WP 42_14

Joshua C.C. Chan
Australian National University, Australia

Rodney Strachan
University of Queensland, Australia
The Rimini Centre for Economic Analysis, Italy

THE ZERO LOWER BOUND: IMPLICATIONS FOR MODELLING THE INTEREST RATE

The Zero Lower Bound:
Implications for Modelling the Interest Rate

Joshua C.C. Chan[*]
Research School of Economics,
and Centre for Applied Macroeconomic Analysis,
Australian National University

Rodney Strachan[†]
School of Economics,
and Centre for Applied Macroeconomic Analysis,
University of Queensland

February 2014

Abstract

The time-varying parameter vector autoregressive (TVP-VAR) model has been used to successfully model interest rates and other variables. As many short interest rates are now near their zero lower bound (ZLB), a feature not included in the standard TVP-VAR specification, this model is no longer appropriate. However, there remain good reasons to include short interest rates in macro models, such as to study the effect of a credit shock. We propose a TVP-VAR that accounts for the ZLB and study algorithms for computing this model that are less computationally burdensome than others yet handle many states well. To illustrate the proposed approach, we investigate the effect of the zero lower bound of interest rate on transmission of a monetary shock.

Keywords: integrated likelihood; accept-reject Metropolis-Hastings; cross-entropy; liquidity trap; zero lower bound

---

[*]Corresponding author. Email: joshua.chan@anu.edu.au
[†]Email: rodney.strachan@anu.edu.au. Strachan is also a Fellow of the Rimini Centre for Economic Analysis.

1 Introduction

The time-varying parameter vector autoregressive (TVP-VAR) model has become prominent in empirical macroeconomics with important applications such as Cogley and Sargent (2001 and 2005) and Primiceri (2005). Interest rates are often included in such models to capture monetary policy changes. As interest rates have recently fallen to be very near the zero lower bound (ZLB), the usual linear Gaussian specification of the TVP-VAR is no longer appropriate since these models do not allow for bounds on the support of variables. As target interest rates have reached the bound of their supports, alternative instruments have been employed to implement monetary policy and such instruments have therefore been included in macroeconometric models (see, for example, Ugai (2007)). However, as interest movements reflect more than just monetary policy changes, there remain good reasons to include short term interest rates in multivariate macroeconometric models and therefore the ZLB must be incorporated into the model specification.

One approach to the problem of the ZLB is to treat interest rates as a censored variable, where the value of the interest rate is not observed below some chosen positive level. This approach was adopted in Iwata and Wu (2006) and Nakajima (2011), where the point of censoring was chosen to be 50 basis points. These studies maintain a focus upon monetary policy transmission and so this censoring is appropriate as it is reasonable to assume there are no further monetary policy changes via interest rates when they are at such low levels. However, interest rates are not censored variables as a censored variable is observed in some ranges of its support but not observed in others. Interest rates are more accurately described, and modelled, as variables with a bounded or truncated support. A variable with a truncated support may always be observed. Interest rates are always observed but always positive[1], and censoring below a positive level excludes inference on behaviour in an important range of the variable as it excludes those observations from the likelihood.

While the focus of many studies has been on the transmission of monetary policy shocks, such shocks are not transmitted via interest rates when the growth is low, inflation is low and interest rates are already pushing up against the ZLB. It is in this state of the economy when interest rates are low that other shocks, such as credit shocks say, would be more evident and have more apparent and immediate effect. Consider, for example, that changes in the U.S. nominal federal funds rate or target rate are reflected in the effective funds rate and other short rates such as the three month treasury bill rate. The target rate was set in 2008 as a range from 0 to 25 basis points and has not changed since this setting until the time of writing, reflecting no further changes in monetary policy via this instrument. Yet, as we see in Figure 1, the three month treasury bill rate has continued to vary over that period due to other shocks manifest that are transmitted via short interest rates. One potential shock we consider is a credit shock. An important feature of interest rates, evident from Figure 1, is that when they are low any shock of say 10 or 25 basis points would be large relative to the conditional variance.

Treating the ZLB on interest rates as a bound on the support rather than a point of censure,

---

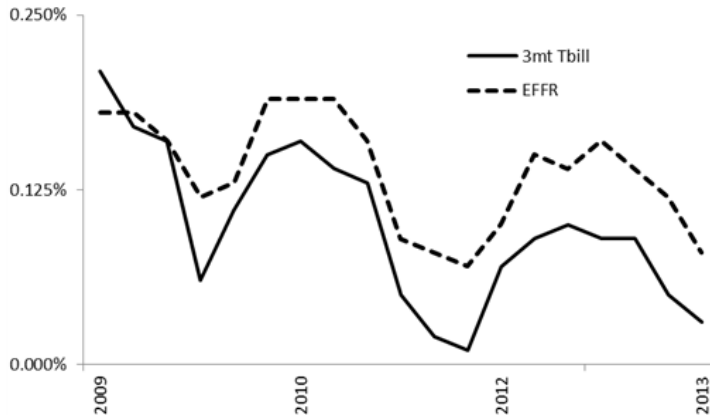[1] (at the frequencies we consider).

Figure 1: Quarterly plot of the Effective Federal Funds Rate (EFFR) and the 3-Month Treasury Bill (3mth TBill) from 1995 Q1 to 2013 Q3.

therefore, has important modelling implications. In this paper we present a specification for a TVP-VAR with stochastic volatility that accounts for the ZLB as a bound on the support of the interest rates. Borrowing the acronym from Nakajima (2011), we denote this model by TVP-VAR-ZLB. The model could be used to study other variables with bounded supports such as exchange rates.

Drawing upon recent developments in precision-based algorithms for the linear Gaussian TVP-VAR, we present three fast sampling schemes for efficient simulation of the TVP-VAR-ZLB. The first algorithm, the baseline algorithm, approximates the conditional distribution of the states by a multivariate Gaussian or t density, which is then used as a proposal density for posterior simulation using Markov chain Monte Carlo (MCMC) methods. This approximating density can also used for evaluating the integrated likelihood — the joint distribution of the observations given the model parameters but integrated over the states — via importance sampling. We then build upon this baseline approach to consider two other more efficient algorithms for posterior simulation. The first of these, our second algorithm, is the accept-reject Metropolis-Hastings (ARMH) algorithm that combines the classic accept-reject sampling and the Metropolis-Hastings algorithm. The third algorithm is a collapsed sampler used in conjunction with the cross-entropy method, where we sample the states and the model parameters jointly to reduce autocorrelations in the posterior simulator.

