

Bayesian Inference on Time-Varying Proportions

William J. McCausland *Université de Montréal, CIREQ and CIRANO

Brahim Lgui †Université de Montréal

Current version: October 31, 2007

Abstract

Time varying proportions arise frequently in economics. Market shares show the relative importance of firms in a market. Labor economists divide populations into different labor market segments. Expenditure shares describe how consumers and firms allocate total expenditure to various categories of spending. We introduce a state space model where unobserved states are Gaussian and observations are conditionally Dirichlet. Markov chain Monte Carlo techniques allow inference for unknown parameters and states. We can draw states as a block using a multivariate normal proposal distribution based on a quadratic approximation of the log conditional density of states given parameters and data. Repeated draws from the proposal distribution are particularly efficient.

Key words: State space models, Proportions data, Time varying parameters.

*Mailing address: Département de sciences économiques, C.P. 6128, succursale Centre-ville, Montréal QC H3C 3J7, Canada. e-mail: william.j.mccausland@umontreal.ca. Web site: www.cirano.qc.ca/~mccauslw.

†e-mail: brahim.lgui@umontreal.ca.

1 Introduction

A vector $\pi_t \equiv (\pi_{t1}, \dots, \pi_{tp})$ of *proportions* observed at time t satisfies $\pi_{ti} \geq 0$ for all $i \in \{1, \dots, p\}$ and $\sum_{i=1}^p \pi_{ti} = 1$. Time varying proportions arise frequently in economics. Market shares show the relative importance of firms in a market. Labor economists divide the working age population into different labor market segments: people are either in the labor force or not, those in the labor force are either employed or unemployed, and the employed may be working full time or part time. Also relevant to labor economists is a person's allocation of time to work, child care, home production, leisure and other pursuits. Expenditure shares describe how consumers and firms allocate total expenditure to various categories of spending. Probabilities are also proportions, and we can study the probabilities of political events implied by market prices of political futures contracts.

Varying proportions also arise outside of economics. Biologists track the time-varying proportions of species in a habitat, such as a lake. Geologists study the variation of the composition of rocks in a spatial dimension.

Modelling time-varying proportions allows us to learn about the joint dynamics of proportions and other observed data, to measure the response of proportions to changes in exogenous variables, and to make predictions.

State space models have proved useful for exploring many kinds of dynamic relationships. With proportions data, difficulties arise because of the positivity and adding-up constraints.

The most common resolution of this difficulty involves modelling suitably transformed data. Aitchison (1986) introduces two different transformations, both of which have been used extensively. The *additive log ratio* (alr) transformation of a proportion vector π is

$$(\log \pi_1 - \log \pi_p, \dots, \log \pi_{p-1} - \log \pi_p).$$

This has the advantage of being a bijective map from the simplex to \mathbb{R}^{p-1} , but it is not symmetric: the transformation depends on the choice of a reference proportion, here π_p .

The *centered log ratio* (clr) transformation is

$$\left(\log \pi_1 - p^{-1} \sum_{i=1}^p \log \pi_i, \dots, \log \pi_p - p^{-1} \sum_{i=1}^p \log \pi_i \right).$$

This transformation is symmetric, but the image of the transformation is $\{x \in \mathbb{R}^p : \sum_{i=1}^p x_i = 0\}$, which is more difficult to work with.

Whatever the transformation used, models of transformed data are more difficult to interpret than models for untransformed proportions. Grunwald, Raftery, and Guttorp (1993) introduce a state-space model where the conditional distribution of proportions given latent states is explicit. They use the Dirichlet distribution, a natural distribution for proportions since its support is the simplex: if π has a Dirichlet distribution with parameter vector $\gamma \equiv (\gamma_1, \dots, \gamma_p) > 0$, then $(\pi_1, \dots, \pi_{p-1})$ has density

$$f(\pi_1, \dots, \pi_{p-1}; \gamma) = \begin{cases} \frac{\Gamma(G)}{\prod_{i=1}^p \Gamma(\gamma_i)} \prod_{i=1}^p \pi_i^{\gamma_i-1} & \pi_i \geq 0 \text{ for all } i \in \{1, \dots, p\} \\ 0 & \text{otherwise,} \end{cases}$$

where $\pi_p = 1 - \sum_{i=1}^{p-1} \pi_i$ and $G \equiv \sum_{i=1}^p \gamma_i$. Another advantage of the Dirichlet distribution is that the mean is a transparent function of the parameters:

$$E[\pi] = \frac{1}{G} \gamma.$$

However, the Dirichlet distribution is fairly inflexible: once we choose the ratios γ_i/γ_j to set the mean of the distribution, only the sum G remains as a degree of freedom. We can interpret G as an overall precision parameter. We cannot freely choose covariances or even individual variances. The elements of π are *neutral*, meaning that for all i and j , π_i and $\pi_j/(1 - \pi_i)$ are independent.

In state space models, state dynamics can compensate for the inflexibility of the Dirichlet distribution: while the conditional distribution of proportions given unobserved states is Dirichlet, the unconditional distribution of proportions may be much richer. In Grunwald, Raftery, and Guttorp (1993), state dynamics follow the *Dirichlet conjugate distribution*, defined in that paper.

Here we introduce a state space model where observations are conditionally Dirichlet and latent state transitions are Gaussian. Thus our method has the same

advantage as that of Grunwald, Raftery, and Guttorp (1993), the transparency of the conditional distribution of proportions given states. This facilitates the interpretation of state dynamics.

Gaussian state transitions have several advantages over the Dirichlet conjugate distribution. They are easier to interpret, more flexible and more familiar. Finally, we do not have to settle for approximations of unknown normalization constants.

