma, \Phi_j (j \leq k))\), for given \( H^* \) and \( \Phi^*_j (j > k) \). Suitable conditional normal densities replace the unconditional normal in (7.22).

Finally, upon letting
\begin{equation}
H^* = \begin{pmatrix} H^*_{11} & H^*_{12} \\ H^*_{21} & H^*_{22} \end{pmatrix},
\end{equation}

where each \( H^*_k \) is a \( 2 \times 2 \) matrix, the first term in (D.1) can be decomposed as follows:
\begin{equation}
p(H^*) = p(H^*_{22}) p(H^*_{12} | H^*_{22}) p(H^*_{11} | H^*_{12}, H^*_{22}). \quad (D.6)
\end{equation}

From (7.21), the full conditional of \( H \) is proportional to \( \mathcal{L}_* \) times a Wishart density with \( T - 5 \) degrees of freedom and scale matrix
\[ \mathcal{S} = [(Y - \Phi X)(Y - \Phi X)^\top]^{-1} = \begin{pmatrix} \mathcal{S}_{11} & \mathcal{S}_{12} \\ \mathcal{S}_{21} & \mathcal{S}_{22} \end{pmatrix}, \]

where the partition of \( \mathcal{S} \) corresponds to the partition in (D.5). Let
\[ \mathcal{S}_{11.2} = \mathcal{S}_{11} - \mathcal{S}_{12} \mathcal{S}_{22}^{-1} \mathcal{S}_{21}. \]

A result proved by Muirhead (1982, p. 93, Theorem 3.2.10) implies that
\begin{equation}
p(H_{11}, \text{vec} H_{12} | \Pi, \Phi, H^*_{22}) \propto \mathcal{L}_* f_W(H_{11} - H_{12} (H^*_{22})^{-1} H_{21}; T - 7, \mathcal{S}_{11.2})
\end{equation}
\[ f_{NO}(\text{vec } H_{12}; \text{vec } H_{22}^g \mathcal{S}_{21}^{-1} \mathcal{S}_{21}, \mathcal{S}_{11.2} \otimes H_{22}^g), \quad (D.7) \]
\[ p(H_{11} | \Pi, \Phi, H_{12}^g, H_{22}^g) \propto \mathcal{L}_* f_w(H_{11} - H_{12}^g(H_{22}^g)^{-1}H_{21}^g; T - 7, \mathcal{S}_{11.2}), \quad (D.8) \]

where \( \mathcal{L}_* \) is given by (7.23) with the appropriate arguments replaced by their conditioning values, where \( f_w(\cdot; \nu, \mathcal{T}) \) denotes the Wishart density with \( \nu \) degrees of freedom and scale matrix \( \mathcal{T} \), and where \( f_{NO}(\cdot; \mu, \Sigma) \) is the multinormal density.

It is then straightforward to estimate the three terms in (D.6) by a multivariate kernel method. \( p(H_{22}^g) \) is estimated from the simulation output for the joint density of \((P, U, H_{11}, H_{12}, H_{22})\). \( p(H_{12}^g | H_{22}^g) \) is estimated using reduced Markov chains on \((P, U, H_{11}, H_{12})\), for given \( H_{22}^g \): in Step 6 of Appendix C, \( H_{12} \) is drawn from the normal density in (D.7), a 2 \times 2 matrix \( W \) is drawn from the Wishart density in (D.7), \( H_{11} \) is computed as \( W + H_{12}(H_{22}^g)^{-1}H_{12} \), and the candidate becomes

\[ \Omega = \begin{pmatrix} H_{11} & H_{12} \\ H_{12}^T & H_{22}^g \end{pmatrix}^{-1}. \]

Similarly, \( p(H_{11}^g | H_{12}^g, H_{22}^g) \) is estimated using reduced Markov chains on \((P, \Phi, H_{11}^g)\), for given \( H_{12}^g \) and \( H_{22}^g \).

For a four-equation model with first-order dynamics, estimating the marginal likelihood in (8.5) thus necessitates the Markov chain Monte Carlo simulation of 16 different densities when \( k = N \), and 14 different densities when \( k = H \) or \( S \).

References