The specific framework we consider is a general state space model where the evolution of the $n \times 1$ vector of observations y_t is governed by the measurement or observation equation characterised by a generic density function $p(y_t \mid \eta_t, \theta)$, where η_t is an $m \times 1$ vector of latent states and θ denotes the set of model parameters. Note that the density $p(y_t \mid \eta_t, \theta)$ may depend on previous observations y_{t-1}, y_{t-2}, etc. and other covariates as is the case in the application in this paper. These observations are suppressed in the conditioning sets for notational convenience. The evolution of the states η_t, in turn, is specified by the state or transition equation summarised by the density function $p(\eta_t \mid \eta_{t-1}, \theta)$. We note in passing that the proposed approach can be

3

easily generalized to the case where the state equation is non-Markovian and the observation y_t depends on previous states η_{t-1}, η_{t-2}, etc.

The contributions of this paper are three-fold. In the first contribution we build upon the recently proposed precision-based sampler in Chan and Jeliazkov (2009b) and McCausland, Millera, and Pelletier (2011) originally developed for linear Gaussian state space models. Combining a Gaussian approximation to the measurement equation with the precision sampler, we present a quick method to obtain a Gaussian or a student t approximation for the conditional density of the states $p(\eta \mid \theta, y)$. We develop an accept-reject Metropolis Hastings (ARMH) algorithm for efficient simulation of the states. The efficiency in the MH step with a Gaussian proposal can be quite low in certain settings, presumably because the Gaussian approximation is not sufficiently accurate. By using the ARMH algorithm, we construct a better approximation and consequently the acceptance rate is substantially higher compared to the baseline MH algorithm. The cost of this increased acceptance rate is that multiple draws from the proposal density might be required and this is why it is essential to have low marginal cost for additional draws.

The second contribution of the paper is to develop a practical way to sample the model parameters θ and the states η jointly. As η and θ are often highly correlated, this motivates sampling θ and η jointly by first drawing from $p(\theta \mid y)$ marginally of the states η followed by a draw from $p(\eta \mid y, \theta)$. To locate a good proposal density for θ, denoted as $q(\theta \mid y)$, we adopt the cross-entropy method (Rubinstein and Kroese, 2004) to obtain the optimal $q(\theta \mid y)$ in a well-defined sense. We show via an empirical example that the efficiency of the sampling scheme is substantially improved.

In the third contribution of the paper we demonstrate the overall approach with a topical application. We investigate the implications for transmission of monetary shocks of accounting for the zero lower bound (ZLB) on interest rates. Recent work using time-varying parameter vector autoregressive models (TVP-VARs) by Cogley and Sargent (2001 and 2005), Primiceri (2005), Sims and Zha (2006) and Koop, Leon-Gonzalez and Strachan (2009) demonstrated the importance of model specification for the evidence on changes in the transmission mechanism for monetary shocks. These studies considered periods of relatively high interest rates, with the exception of the period after the dot com bubble when interest rates fell as low as 1%. This latter event, and the history of Japan in the 1990s, led to an increase in interest in the effect of the ZLB on the conduct of monetary policy (see, for example, Iwata and Wu (2006), Reifschneider and Williams (2000) and Svensson (2003)). We study the effect of the ZLB on the transmission of a contractionary monetary shock in a low growth, low interest rate environment. We use a TVP-VAR with a truncated support for an interest rate variable and multivariate stochastic volatility. Our interest is in how estimates of the impulse responses to positive monetary shocks change when we allow for the ZLB.

The rest of this article is orgainsed as follows. Section 2 first introduces the TVP-VAR-ZLB model. In Section 3 we consider a general form of the state space model for which we propose an approximation to the conditional density for the states and the three efficient simulation schemes for the states using the precision sampler. In Section 4 we apply the sampler to estimate a standard model used a number of times in the literature to investigate the evolution

of the transmission of monetary shocks, but we incorporate the ZLB on interest rates which results in a non-linear measurement equation. Section 5 concludes and discusses various future research directions.

2   The TVP-VAR-ZLB Model

To illustrate the proposed approach we estimate a VAR with a lower bound restriction on one of the variables; this restriction implies a measurement equation that is non-linear in the states. Specifically, we investigate the implications for the transmission of monetary shocks of accounting for the zero lower bound (ZLB) on interest rates. With time varying parameters, incorporating the lower bound on interest rates introduces a non-linearity in the states into the measurement equation. Recent work using time-varying parameter vector autoregressive models (TVP-VARs) on changes in the transmission mechanism for monetary policy shocks (see for example, Cogley and Sargent, 2001, 2005, Primiceri, 2005, Sims and Zha, 2006, and Koop, Leon-Gonzalez and Strachan, 2009) has ignored the lower bound on interest rates. Not accounting for the ZLB is reasonable when interest rates are relatively high and far from zero. However, episodes of low interest rates have occurred often in recent history including, as examples, in the U.S. just after the dot com bubble of 2001, during the 1990s in Japan, or since 2009 in much of the developed world. The prevalence of low interest rates suggests it is important to know whether transmission of monetary shocks is affected and, if so, to understand how the transmission mechanism is affected. Our focus is upon the effect of a contractionary monetary shock when interest rates are on the ZLB. Such a situation might arise for the U.S. if several rating agencies were to downgrade the rating of U.S. government debt and creditors then began to demand a premium to compensate for the risk of default, or if the cost of funds to banks increased independently of moves in the Federal Funds rate inducing an effective, unintended tightening of monetary policy.[2]

2.1   The Model

The framework we consider is the following time-varying parameter vector autoregressive (TVP-VAR) model with l lags:

$$y_t = \mu_t + A_{1t}y_{t-1} + \cdots + A_{lt}y_{t-l} + \epsilon_t, \quad \epsilon_t \sim \mathsf{N}(0, \Sigma_t^{-1}),$$

where μ_t is an $n \times 1$ vector of time-varying intercepts, $A_{t1}, \ldots, A_{tl}$ are $n \times n$ matrices of VAR lag coefficients at time t, and Σ_t is a time-varying precision matrix, i.e., the inverse of the covariance matrix. For the purpose of estimation, we write the VAR system in the form of seemingly unrelated regressions:

$$y_t = x_t\beta_t + \epsilon_t, \quad \epsilon_t \sim \mathsf{N}(0, \Sigma_t^{-1}), \tag{1}$$

---

[2]This effect was observed in February 2012 in Australia when, after the central bank kept its rate unchanged, all banks increased their lending rates in response to increased costs of wholesale funding costs.