Durbin and Koopman (1997) describe an approach to inference in state space models with Gaussian states and non-Gaussian observations. Given such a semi-Gaussian state space model, they construct, as a device, a fully Gaussian state space model where the posterior distribution of the latent states approximates that of the semi-Gaussian model. While they use this distribution as an importance distribution to compute the likelihood function using importance sampling, it could just as well be used as a proposal distribution in an MCMC method for posterior simulation. (Likewise, the proposal distributions we propose here could be used as an importance distribution.)

What we do in this paper is similar to what Durbin and Koopman (1997) do in one respect but quite different in another. Like Durbin and Koopman (1997), we approximate the conditional posterior distribution of states as a multivariate normal distribution. We could also frame this in terms of a fully Gaussian state space model as Durbin and Koopman (1997) do. The basic proposal distribution we describe below is identical to what the Durbin and Koopman (1997) approach gives for the state space model we propose. We use an entirely different method for drawing states, however. Our procedure is not based on the Kalman filter, but rather the algorithm described in McCausland, Miller, and Pelletier (2007). McCausland, Miller, and Pelletier (2007) show that for many important cases, their method is more computationally efficient than methods, such as that of Durbin and Koopman (1997), that use the Kalman filter. The McCausland, Miller, and Pelletier (2007) algorithm is particularly efficient for repeated draws.

In Section 2, we describe in detail a state space model for time-varying proportions. Then we describe Markov chain Monte Carlo methods for simulating states and parameters from their joint posterior distribution. We draw states and parameters in separate Gibbs blocks. We show how to draw latent states as a block from a proposal distribution approximating their conditional posterior dis-

tribution in Section 3. We show how to draw parameters from their conditional posterior distribution in Section 4, using a fairly standard approach. We illustrate our methods using an empirical example in Section 5. We conclude in Section 6.

2 The Model

We use state space models to describe the dynamics of time-varying proportions. There are p latent states, and the state vector α_t is a first order Gaussian vector autoregressive process. The observed vectors π_t of proportions are conditionally independent and Dirichlet given the state process.

The state vector process $\{\alpha_t\}_{t=1}^{\infty}$ is defined by

$$\alpha_1 \sim N(\mu_1, H_1^{-1}) \quad \alpha_t | \alpha_1, \dots, \alpha_{t-1} \sim N(\delta + \Phi \alpha_{t-1}, H_\alpha^{-1}).$$

For the purposes of inference, we can distinguish two cases. In the non-stationary case, the parameter vector is $\theta \equiv (\mu_1, H_1, \delta, \Phi, H_\alpha)$. In the stationary case, the parameter vector is $\theta \equiv (\delta, \Phi, H_\alpha)$, and μ_1 and H_1 are the following functions of δ , Φ and H_α :

$$\mu_1 = (I - \Phi)^{-1} \delta \quad \text{vec } H_1^{-1} = (I_{p^2} - \Phi \otimes \Phi)^{-1} \text{vec } H_\alpha^{-1}.$$

See Hamilton (1994, p.265) for details on the computation of H_1^{-1} .

Given the state sequence, the π_t are conditionally independent and each π_t has a Dirichlet distribution with parameter vector given by $\gamma_t = X_t \beta + Z_t \alpha_t$, where X_t is a $p \times k$ matrix of covariates, β is a $k \times 1$ vector of coefficients and Z_t is a $p \times m$ matrix. This allows for covariates (X_t) in the observation distribution and factor models (Z_t gives factor loadings).

Thus we have the following conditional density of π given states:

$$f(\pi_t | \alpha_1, \dots) = \frac{\Gamma(\sum_{i=1}^m \gamma_{ti})}{\prod_{i=1}^m \Gamma(\gamma_{ti})} \prod_{i=1}^m \pi_{ti}^{\gamma_{ti}-1}. \quad (1)$$

The notation for the observation distribution corresponds to the notation used by de Jong and Shephard (1995) for fully Gaussian state space models. We take X_t , β and Z_t as fixed and do not consider inference for β or Z_t here.

3 Drawing states

For posterior inference, we use a Metropolis-Hastings-Greene chain (Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller (1953), Hastings (1970), Green (1995)) whose stationary distribution is the joint posterior distribution of θ and α . In this section, we describe an update block for α , so the target distribution is the conditional distribution of α given θ and π . We describe a proposal distribution approximating this target distribution.

The basic proposal has a multivariate normal distribution based on a quadratic approximation of the log target density $\log f(\alpha|\theta, \pi)$ at the mode α° of the target distribution. We can write

$$\log f(\alpha|\theta, \pi) = \log f(\alpha|\theta) + \log f(\pi|\alpha) + k,$$

where k is a term not depending on α .

The first term is already quadratic in α . It is convenient to write it in terms of the precision \bar{H} and covector \bar{c} of the conditional distribution of α given θ , rather than the more familiar mean and variance. This is because precisions and covectors combine additively when quadratic forms are summed. Therefore we write

$$\log f(\alpha|\theta) = \frac{1}{2} [\log |\bar{H}| - np \log 2\pi - \alpha^\top \bar{H} \alpha + 2\bar{c}^\top \alpha + \bar{c} \bar{H}^{-1} \bar{c}]. \quad (2)$$