It is for this reason we do not term the shock a monetary policy shock, but an unintended shock to monetary conditions.

where $x_t = I_n \otimes [1, y'_{t-1}, \ldots, y'_{t-l}]$ and $\beta_t = \text{vec}([\mu_t : A_{1t} : \cdots : A_{lt}]')$ is a $k \times 1$ vector of VAR coefficients with $k = n^2 l + n$. To model the time-varying precision matrix Σ_t, we follow the approach proposed in Primiceri (2005) by first factoring the precision matrix as $\Sigma_t = L'_t D_t^{-1} L_t$, where $D_t = \text{diag}(e^{h_{1t}}, \ldots, e^{h_{nt}})$ is a diagonal matrix, and L_t is a lower triangular matrix with ones on the main diagonal, i.e.,

$$D_t = \begin{pmatrix} e^{h_{1t}} & 0 & \cdots & 0 \\ 0 & e^{h_{2t}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{h_{nt}} \end{pmatrix}, \quad L_t = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{21,t} & 1 & 0 & \cdots & 0 \\ a_{31,t} & a_{32,t} & 1 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1,t} & a_{n2,t} & \cdots & a_{n(n-1),t} & 1 \end{pmatrix}.$$

This decomposition has been employed in various applications, especially in the context of efficient estimation of covariance matrices (Pourahmadi, 1999, 2000, Smith and Kohn, 2002, Chan and Jeliazkov, 2009a, among others). In the setting of VAR models with time-varying volatility, this approach is first considered in Cogley and Sargent (2005). For notational convenience, we let $h_{\bullet t} = (h_{1t}, \ldots, h_{nt})'$ and $h_{i\bullet} = (h_{i1}, \ldots, h_{iT})'$. That is, $h_{\bullet t}$ is the $n \times 1$ vector obtained by stacking h_{it} by the first subscript, whereas $h_{i\bullet}$ is the $T \times 1$ vector obtained by stacking h_{it} by the second subscript. The log-volatilities $h_{\bullet t}$ evolve according to the state equation

$$h_{\bullet t} = h_{\bullet t-1} + \xi_t, \quad \xi_t \sim \mathsf{N}(0, \Omega_h^{-1}) \tag{2}$$

for $t = 2, \ldots, T$, where $\Omega_h = \text{diag}(\omega_{h1}, \ldots, \omega_{hn})$ is a diagonal matrix. The process is initialised with $h_{\bullet 1} \sim \mathsf{N}(0, V_h^{-1})$ for some known diagonal precision matrix V_h. Let a_t denote the free elements in L_t ordered by rows, i.e., $a_t \equiv a_{\bullet t} = (a_{21,t}, a_{31,t}, a_{32,t}, \ldots, a_{n(n-1),t})'$, so that a_t is an $m \times 1$ vector of parameters where $m = n(n-1)/2$. The evolution of a_t is modelled as a random walk

$$a_t = a_{t-1} + \zeta_t, \quad \zeta_t \sim \mathsf{N}(0, \Omega_a^{-1}) \tag{3}$$

for $t = 2, \ldots, T$, where $\Omega_a = \text{diag}(\omega_{a1}, \ldots, \omega_{am})$ is a diagonal precision matrix. The process is initialised with $a_1 \sim \mathsf{N}(0, V_a^{-1})$ for some known diagonal precision matrix V_a. In what follows we use these two parameterizations, namely, Σ_t and $(h_{\bullet t}, a_t)$, interchangeably. To complete the specification of the model, it remains to specify the evolution of the VAR coefficients β_t. We follow the standard approach of modelling the VAR coefficients β_t as a random walk process:

$$\beta_t = \beta_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathsf{N}(0, \Omega_\beta^{-1}) \tag{4}$$

for $t = 2, \ldots, T$, where $\Omega_\beta = \text{diag}(\omega_{\beta 1}, \ldots, \omega_{\beta k})$ is a diagonal precision matrix. The process is initialised with $\beta_1 \sim \mathsf{N}(0, V_\beta^{-1})$ for some known precision matrix V_β.

After presenting the basic setup of a TVP-VAR model with stochastic volatility, we now wish to impose the restriction that the nominal interest rate is always non-negative. For this purpose, arrange the data y_t so that y_{1t}, the first element of y_t, is the nominal interest rate, and let x_{1t} be the first row of x_t. We assume that $y_{1t} \geqslant 0$. Consequently, given β_t and Σ_t, y_t follows a multivariate Gaussian distribution with the first element restricted to be positive. To derive the likelihood function, first note that since only y_{1t} is constrained while other elements of y_t

are not, the marginal distribution of y_{1t} is a univariate Gaussian variable truncated below at 0. In fact, it can be shown that

$$(y_{1t} \mid \beta_t, \Sigma_t) \sim \mathsf{N}(x_{1t}\beta_t, \mathrm{e}^{h_{1t}})\mathbb{1}(y_{1t} \geqslant 0).$$

It follows that given β_t and Σ_t, we have

$$\mathbb{P}(y_{1t} \geqslant 0 \mid \beta_t, \Sigma_t) = 1 - \Phi\left(-x_{1t}\beta_t/\mathrm{e}^{\frac{1}{2}h_{1t}}\right) = \Phi\left(x_{1t}\beta_t\mathrm{e}^{-\frac{1}{2}h_{1t}}\right),$$

where $\Phi(\cdot)$ denotes the standard Gaussian cumulative distribution function. Letting $y = (y_1', \ldots, y_T')'$, $\beta = (\beta_1', \ldots, \beta_T')'$ and $\Sigma = (\Sigma_1, \ldots, \Sigma_T)$, the log-likelihood function is thus

$$\log p(y \mid \beta, \Sigma) = \sum_{t=1}^{T} \log p(y_t \mid \beta_t, \Sigma_t), \tag{5}$$

where

$$p(y_t \mid \beta_t, \Sigma_t) \propto -\frac{1}{2}\log|\Sigma_t^{-1}| - \frac{1}{2}(y_t - x_t\beta_t)'\Sigma_t(y_t - x_t\beta_t) - \log\Phi\left(x_{1t}\beta_t\mathrm{e}^{-\frac{1}{2}h_{1t}}\right).$$

Given the measurement equation (5) and the state equations (2)–(4), we present three efficient Markov samplers to sample from the posterior distribution. To this end, we first specify the priors for the remaining parameters: $\omega_\beta = (\omega_{\beta 1}, \ldots, \omega_{\beta k})'$, $\omega_h = (\omega_{h1}, \ldots, \omega_{hn})'$ and $\omega_a = (\omega_{a1}, \ldots, \omega_{am})'$. Specifically, the elements of ω_β, ω_h and ω_a follow independently Gamma distributions: $\omega_{\beta i} \sim \mathsf{Gamma}(r_{\beta i}, s_{\beta i})$ for $i = 1, \ldots, k$, $\omega_{hi} \sim \mathsf{Gamma}(r_{hi}, s_{hi})$, for $i = 1, \ldots, n$, and $\omega_{ai} \sim \mathsf{Gamma}(r_{ai}, s_{ai})$, for $i = 1, \ldots, m$. For later reference, we stack $h = (h_{\cdot 1}', \ldots, h_{\cdot T}')'$ and $a = (a_1', \ldots, a_T')'$. In the following section we present the sampling algorithms. As the blocks of states β, h and a are drawn sequentially, we use η to denote the current state being drawn and let θ denote the set of parameters except the current latent state. We will refer to the relevant measurement and state equations then as $p(y|\eta, \theta)$ and $p(\eta|\theta)$ respectively.

3 General State Space Model

The algorithms we propose for estimating the TVP-VAR-ZLB are quite general in that they are applicable to general non-linear and non-Gaussian models. More specifically, we consider a general state space model where the measurement equation is characterised by a generic density function $p(y_t \mid \eta_t, \theta)$, whereas the state equation, $p(\eta|\theta)$, is linear Gaussian as in previous studies. We give the explicit details on the measurement and state equations for the TVP-VAR-ZLB in Section 2.

A few comments are in order on the scope of the approaches we propose. In the application in Section 2 the non-linearity and non-Gaussianity are induced by the truncated support of interest rates, for which the conditional distribution is truncated Gaussian. However, our approaches may be applicable in other non-linear and or non-Gaussian settings. In our application, as interest rates move away from the ZLB, the model is very well approximated by the

standard TVP-VAR. As interest rates are far from zero for much of the sample, this gives us a benchmark for comparison between the two specifications over these periods. Using the algorithms presented below, one can estimate the integrated likelihood, $p(y \,|\, \theta)$ — an ingredient for maximum likelihood estimation and efficient MCMC design — via importance sampling using multiple draws from the proposal density.[3] We demonstrate this only in the final algorithm where it is integral to the procedure. Finally, we should point out that these approaches will probably work better in macroeconomic applications with sample sizes of a few hundred observations, rather than in, say, finance with sample sizes of tens of thousands. This conjecture is based upon the likely fall in acceptance rates in large samples.

3.1 Gaussian Approximation

We first discuss a quick method to obtain a Gaussian approximation for the conditional density $p(\eta \,|\, y, \theta)$. This approach builds upon the precision-based algorithm detailed in Chan and Jeliazkov (2009b) and McCausland, Millera, and Pelletier (2011). To begin, let f_t and G_t denote respectively the gradient and negative Hessian of $\log p(y_t \,|\, \eta_t, \theta)$ evaluated at $\eta_t = \widetilde{\eta}_t$, i.e.,

$$f_t \equiv \frac{\partial}{\partial \eta_t} \log p(y_t \,|\, \eta_t, \theta)\bigg|_{\eta_t = \widetilde{\eta}_t}, \quad G_t \equiv -\frac{\partial^2}{\partial \eta_t \eta_t'} \log p(y_t \,|\, \eta_t, \theta)\bigg|_{\eta_t = \widetilde{\eta}_t}.$$

Stacking these terms and define the following vector and matrix:

$$f = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_T \end{bmatrix}, \quad G = \begin{bmatrix} G_1 & 0 & \cdots & 0 \\ 0 & G_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & G_T \end{bmatrix}.$$

We then expand the log-likelihood $\log p(y \,|\, \eta, \theta) = \sum_{t=1}^{T} \log p(y_t \,|\, \eta_t, \theta)$ around $\widetilde{\eta} = (\widetilde{\eta}_1', \ldots, \widetilde{\eta}_T')'$ to obtain the expression

$$\begin{aligned} \log p(y \,|\, \eta, \theta) &\approx \log p(y \,|\, \widetilde{\eta}, \theta) + (\eta - \widetilde{\eta})' f - \frac{1}{2}(\eta - \widetilde{\eta})' G (\eta - \widetilde{\eta}) \\ &= -\frac{1}{2}\left[\eta' G \eta - 2\eta'(f + G\widetilde{\eta}) \right] + c_1, \end{aligned} \tag{6}$$

where c_1 is some unimportant constant independent of η.