We compute

$$\bar{H} = \begin{bmatrix} \bar{H}_{11} & \bar{H}_{12} & 0 & \cdots & 0 & 0 \\ \bar{H}_{21} & \bar{H}_{22} & \bar{H}_{23} & \cdots & 0 & 0 \\ 0 & \bar{H}_{32} & \bar{H}_{33} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \bar{H}_{n-1,n-1} & \bar{H}_{n-1,n} \\ 0 & 0 & 0 & \cdots & \bar{H}_{n,n-1} & \bar{H}_{n,n} \end{bmatrix}, \quad \bar{c} = \begin{bmatrix} \bar{c}_1 \\ \bar{c}_2 \\ \bar{c}_3 \\ \vdots \\ \bar{c}_{n-1} \\ \bar{c}_n \end{bmatrix}, \quad (3)$$

where

$$\begin{aligned} \bar{H}_{11} &= H_1 + \Phi' H_\alpha \Phi, & \bar{H}_{nn} &= H_\alpha, \\ \bar{H}_{tt} &= H_\alpha + \Phi' H_\alpha \Phi \text{ for } t = 2, \dots, n-1, \end{aligned}$$

$$\begin{aligned}\bar{H}_{t-1,t} &= -\Phi' H_\alpha \text{ and } \bar{H}_{t,t-1} = -H_\alpha \Phi \text{ for } t = 1, \dots, n-1, \\ \bar{c}_1 &= H_1 \mu_1 - \Phi' H_\alpha \delta, \quad \bar{c}_n = H_n \delta, \\ \bar{c}_t &= (I - \Phi)' H_\alpha \delta \text{ for } t = 2, \dots, n-1.\end{aligned}$$

The second term, $\log f(\pi|\alpha)$, is not quadratic but we can approximate it as such. A second order Taylor expansion of $\log f(\pi|\alpha)$ around some reference value α^\bullet gives

$$\log f(\pi|\alpha) \approx -\frac{1}{2} [\alpha^\top H(\alpha^\bullet) \alpha - 2c(\alpha^\bullet)^\top \alpha] + k', \quad (4)$$

where

$$H(\alpha^\bullet) \equiv \begin{bmatrix} h_1(\alpha^\bullet) & 0 & \cdots & 0 \\ 0 & h_2(\alpha^\bullet) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_n(\alpha^\bullet) \end{bmatrix}, \quad c(\alpha^\bullet) \equiv \begin{bmatrix} c_1(\alpha^\bullet) \\ \vdots \\ c_n(\alpha^\bullet) \end{bmatrix},$$

and for $t = 1, \dots, n$,

$$h_t(\alpha^\bullet) = -\frac{\partial^2 \log f(\pi_t|\alpha_t^\bullet)}{\partial \alpha_t \partial \alpha_t'}, \quad (5)$$

$$c_t(\alpha^\bullet) = \frac{\partial \log f(\pi_t|\alpha_t^\bullet)}{\partial \alpha_t} - \frac{\partial^2 \log f(\pi_t|\alpha_t^\bullet)}{\partial \alpha_t \partial \alpha_t'} \alpha_t^\bullet, \quad (6)$$

and k' is a term not depending on α . We derive expressions for $h_t(\alpha^\bullet)$ and $c_t(\alpha^\bullet)$ in Appendix A.

Combining the quadratic expression for $\log f(\alpha|\theta)$ in (2) with the quadratic approximation of $\log f(\pi|\alpha)$ in (4), we obtain the following quadratic approximation of the log conditional posterior density of α :

$$\log f(\alpha|\theta, r) \approx -\frac{1}{2} [\alpha^\top \bar{\bar{H}}(\alpha^\bullet) \alpha - 2\bar{\bar{c}}(\alpha^\bullet)^\top \alpha] + k'',$$

where $\bar{\bar{H}}(\alpha^\bullet) \equiv \bar{H} + H(\alpha^\bullet)$, $\bar{\bar{c}}(\alpha^\bullet) \equiv \bar{c} + c(\alpha^\bullet)$ and k'' is a term not depending on α .

Since the mode α° of the conditional posterior density maximizes $\log f(\alpha|\theta, \pi)$, $\bar{\bar{H}}(\alpha^\circ)$ must be positive definite and therefore the approximation of $\log f(\alpha|\theta, \pi)$ at α° is the log of a multivariate Gaussian density.

3.1 Implementation

Drawing from the proposal distribution involves finding α° , $\bar{\bar{H}}(\alpha^\circ)$ and $\bar{\bar{c}}(\alpha^\circ)$ and then drawing $\alpha^* \sim N(\bar{\bar{H}}(\alpha^\circ)^{-1}\bar{\bar{c}}(\alpha^\circ), \bar{\bar{H}}(\alpha^\circ)^{-1})$. The quantities α° , $\bar{\bar{H}}(\alpha^\circ)$ and $\bar{\bar{c}}(\alpha^\circ)$ satisfy $\alpha^\circ = \bar{\bar{H}}(\alpha^\circ)^{-1}\bar{\bar{c}}(\alpha^\circ)$, since $\bar{\bar{H}}(\alpha^\circ)^{-1}\bar{\bar{c}}(\alpha^\circ)$ is the mean, and therefore mode, of the Gaussian approximation with precision $\bar{\bar{H}}(\alpha^\circ)$ and covector $\bar{\bar{c}}(\alpha^\circ)$. We find these values by iterating the following computation until numerical convergence:

$$\alpha^{\text{new}} := \bar{\bar{H}}(\alpha^{\text{old}})^{-1}\bar{\bar{c}}(\alpha^{\text{old}}).$$

It sometimes happens that $\bar{\bar{H}}(\alpha^{\text{old}})$ is not positive definite, in which case we replace the step with

$$\alpha^{\text{new}} := \bar{\bar{H}}^{\text{safe}}(\alpha^{\text{old}})^{-1}\bar{\bar{c}}^{\text{safe}}(\alpha^{\text{old}}),$$

where $\bar{\bar{H}}^{\text{safe}}(\alpha^{\text{old}})$ is positive definite, $\bar{\bar{H}}^{\text{safe}}(\alpha^{\text{old}}) - \bar{\bar{H}}(\alpha^{\text{old}})$ is positive semi-definite, and the gradient of the quadratic approximation $\alpha^\top \bar{\bar{H}}^{\text{safe}}(\alpha^{\text{old}})\alpha_t - 2\alpha^\top \bar{\bar{c}}^{\text{safe}}(\alpha^{\text{old}})$ is still equal to the gradient of $\log f(\pi|\alpha)$ at α^{old} . We construct $\bar{\bar{H}}^{\text{safe}}(\alpha^{\text{old}})$ $\bar{\bar{c}}^{\text{safe}}(\alpha^{\text{old}})$ by replacing $h_t(\alpha_t^{\text{old}})$ and $c_t(\alpha_t^{\text{old}})$ with $h_t^{\text{safe}}(\alpha_t^{\text{old}})$ and $c_t^{\text{safe}}(\alpha_t^{\text{old}})$, whose precise definitions are given in Appendix A.