The evolution of the states is governed by the following transition equation

$$\eta_t = \Gamma_t \eta_{t-1} + u_t \tag{7}$$

for $t = 1, \ldots, T$, where $u_t \sim \mathsf{N}(0, \Omega_t^{-1})$. The directed conditional structure for $p\left(\eta_t | \theta, \eta_{t-1}\right)$

---

[3]A further advantage of the proposed method is that it can be applied to non-Markovian state equations, which arise in, e.g., various DSGE models, and they are more difficult to handle under other approaches.

in (7) implies that the joint density for η is also Gaussian. To see this, define

$$K = \begin{pmatrix} I_m & & & & \\ -\Gamma_2 & I_m & & & \\ & -\Gamma_3 & I_m & & \\ & & \ddots & \ddots & \\ & & & -\Gamma_T & I_m \end{pmatrix} \quad \text{and} \quad \Omega = \begin{pmatrix} \Omega_1 & & & & \\ & \Omega_2 & & & \\ & & \Omega_3 & & \\ & & & \ddots & \\ & & & & \Omega_T \end{pmatrix},$$

so that (7) can be written as $K\eta = \gamma + \zeta$, where

$$\gamma = \begin{bmatrix} \Gamma_1 \eta_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{and} \quad \zeta = \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \vdots \\ \zeta_T \end{bmatrix} \sim \mathsf{N}(0, \Omega^{-1}).$$

Noting that $|K| = 1$, by a change of variable from ζ to η, we have

$$\log p(\eta \mid \theta) \propto -\frac{1}{2} \log |\Omega^{-1}| - \frac{1}{2} (\eta - \eta^0)' K'\Omega K(\eta - \eta^0)), \tag{8}$$

where $\eta^0 = K^{-1}\gamma$ is the prior mean. Note that the $Tm \times Tm$ precision matrix $K'\Omega K$ is banded, i.e., it only contains a small number of non-zero elements on a narrow band around the main diagonal. Combining (6) and the prior density in (8), we have

$$\begin{aligned} \log p(\eta \mid y, \theta) &\propto \log p(y \mid \eta, \theta) + \log p(\eta \mid \theta) \\ &\approx -\frac{1}{2} \left[\eta'(G + K'\Omega K)\eta - 2\eta'(f + G\widetilde{\eta} + K'\Omega K\eta^0) \right] + c_2. \end{aligned} \tag{9}$$

where c_2 is some unimportant constant independent of η. In other words, the approximating distribution is Gaussian with precision $H \equiv G + K'\Omega K$ and mean vector $H^{-1}\left(f + G\widetilde{\eta} + K'\Omega K\eta^0\right)$. Since G and $X'\Sigma X$ are banded, it follows that H is also banded. An important consequence is that its Cholesky decomposition can be obtained in $O(N)$ operations instead of $O(N^3)$ operations for full matrices, where N is the dimension of the matrix. By exploiting this fact, one can sample $(\eta \mid y, \theta)$ without the need to carry out an inversion to obtain H^{-1} and $\widehat{\eta}$. More specifically, the mean $\widehat{\eta}$ can be found in two steps. First, we compute the (banded) Cholesky decomposition C_H of H such that $C_H'C_H = H$. Second, we solve

$$C_H'C_H\widehat{\eta} = f + G\widetilde{\eta} + K'\Omega K\eta^0.$$

Further details on this sampler can be found in Chan and Jeliazkov (2009b). The important point is that additional draws from this Gaussian approximation can be obtained with low marginal costs.

It remains to choose the point $\widetilde{\eta}$ around which to construct the Taylor expansion. One obvious choice is the posterior mode, denoted as $\widehat{\eta}$, which has the advantage that it can be easily obtained via the Newton-Raphson method. More specifically, it follows from (9) that the negative Hessian of $\log p(\eta \mid y, \theta)$ evaluated at $\eta = \widetilde{\eta}$ is H, while the gradient at $\eta = \widetilde{\eta}$ is given by

$$\frac{\partial}{\partial \eta} \log p(\eta \mid y, \theta) \bigg|_{\eta = \widetilde{\eta}} = -H\widetilde{\eta} + 2(f + G\widetilde{\eta} + K'\Omega K\eta^0).$$

9

Hence, we can implement the Newton-Raphson method as follows: initialise with $\eta = \eta^{(1)}$. For $s = 1, 2, \ldots$, use $\widetilde{\eta} = \eta^{(s)}$ in the evaluation of f, G and H, and denote them as $f(\eta^{(s)})$, $G(\eta^{(s)})$ and $H(\eta^{(s)})$ respectively, where the dependence on $\eta^{(s)}$ is made explicit. Compute $\eta^{(s+1)}$ as

$$
\begin{aligned}
\eta^{(s+1)} &= \eta^{(s)} + H(\eta^{(s)})^{-1} \frac{\partial}{\partial \eta} \log p(\eta \,|\, y, \theta) \bigg|_{\eta = \eta^{(s)}} \\
&= H(\eta^{(s)})^{-1} \left(f(\eta^{(s)}) + G(\eta^{(s)})\eta^{(s)} + K'\Omega K \eta^0 \right).
\end{aligned}
\tag{10}
$$

If $||\eta^{(s+1)} - \eta^{(s)}|| > \epsilon$ for some pre-fixed tolerance level ϵ, then continue; otherwise stop and set $\widehat{\eta} = \eta^{(s+1)}$. Again, it is important to note that because the precision H is banded, and its Cholesky decomposition C_H can be readily obtained. Hence, (10) can be efficiently evaluated without inverting any high-dimensional matrix. More precisely, we compute $\eta^{(s+1)}$ as follows: given the Cholesky decomposition C_H for $H(\eta^{(s)})$, first solve $C_H^{(s)'} x = f\left(\eta^{(s)}\right) + G(\eta^{(s)})\eta^{(s)} + K'\Omega K \eta^0$ for x by forward-substitution. Then given x, solve $C_H \eta^{(s+1)} = x$ for $\eta^{(s+1)}$ by back-substitution. Finally, given the mode $\widehat{\eta}$, the negative Hessian H at $\widehat{\eta}$ can be easily computed.

3.2 Efficient Simulation for the States

Building upon the Gaussian approximation presented previously, we propose three different sampling schemes for drawing the states efficiently.

3.2.1 Metropolis-Hastings with Gaussian and t proposals

A simple sampling scheme is to implement a Metropolis-Hastings step with proposal density $\mathsf{N}(\widehat{\eta}, H^{-1})$. The mode $\widehat{\eta}$ and the negative Hessian at $\widehat{\eta}$ of the conditional density $p(\eta \,|\, y, \theta)$ can be computed as discussed in earlier. Moreover, a draw from the proposal can be obtained using the precision-based sampler as in Chan and Jeliazkov (2009b). Using a Gaussian approximation would be adequate in models where either the measurement or the state equations is Gaussian, as the resulting conditional posterior in either case has exponentially decaying tails. We summarise this basic sampling scheme as follows:

Algorithm 1. *Metropolis-Hastings with the Gaussian Proposal* $\mathsf{N}(\widehat{\eta}, H^{-1})$

1. *Obtain $\widehat{\eta}$ iteratively via* (10). *Given H, compute its Cholesky decomposition C_H such that $H = C_H' C_H$.*

2. *Sample $u \sim \mathsf{N}(0, I_{Tm})$, and solve $C_H x = u$ for x by back-substitution. Take $\eta = \widehat{\eta} + x$, so that $\eta \sim \mathsf{N}(\widehat{\eta}, H^{-1})$.*

In implementing the Metropolis-Hastings algorithm, it is often suggested that the proposal density $q(\eta \,|\, y, \theta)$ should have heavier tails than the posterior distribution $p(\eta \,|\, y, \theta)$, so that the likelihood ratio $p(\eta \,|\, y, \theta)/q(\eta \,|\, y, \theta)$ is bounded. This is important because a bounded likelihood

ratio ensures the geometric ergodicity of the Markov chain (Roberts and Rosenthal, 2004). In the context of estimating the integrated likelihood, this guarantees the estimator has finite variance. Thus, one concern of using a Gaussian proposal is that it has exponentially decaying tails, and consequently, the likelihood ratio might not be bounded. This motivates using a proposal density with heavier tails, such a t distribution. We note that one can easily modify the above Gaussian approximation to obtain a t proposal density instead. More explicitly, consider the t proposal $\eta \sim q(\eta \,|\, y, \theta) \equiv \mathsf{t}(\nu, \widehat{\eta}, H^{-1})$ with degree of freedom parameter ν, location vector $\widehat{\eta}$ and scale matrix H^{-1}. Note that sampling from $\mathsf{t}(\nu, \widehat{\eta}, H^{-1})$ involves only Tm iid standard Gaussian draws and a draw from the $\mathsf{Gamma}(\nu/2, \nu/2)$ distribution. We summarise the algorithm as follows:

Algorithm 2. *Metropolis-Hastings with the t proposal $\mathsf{t}(\nu, \widehat{\eta}, H^{-1})$*

1. *Given the posterior mode $\widehat{\eta}$ and negative Hessian H, obtain the Cholesky decomposition C_H such that $H = C_H' C_H$.*

2. *Sample $u \sim \mathsf{N}(0, I_{Tm})$ and $r \sim \mathsf{Gamma}(\nu/2, \nu/2)$. Then $v \equiv u/\sqrt{r} \sim \mathsf{t}(\nu, 0, I_{Tm})$.*

3. *Solve $C_H x = v$ for x by back-substitution and take $\eta = \widehat{\eta} + x$, so that $\eta \sim \mathsf{t}(\nu, \widehat{\eta}, H^{-1})$.*

3.2.2 Accept-Reject Metropolis-Hastings

As its name suggests, the accept-reject Metropolis-Hastings (ARMH) algorithm (Tierney, 1994; Chib and Greenberg, 1995) is an MCMC sampling procedure that combines classic accept-reject sampling with the Metropolis-Hastings algorithm. In the our setting the target density is the conditional density of the states $p(\eta \,|\, y, \theta) \propto p(y \,|\, \eta, \theta) p(\eta \,|\, \theta)$. Suppose we have a proposal density $q(\eta \,|\, y, \theta)$ from which we generate candidate draws (e.g. $q(\eta \,|\, y, \theta)$ can be the Gaussian or t density discussed in the previous section). In the classic accept-reject sampling a key requirement is that there exists a constant c such that

$$p(y \,|\, \eta, \theta) p(\eta \,|\, \theta) \leqslant c q(\eta \,|\, y, \theta) \tag{11}$$

for all η in the support of $p(\eta \,|\, y, \theta)$. When η is a high-dimensional vector, as in the present case, such a constant c, if it exists, is usually difficult to obtain. To make matters worse, the target density $p(\eta \,|\, y, \theta)$ depends on other model parameters θ that are revised at every iteration. Finding a new value of c for each new set of parameters might significantly increase the computational costs. The ARMH relaxes the domination condition (11) such that when it is not satisfied for some η, we resort to the MH algorithm. To present the algorithm, it is convenient to first define the set

$$\mathcal{D} = \{\eta : p(y \,|\, \eta, \theta) p(\eta \,|\, \theta) \leqslant c q(\eta \,|\, y, \theta)\},$$

and let $\mathcal{D}^c$ denote its complement. Then the ARMH algorithm proceeds as follows:

Algorithm 3. *Accept-Reject Metropolis-Hastings with Gaussian or t proposal*

1. *AR step: Generate a draw $\eta^* \sim q(\eta \,|\, y, \theta)$, where $q(\eta \,|\, y, \theta)$ is the Gaussian or t proposal obtained in Algorithms 1 or 2. Accept η^* with probability*

$$\alpha_{AR}(\eta^* \,|\, y, \theta) = \min\left\{1, \frac{p(y \,|\, \eta^*, \theta)p(\eta^* \,|\, \theta)}{cq(\eta^* \,|\, y, \theta)}\right\}.$$

Continue the above process until a draw η^* is accepted.