We use the algorithm described in McCausland, Miller, and Pelletier (2007) to solve for α in the equation $\bar{\bar{H}}\alpha = \bar{\bar{c}}$. The steps are the following, where $\bar{\bar{H}}_{tt}$, $\bar{\bar{H}}_{t,t+1}$ and $\bar{\bar{c}}_t$ are defined by partitions of $\bar{\bar{H}}$ and $\bar{\bar{c}}$ analogous to those of \bar{H} and \bar{c} in (3):

1. Compute $\Sigma_1 = (\bar{\bar{H}}_{11})^{-1}$, $m_1 = \Sigma_1\bar{\bar{c}}_1$.
2. For $t = 2, \dots, n$, compute

$$\Sigma_t = (\bar{\bar{H}}_{tt} - \bar{\bar{H}}_{t,t-1}\Sigma_{t-1}\bar{\bar{H}}_{t-1,t})^{-1}, \quad m_t = \Sigma_t(\bar{\bar{c}}_t - \bar{\bar{H}}_{t,t-1}m_{t-1}).$$

3. Compute $\alpha_n = m_n$.
4. For $t = n - 1, \dots, 1$, compute

$$\alpha_t = m_t - \Sigma_t\bar{\bar{H}}_{t,t+1}\alpha_{t+1}.$$

In practice, we never actually construct the Σ_t , only the Σ_t^{-1} and their Cholesky factors. This is more efficient and delivers everything we need.

Once we have α° , μ_t and the Cholesky decompositions of the Σ_t^{-1} , we use the algorithm in McCausland, Miller, and Pelletier (2007) to draw the proposal α^* as follows:

1. Draw $\alpha_n \sim N(m_n, \Sigma_n)$.
2. For $t = n - 1, \dots, 1$, draw

$$\alpha_t \sim N(m_t - \Sigma_t \bar{\bar{H}}_{t,t+1} \alpha_{t+1}, \Sigma_t).$$

We accept the proposal with probability

$$\min \left(1, \frac{f(\alpha^*|\theta)f(\pi|\alpha^*)}{f(\alpha|\theta)f(\pi|\alpha)} \cdot \frac{g(\alpha)}{g(\alpha^*)} \right),$$

where g is the proposal density. We evaluate $g(\alpha^*)$ as

$$g(\alpha^*) = g(\alpha_n^*)g(\alpha_{n-1}^*|\alpha_n^*) \cdots g(\alpha_1^*|\alpha_2^*, \dots, \alpha_n^*)$$

using the conditional means and variances computed using the algorithm above. To evaluate $g(\alpha)$ we use the same algorithm to compute the appropriate conditional means and variances, but without drawing any values.

We note that multiple draws for a given value of θ are particularly efficient. For each t , we only need to compute $\Sigma_t \bar{\bar{H}}_{t,t+1}$ and the Cholesky decomposition of Σ_t once. Drawing α_t only requires two operations of order $O(m^2)$: premultiplying α_{t+1} by $(\Sigma_t \bar{\bar{H}}_{t,t+1})$ and premultiplying a vector of independent normals by the Cholesky decomposition of Σ_t . The methods of de Jong and Shephard (1995) and Durbin and Koopman (2002) are both more computationally demanding. See McCausland, Miller, and Pelletier (2007) for a detailed comparison of computational costs for repeated draws of α .

4 Conditional Inference for parameters

In the non-stationary case, we use standard methods exploiting conditionally conjugate priors to draw the parameters from their exact conditional posterior distri-

butions. All conditional distributions below are well known and we can draw from them using standard methods.

We suppose that μ_1 , H_1 , (δ, Φ) and H_α are *a priori* independent. We let

$$a_+ \equiv \text{vec} \begin{bmatrix} \delta^\top \\ \Phi^\top \end{bmatrix}.$$

The initial mean parameter μ_1 is multivariate normal:

$$\mu_1 \sim N(\bar{r}_1, \bar{P}_1^{-1}),$$

and its conditional conjugacy leads to the following conditional posterior distribution:

$$\mu_1 | \alpha, H_1, \delta, \Phi, H_\alpha \sim N(\bar{r}_1, \bar{P}_1^{-1}),$$

where

$$\bar{P}_1 \equiv \bar{P}_1 + H_1 \quad \text{and} \quad \bar{r}_1 \equiv \bar{P}_1^{-1} (\bar{P}_1 \bar{r}_1 + H_1 \alpha_1).$$

The initial precision parameter H_1 is Wishart:

$$H_1 \sim \text{Wi}(\bar{\nu}_1, \bar{A}_1),$$

giving the following conditional posterior distribution:

$$H_1 | \alpha, \mu_1, \delta, \Phi, H_\alpha \sim \text{Wi}(\bar{\nu}_1, \bar{A}_1),$$

where

$$\bar{\nu}_1 = \bar{\nu} + 1 \quad \text{and} \quad \bar{A}_1^{-1} = \bar{A}_1^{-1} + (\alpha_1 - \mu_1)(\alpha_1 - \mu_1)^\top.$$

With priors for μ_1 and H_1 , there is more than one way to express a large degree of prior uncertainty about α_1 . One way is to choose fairly precise priors for μ_1 and H_1 , favoring values of H_1 with small eigenvalues. Such values of H_1 imply a high degree of uncertainty about α_1 . In the limit, one can choose fixed values for μ_1 and H_1 . Another way is to choose an imprecise prior for μ_1 , and a precise prior for H_1 favoring values of H_1 with large eigenvalues. In the choice between the two, there is a computational trade-off. Values of H_1 with large eigenvalues ensure that \bar{H} , and thus $\bar{\bar{H}}$, do not have any small eigenvalues. This leads to faster

convergence to α° and a conditional distribution of α given π and θ that is more nearly Gaussian. On the other hand, values of H_1 with large eigenvalues lead to higher posterior correlation between α_1 and μ_1 , which implies slower mixing.