2. *MH step: Given the current draw η and the proposal η^**

   (a) *if $\eta \in \mathcal{D}$, set $\alpha_{MH}(\eta, \eta^* \,|\, y, \theta) = 1$;*
   (b) *if $\eta \in \mathcal{D}^c$ and $\eta^* \in \mathcal{D}$, set*

$$\alpha_{MH}(\eta, \eta^* \,|\, y, \theta) = \frac{cq(\eta \,|\, y, \theta)}{p(y \,|\, \eta, \theta)p(\eta \,|\, \theta)};$$

   (c) *if $\eta \in \mathcal{D}^c$ and $\eta^* \in \mathcal{D}^c$, set*

$$\alpha_{MH}(\eta, \eta^* \,|\, y, \theta) = \min\left\{1, \frac{p(y \,|\, \eta^*, \theta)p(\eta^* \,|\, \theta)q(\eta \,|\, y, \theta)}{p(y \,|\, \eta, \theta)p(\eta \,|\, \theta)q(\eta^* \,|\, y, \theta)}\right\}.$$

   Return η^* with probability $\alpha_{MH}(\eta, \eta^* \,|\, y, \theta)$; otherwise return η.

As shown in Chib and Greenberg (1995), the draws produced at the completion of the AR step have the density

$$q_{\mathrm{AR}}(\eta \,|\, y, \theta) = d^{-1}\alpha_{\mathrm{AR}}(\eta \,|\, y, \theta)q(\eta \,|\, y, \theta),$$

where d is the normalising constant (which needs not be known for implementing the algorithm). In other words, one might view the AR step as a means to sample from the density $q_{\mathrm{AR}}(\eta \,|\, y, \theta)$. By adjusting the original proposal density $q(\eta \,|\, y, \theta)$ by the function $\alpha_{\mathrm{AR}}(\eta \,|\, y, \theta)$, a better approximation of the target density is achieved. In fact, we have

$$q_{\mathrm{AR}}(\eta \,|\, y, \theta) = \begin{cases} p(y \,|\, \eta, \theta)p(\eta \,|\, \theta)/cd, & \eta \in \mathcal{D}, \\ q(\eta \,|\, y, \theta)/d, & \eta \in \mathcal{D}^c, \end{cases}$$

i.e., the new proposal density coincides with the target density on the set $\mathcal{D}$ (albeit with different normalising constants), whereas on $\mathcal{D}^c$ the new proposal density is reduced to the original one. To give a feeling for the improvement this approach brings, consider Figure 2. In this figure, the true density is shown as a grey shaded area and the Gaussian candidate by the dotted line. The candidate density $q_{\mathrm{AR}}(\eta \,|\, y, \theta)$ is shown as the solid line which fits the true density for values less than 1.6, then differs above this point but still fits better than the Gaussian density. The better approximation, of course, comes at a cost, because multiple draws from the proposal density $q(\eta \,|\, y, \theta)$ might be required in the AR step. This is where the precision-based method (as in Algorithms 1 or 2) comes in. As we have emphasized before, the marginal cost of generating additional draws using the precision-based method is low, and is substantially lower than generating candidate draws via Kalman filter-based algorithms. In fact, as demonstrated
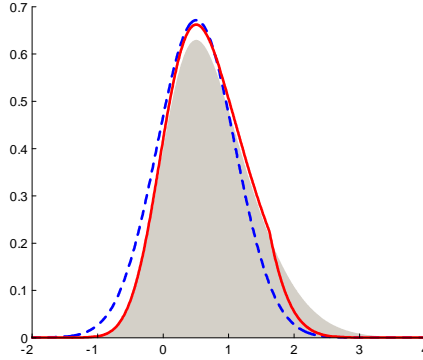
Figure 2: Illustration of the two approximations for a skew normal distribution (shaded): Gaussian (dotted line) and the same Gaussian with AR adjustment (solid line).

in the application, the gain in efficiency under the ARMH sampling scheme more than justifies its additional cost compared to a plain MH step.

Chib and Jeliazkov (2005) present a practical way to select the constant c and the trade-off in such a choice which we outline here. Notice that if a bigger c is chosen, then the set $\mathcal{D}$ is larger and we are more likely to accept the candidate η^*. The cost, on the other hand, of selecting a larger c is that more draws from $q(\eta \mid y, \theta)$ are required in the AR step. A practical way to strike a balance between these two conflicting considerations is to set $c = rp(y \mid \widehat{\eta}, \theta)p(\widehat{\eta} \mid \theta)/q(\widehat{\eta} \mid y, \theta)$, where $\widehat{\eta}$ is the mode of the conditional density $p(\eta \mid y, \theta)$ and r is, say, between 1 and 5. Such a choice would ensure that c is sufficiently small to reduce the required number of draws from $q(\eta \mid y, \theta)$, while big enough so that the set $\mathcal{D}$ contains the mode $\widehat{\eta}$ and its neighboring points.

3.2.3   Collapsed Sampling with the Cross-entropy Method

We have so far discussed two sampling schemes for efficient simulation from the conditional density $p(\eta \mid y, \theta)$: the MH and the ARMH algorithms with either a Gaussian or a t proposal. In performing a full Bayesian analysis, one often sequentially draws from $p(\eta \mid y, \theta)$ followed by sampling from $p(\theta \mid y, \eta)$. In typical situations where θ contains parameters in the state equation, η and θ are expected to be highly correlated. Consequently, the conventional sampling scheme might induce high autocorrelation and slow mixing in the Markov chain, especially in high-dimensional settings. For this reason, we seek to sample (θ, η) jointly by first drawing from $p(\theta \mid y)$ marginally of the states η followed by a draw from $p(\eta \mid y, \theta)$, where the latter step can be accomplished by either the MH or ARMH algorithm previously discussed. To sample from $p(\theta \mid y)$, we again implement a MH step: we first generate a candidate draw θ^* from the proposal density $g(\theta)$, then we decide whether to accept θ^* or not according to the acceptance probability. Hence, we need two ingredients: (1) a quick routine to evaluate the integrated likelihood $p(y \mid \theta)$, which arises in computing the acceptance probability; and (2) a good proposal density $g(\theta)$ for generating candidate draws for the MH step.