The coefficient vector a_+ is multivariate normal:

$$a_+ \sim N(\bar{r}, \bar{P}^{-1}),$$

leading to the following conditional posterior distribution:

$$a_+ | \alpha, \mu_1, H_1, H_\alpha \sim N(\bar{r}, \bar{P}^{-1}),$$

where

$$\begin{aligned} \bar{P} &\equiv \bar{P} + H_\alpha \otimes \begin{bmatrix} n & \sum_{t=2}^n \alpha_{t-1}^\top \\ \sum_{t=2}^n \alpha_{t-1} & \sum_{t=2}^n \alpha_{t-1} \alpha_{t-1}^\top \end{bmatrix}, \\ \bar{r} &\equiv \bar{P}^{-1} \left(\bar{P} \bar{r} + \text{vec} \left(\begin{bmatrix} \sum_{t=2}^n \sum_{i=1}^m \alpha_{ti} \\ \sum_{t=2}^n \alpha_{t-1} \sum_{i=1}^m \alpha_{ti} \end{bmatrix} H_\alpha \right) \right). \end{aligned}$$

The state conditional precision parameter H_α has the following Wishart prior:

$$H_\alpha \sim \text{Wi}(\bar{\nu}_\alpha, \bar{A}_\alpha),$$

so that its conditional posterior distribution is

$$H_\alpha | \alpha, \mu_1, H_1, \delta, \Phi \sim \text{Wi}(\bar{\nu}, \bar{A}_\alpha),$$

where

$$\bar{\nu}_\alpha = \bar{\nu}_\alpha + n - 1 \quad \text{and} \quad \bar{A}_\alpha^{-1} = \bar{A}_\alpha^{-1} + \sum_{t=2}^n (\alpha_t - \delta - \Phi \alpha_{t-1})(\alpha_t - \delta - \Phi \alpha_{t-1})^\top.$$

5 An Empirical Example

As an empirical example, we analyze prices in a political futures market. We use publicly available data from the Iowa Electronic Markets (IEM), which are futures markets for contracts whose payoffs are contingent on economic and political events. The web site of IEM is <http://www.biz.uiowa.edu/iem/>.

We use data from the market DCONV08, based on the result (unknown at the time of writing) of the 2008 Democratic Party nomination of candidate for President of the United States. There are four contracts. CLIN_NOM pays one dollar if Hillary Clinton wins the nomination, and nothing otherwise. OBAM_NOM and EDWA_NOM are similar contracts for Barack Obama and John Edwards, respectively. The contract DROF_NOM pays one dollar if another candidate wins and nothing otherwise. There is similar data for the Republican nomination, but one of the contracts is split in two during the sample. For simplicity, we consider only the Democratic nomination data. Figure 1 shows daily prices for the first 235 days of the market: March 2 through October 28, 2007.

We suspect that the data are somewhat noisy. The total price of all contracts differs from one by more than 0.01 on 124 days out of 235; by more than 0.02 on 65 days; by more than 0.03 on 42 days; and by more than 0.05 on 20 days. First order sample autocorrelations are consistently less than and significantly different from zero, but decay quickly. These two observations suggest the presence of bid-ask bounce and the microstructure of the market is such that this is quite possible. Traders can post bid or ask prices, buy at the currently lowest ask price, or sell at the currently highest bid price. We do not have historical data on bid-ask spreads, but we observe that the spread is usually, but not always, less than 0.01.

To mitigate microstructure noise, we construct weekly data. We take the closing price on Saturday, since Saturday is the day of the week with highest trading volume, and thus the day we might expect a narrower bid-ask spread. We construct implied probabilities by dividing each contract's closing price by the total price of all contracts. Where the total closing price is less than 0.97 or greater than 1.03, which occurs on four of 35 Saturdays, we treat the data for that day as missing.

We consider a simple model for implied probabilities. We take $Z_t = I$, so that there is one state for each candidate. We also ignore covariates and omit the term $X'\beta$. For this special case we obtain, using the derivation in Appendix A, the following expressions for h_t and c_t :

$$h_t = -\psi_1(G_t)\gamma_t\gamma_t' - \text{diag} \{ [\psi(G_t) - \psi(\gamma_{ti}) + \log \pi_{ti}] \gamma_{ti} - \psi_1(\gamma_{ti})\gamma_{ti}^2 \},$$

$$c_{ti} = [\psi(G_t) - \psi(\gamma_{ti}) + \log \pi_{ti}] \gamma_{ti}(1 - \alpha_{ti}) - \psi_1(G_t)(\gamma_t^\top \alpha_t)\gamma_{ti} + \psi_1(\gamma_{ti})\gamma_{ti}^2\alpha_{ti}.$$

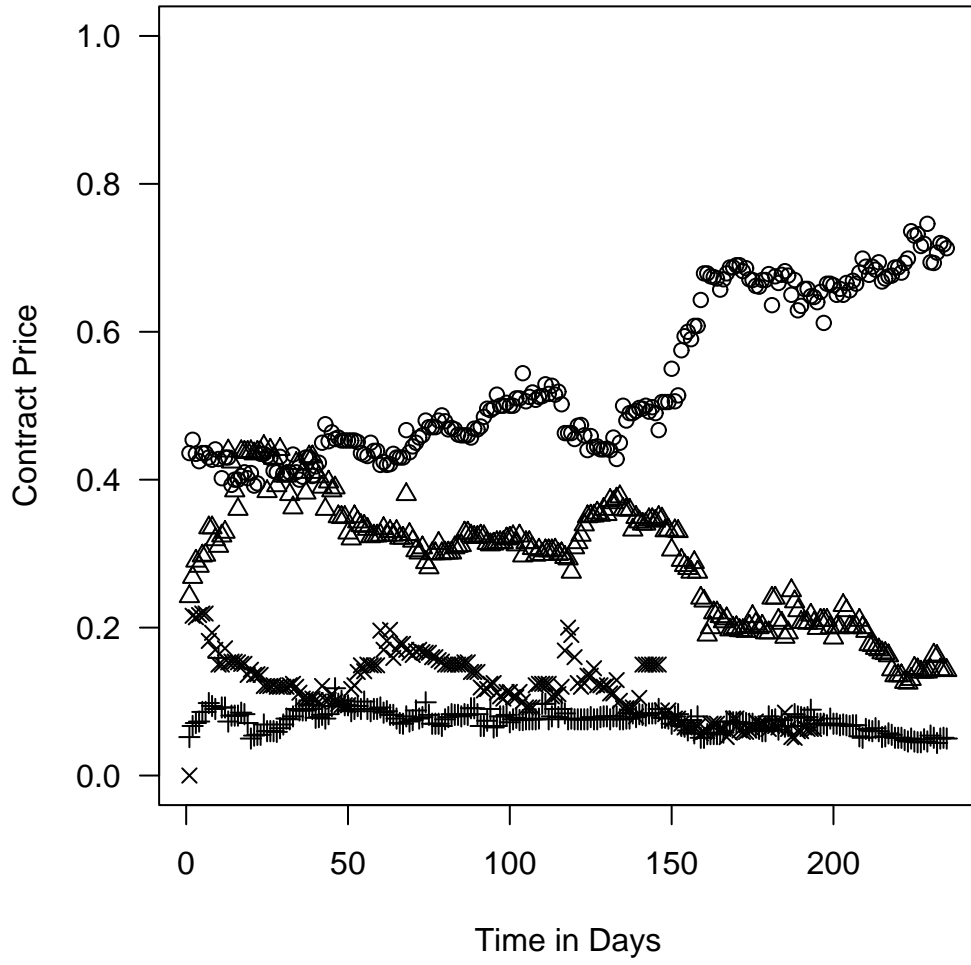


Figure 1: Democratic Candidate Prices

For weeks t where the data is considered missing, $h_t = 0$ and $c_t = 0$.

We set prior hyperparameters to the following values to specify a prior distribution for μ_1 , δ , Φ , H_1 and H_α .

$$\begin{aligned} \bar{r}_1 &= 7.0 \cdot \iota, & \bar{P}_1 &= 4.0 \cdot I, \\ \bar{r} &= \text{vec} \left(\begin{bmatrix} 0.0 \cdot \iota^\top \\ I \end{bmatrix} \right), & \bar{P} &= \text{diag} \left[\text{vec} \left(\begin{bmatrix} 0.0025 \cdot \iota^\top \\ 0.0025 \cdot \iota \iota^\top \end{bmatrix} \right) \right], \\ \bar{\nu}_1 &= 100.0, & \bar{A}_1 &= 0.1 \cdot I, \\ \bar{\nu}_\alpha &= 10.0, & \bar{A}_\alpha &= 0.002 \cdot I, \end{aligned}$$

where ι is a row vector of four ones and I is the 4×4 identity matrix. The elements of μ and mutually independent and normal with mean 7.0 and standard deviation 2.0. The elements of δ and Φ are mutually independent, with $\delta_i \sim N(0, (0.05)^2)$, $\Phi_{ii} \sim N(1, (0.05)^2)$ and $\Phi_{ij} \sim N(0, (0.05)^2)$ for $j \neq i$. H_1 has mean $5000.0 \cdot I$ with degrees of freedom 10.0, and H_α has mean $1000.0 \cdot I$ with degrees of freedom 100.0.

Using the basic proposal to update α , we obtain the following posterior results through simulation. Table 1 shows the posterior median and interquartile range for elements of δ . Table 2 does the same for the elements of Φ . We see that the mass of the joint posterior distribution of δ and Φ is close to the values $\delta = 0$ and $\Phi = I_p$ which correspond to independent random walks without drift.

Since the conditional precision H_α of the α_t is more difficult to interpret, we report features of the conditional variance $\Sigma_\alpha \equiv H_\alpha^{-1}$. Figure 3 shows the posterior median and interquartile range for the diagonal elements of Σ_α . We see that the Clinton, Edwards and Other contracts have similar volatility, but that the Obama contract has a somewhat higher range of plausible volatility. Figure 4 shows the posterior median and interquartile range for the conditional correlations implied by the conditional variance Σ_α . There little evidence for any correlation of a particular sign.

The estimated acceptance probability for α^* proposals is 0.235. We draw α^* five times for each new draw of θ and estimate the probability of at least one acceptance in five as 0.646. The comparison of this value to $1 - (1 - 0.235)^5 = 0.738$ suggests a moderately decreasing hazard rate.

Table 1: Posterior Median and Interquartile Range for Elements of δ

Clinton	Obama	Edwards	Other
0.0116	0.0035	-0.0014	-0.0094
(-0.0211, 0.0438)	(-0.0297, 0.0357)	(-0.0340, 0.0316)	(-0.0448, 0.0242)

Table 2: Posterior Median and Interquartile Range for Elements of Φ

Clinton	Obama	Edwards	Other
1.016	-0.017	-0.000	-0.003
(1.007, 1.024)	(-0.028, -0.005)	(-0.013, 0.013)	(-0.014, 0.008)
0.015	0.989	-0.003	-0.013
(0.005, 0.025)	(0.978, 1.000)	(-0.016, 0.010)	(-0.024, -0.001)
-0.006	0.004	0.995	0.005
(-0.016, 0.003)	(-0.008, 0.015)	(0.983, 1.008)	(-0.007, 0.016)
-0.028	0.023	0.005	1.008
(-0.038, -0.018)	(0.011, 0.034)	(-0.008, 0.018)	(0.997, 1.019)

Table 3: Posterior Median and Interquartile Range for Diagonal Elements of Σ_α

Clinton	Obama	Edwards	Other
0.0153	0.0175	0.0153	0.0159
(0.0132, 0.0184)	(0.0146, 0.0223)	(0.0133, 0.0180)	(0.0135, 0.0193)

Table 4: Posterior Median and Interquartile Range for Elements of C_α

Clinton	Obama	Edwards	Other
1.000	-0.015	0.003	-0.043
(1.000, 1.000)	(-0.174, 0.148)	(-0.149, 0.159)	(-0.197, 0.113)
-0.015	1.000	-0.016	-0.037
(-0.174, 0.148)	(1.000, 1.000)	(-0.180, 0.140)	(-0.210, 0.128)
0.003	-0.016	1.000	0.025
(-0.149, 0.159)	(-0.180, 0.140)	(1.000, 1.000)	(-0.130, 0.186)
-0.043	-0.037	0.025	1.000
(-0.197, 0.113)	(-0.210, 0.128)	(-0.130, 0.186)	(1.000, 1.000)

6 Conclusions

This work builds on the paper by McCausland, Miller, and Pelletier (2007), which describes new methods for drawing states in Gaussian state space models. Here we draw states in a semi-Gaussian state space model where the state is Gaussian and observations are Dirichlet. The general approach here can be applied to other semi-Gaussian state space models.

Several future extensions are possible. We can generalize the state dynamics as follows:

$$\alpha_{t+1} = W_t \beta + T_t \alpha_t + \epsilon_t \quad \epsilon_t \sim N(0, \Omega_t^{-1}).$$

The expressions for the Hessian \bar{H} and covector \bar{c} of the Gaussian conditional distribution of α given θ change, but the Hessian retains the block band structure illustrated in equation (3).

We are currently working on refinements to the proposal distribution so that it more closely approximates the conditional distribution of states given parameters and observed data. This will facilitate inference in applications with larger data sets. Initial work with a univariate stochastic volatility model suggests that the methods of McCausland, Miller, and Pelletier (2007) can be modified to generate proposals that are much closer to the target distribution.

A Derivations of h_t and c_t

In this appendix we derive expressions for h_t and c_t used in Section 3 to draw states. We use the following notational conventions. The vector ι has all elements equal to 1. Vector e_i has elements equal to zero except the i 'th, which is 1. For vector arguments, the functions ψ , ψ_1 , \log , \exp , \min and \max are applied element-by-element and return vectors of the same length. The Hadamard (element-by-element) product of matrices A and B is denoted $A \bullet B$ and the Kronecker product is denoted $A \otimes B$. The digamma function ψ and the trigamma function ψ_1 are defined by

$$\psi(x) \equiv \frac{\partial \log \Gamma(x)}{\partial x} \quad \text{and} \quad \psi_1(x) \equiv \frac{\partial^2 \log \Gamma(x)}{\partial x^2},$$

where Γ is the gamma function. For matrix and vector calculus, we use the notational conventions in Dhrymes (2000).

We begin with the most general results. Suppose $f(y_t|\gamma_t)$ is a density or probability mass function for observed data given a vector of parameters γ_t , and that γ_t is a deterministic twice differentiable function of α_t . Then the gradient and Hessian of $\log f$ with respect to α_t are given by

$$\frac{\partial \log f}{\partial \alpha_t} = \frac{\partial \log f}{\partial \gamma_t} \frac{\partial \gamma_t}{\partial \alpha_t}$$

and

$$\begin{aligned} \frac{\partial^2 \log f}{\partial \alpha_t \partial \alpha_t^\top} &= \frac{\partial}{\partial \alpha_t} \left[\left(\frac{\partial \gamma_t}{\partial \alpha_t} \right)^\top \left(\frac{\partial \log f}{\partial \gamma_t} \right)^\top \right] \\ &= \left(\frac{\partial \log f}{\partial \gamma_t} \otimes I_m \right) \frac{\partial^2 \gamma_t}{\partial \alpha_t \partial \alpha_t^\top} + \left(\frac{\partial \gamma_t}{\partial \alpha_t} \right)^\top \frac{\partial^2 \log f}{\partial \gamma_t \partial \gamma_t^\top} \left(\frac{\partial \gamma_t}{\partial \alpha_t} \right). \end{aligned}$$

We now specialize to the case that $\gamma_t = \exp(X_t\beta + Z_t\alpha_t)$. We define $\eta_t \equiv X_t\beta + Z_t\alpha_t$. We compute the following derivatives of γ_t :

$$\frac{\partial \gamma_t}{\partial \alpha_t} = \text{diag}(\gamma_t) Z_t.$$

$$\begin{aligned}
\frac{\partial^2 \gamma_t}{\partial \alpha_t \partial \alpha_t^\top} &= \frac{\partial}{\partial \alpha_t} [Z_t^\top \text{diag}(\gamma_t)] \\
&= (I_m \otimes Z_t^\top) \frac{\partial}{\partial \alpha_t} \text{diag}(\gamma_t) \\
&= (I_m \otimes Z_t^\top) \left[\frac{\partial}{\partial \eta_t} \text{diag}(\gamma_t) \right] Z_t \\
&= (I_m \otimes Z_t^\top) \begin{bmatrix} \gamma_{t1} e_1 e_1^\top \\ \vdots \\ \gamma_{tp} e_p e_p^\top \end{bmatrix} Z_t.
\end{aligned}$$

In the case where f is the Dirichlet density in (1), we have

$$\frac{\partial \log f}{\partial \gamma_t} = [\psi(G_t)\iota - \psi(\gamma_t) + \log \pi_t]^\top \quad (7)$$

$$\frac{\partial^2 \log f}{\partial \gamma_t \partial \gamma_t^\top} = \psi_1(G_t)\iota \iota^\top - \text{diag}(\psi_1(\gamma_t)) \quad (8)$$

Putting it all together, we have

$$\frac{\partial \log f}{\partial \alpha_t} = [\psi(G_t)\iota - \psi(\gamma_t) + \log \pi_t]^\top \text{diag}(\gamma_t) Z_t$$

$$\begin{aligned}
\frac{\partial^2 \log f}{\partial \alpha_t \partial \alpha_t^\top} &= \left\{ [\psi(G_t)\iota - \psi(\gamma_t) + \log \pi_t]^\top \otimes I_m \right\} (I \otimes Z_t^\top) \frac{\partial \text{diag}(\gamma_t)}{\partial \eta_t} Z_t \\
&+ Z_t^\top \text{diag}(\gamma_t) [\psi_1(G_t)\iota \iota^\top - \text{diag}(\psi_1(\gamma_t))] \text{diag}(\gamma_t) Z_t \\
&= \left\{ [\psi(G_t)\iota - \psi(\gamma_t) + \log \pi_t]^\top \otimes Z_t^\top \right\} \frac{\partial \text{diag}(\gamma_t)}{\partial \eta_t} Z_t \\
&+ Z_t^\top [\psi_1(G_t)\gamma_t \gamma_t^\top - \text{diag}(\gamma_t \bullet \psi_1(\gamma_t) \bullet \gamma_t)] Z_t \\
&= Z_t^\top [\text{diag}((\psi(G_t)\iota - \psi(\gamma_t) + \log \pi_t) \bullet \gamma_t) + \psi_1(G_t)\gamma_t \gamma_t^\top] Z_t \\
&+ Z_t^\top [\psi_1(G_t)\gamma_t \gamma_t^\top - \text{diag}(\gamma_t \bullet \psi_1(\gamma_t) \bullet \gamma_t)] Z_t.
\end{aligned}$$

Finally, h_t and c_t are given by equations (5) and (6).

We can also compute values h_t^{safe} and c_t^{safe} where

1. h_t^{safe} is guaranteed to be positive semi-definite;
2. the gradient of the log density of the Gaussian approximation with precision h_t^{safe} and covector c_t^{safe} equals the gradient of $\log f$ with respect to α_t ; and
3. $h_t^{\text{safe}} - h_t$ is positive semi-definite.

We compute h_t^{safe} and c_t^{safe} as h_t and c_t , but we replace the Hessian of $\log f$ with respect to γ_t , given by (8), with the negative semi-definite matrix A defined by

$$A = \begin{cases} \lambda^* \psi_1(G_t) \iota^\top - \text{diag}(\psi_1(\gamma_t)) & \gamma_t > 0 \\ \min(0, \psi_1(G_t)) \iota^\top - \text{diag}(\max(0, \psi_1(\gamma_t))) & \text{otherwise,} \end{cases}$$

where

$$\lambda^* = \frac{\sum_{i=1}^p \psi_1(\gamma_{ti})}{\psi_1(G_t)}.$$

If $\gamma_t > 0$, λ^* solves the following equation in λ :

$$\begin{aligned} |A| &= |\lambda \psi_1(G_t) \iota^\top - \text{diag}(\psi_1(\gamma_t))| \\ &= |\text{diag}(\psi_1(\gamma_t))| \cdot |1 - \lambda \psi_1(G_t) \iota^\top (\text{diag}(\psi_1(\gamma_t)))^{-1} \iota| = 0. \end{aligned}$$

We use Dhrymes (2000, Corollary 2.6) to pass from the second to the third line.

Given this result, it is easy to see that A is negative semi-definite and that the matrix $A - \frac{\partial^2 \log f}{\partial \gamma_t \partial \gamma_t}$ is negative semi-definite. Therefore h_t^{safe} and $h_t^{\text{safe}} - h_t$ are positive semi-definite.

References

- Aitchison, J. (1986). *The statistical analysis of compositional data: Monographs on statistics and applied probability*. Chapman & Hall Ltd., London, UK.
- de Jong, P., and Shephard, N. (1995). ‘The Simulation Smoother for Time Series Models’, *Biometrika*, 82(1): 339–350.
- Dhrymes, P. J. (2000). *Mathematics for Econometrics, Third Edition*. Springer–Verlag, New York.

- Durbin, J., and Koopman, S. J. (1997). 'Monte Carlo maximum likelihood estimation for non-Gaussian state space models', *Biometrika*, 84(3): 669–684.
- (2002). 'A Simple and Efficient Simulation Smoother for State Space Time Series Analysis', *Biometrika*, 89(3): 603–615.
- Green, P. J. (1995). 'Reversible Jump Markov chain Monte Carlo computation and Bayesian model determination', *Biometrika*, 82: 711–732.
- Grunwald, G. K., Raftery, A. E., and Guttorp, P. (1993). 'Time Series of Continuous Proportions', *Journal of the Royal Statistical Society Series B*, 55: 103–116.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton, NJ.
- Hastings, W. K. (1970). 'Monte Carlo Sampling Methods Using Markov Chains and Their Applications', *Biometrika*, 57: 97–109.
- McCausland, W. J., Miller, S., and Pelletier, D. (2007). 'A New Approach to Drawing States in State Space Models', *Cahiers de recherche du Département de sciences économiques, Université de Montréal*, no. 2007-06.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). 'Equation of State Calculations by Fast Computing Machines', *J. Chem. Phys*, 21: 1087–1092.