13

The first ingredient, an efficient method to evaluate the integrated likelihood, is provided by the importance sampling estimator $\widehat{p}(y|\theta)$ discussed in Appendix C (or see, for example, Geweke (1989) or Kroese *et al.* (2011, ch. 9)). And this in turn gives us an estimator for the acceptance probability

$$\alpha\left(\theta\,|\,y\right) = \min\left\{1, \frac{p(y\,|\,\theta^*)p\left(\theta^*\right)g(\theta)}{p(y\,|\,\theta)p(\theta)g(\theta^*)}\right\}.$$

One might raise the concern that the simulation error may affect the convergence properties of the Markov chain, as the candidate draws are accepted or rejected according to *estimated* acceptance probabilities rather than the actual values. However, since the importance sampling estimator $\widehat{p}(y\,|\,\theta)$ is unbiased, the results in Andrieu, Berthelsen, Doucet, and Roberts (2007) and Flury and Shephard (2008) show that the stationary distribution of the constructed Markov chain is the posterior distribution as desired.

The second ingredient is a proposal density for generating candidate draws for θ. Of course, one may generate candidates via a random walk, but this strategy is not recommended as a random walk chain is typically inefficient, and it would defeat the purpose — to improve the mixing properties of the Markov chain — of the whole exercise. Therefore, it is essential to locate a good proposal density $g(\theta)$ to implement an independence-chain MH step. We adopt the so-called cross-entropy adaptive independence sampler introduced in Keith, Kroese, and Sofronov (2008). Specifically, the proposal density is chosen such that the *Kullback-Leibler divergence*, or the *cross-entropy* (CE)*distance* between the proposal density and the target (the posterior density) is minimal, where the CE distance between the densities g_1 and g_2 is defined as:

$$\mathcal{D}(g_1, g_2) = \int g_1(x) \log \frac{g_1(x)}{g_2(x)} \mathrm{d}x.$$

Let $\mathcal{G}$ be a parametric family of densities $g(\cdot;\, v)$ indexed by the parameter vector v. Minimizing the CE distance is equivalent to finding

$$v_{\mathrm{ce}} = \operatorname*{argmax}_{v} \int p(\theta\,|\,y) \log g(\theta;\, v)\,\mathrm{d}\theta.$$

As in the CE method (Rubinstein and Kroese, 2004; Kroese, Taimre, and Botev, 2011, ch. 13), we can estimate the optimal solution v_{ce} by

$$\widehat{v}_{\mathrm{ce}} = \operatorname*{argmax}_{v} \frac{1}{N} \sum_{i=1}^{N} \log g(\theta_i;\, v), \tag{12}$$

where $\theta_1, \ldots, \theta_N$ are draws from the marginal posterior density $p(\theta\,|\,y)$. The solution to the maximization problem in (12) is typically easy to obtain; in fact, analytic solutions are often available. On the other hand, finding $\widehat{v}_{\mathrm{ce}}$ requires a pre-run to obtain a small sample from $p(\theta\,|\,y)$. This can be achieved by sequentially drawing from $p(\eta\,|\,y,\theta)$ and $p(\theta\,|\,y,\eta)$, as discussed in the previous section. It is important to note that although the sample obtained in this pre-run may exhibit slow mixing, we only use it to obtain the proposal density, and thus it has little adverse effect on the main collapsed sampler. Once we find $\widehat{v}_{\mathrm{ce}}$, we then use the proposal density $g(\theta; \widehat{v}_{\mathrm{ce}})$ to implement the independence-chain MH step. We discuss in more details the implementation in Appendix B.

3.3 Estimation of the TVP-VAR-ZLB

In what follows, we briefly discuss the implementation of the three samplers to estimate the TVP-VAR-ZLB; we refer the readers to Appendices A and B for more details. The first sampling scheme is the baseline Metropolis-Hastings sampler that involves sequentially drawing from:

a. $p(\beta \,|\, y, h, a, \theta)$ via an MH step;

b. $p(h \,|\, y, \beta, a, \theta)$ via an MH step;

c. $p(a \,|\, y, \beta, h, \theta)$ via a Gibbs step;

d. $p(\theta \,|\, y, \beta, h, a)$ via a Gibbs step.

To efficiently sample the states β in the non-linear state space model (4) and (5), we consider implementing an independence-chain MH step by approximating the conditional distribution $p(\beta \,|\, y, h, a, \theta)$ via a Gaussian distribution as discussed in Section 3.1. The next step is to sample from the conditional distribution $p(h \,|\, y, \beta, a, \theta)$. Recall that h_{it} is the i-th diagonal element in D_t, $h_{\bullet t} = (h_{1t}, \ldots, h_{nt})'$ and $h_{i\bullet} = (h_{i1}, \ldots, h_{iT})'$. Note that we are able to write $\log p(h \,|\, y, a, \beta, \theta) = \sum_{i=1}^{n} \log p(h_{i\bullet} \,|\, y, a, \beta, \theta)$. What this means is that to obtain a draw from $p(h \,|\, y, a, \beta, \theta)$, we can instead sample from $p(h_{i\bullet} \,|\, y, a, \beta, \theta)$ sequentially without adversely affecting the efficiency of the sampler. Now, a draw from $p(h_{i\bullet} \,|\, y, a, \beta, \theta)$ can be obtained via an independence-chain Metropolis-Hastings step with a Gaussian proposal density; more details are given in Appendix A. Thirdly, it can be easily shown that $p(a \,|\, y, \beta, h, \theta)$ is a Gaussian distribution (see, e.g. Primiceri, 2005), and a draw from which can be obtained using the precision-based sampler. Finally, $p(\theta \,|\, y, \beta, h, a)$ is a product of Gamma densities, and a draw from which is standard (see Koop, 2003, p. 61-62).

In the second sampling scheme, we also sequentially draw from the four full conditional densities as before. The only difference is that instead of using the MH algorithm to sample $p(\beta \,|\, y, h, a, \theta)$ and $p(h \,|\, y, \beta, a, \theta)$, we use the ARMH algorithm described in Section 3.3.2. Finally, in the third sampling scheme, we sample

a. $p(\omega_\beta \,|\, y, h, a)$ marginally via an MH step, followed by $p(\beta \,|\, y, h, a, \theta)$ via an ARMH step;

b. $p(\omega_h \,|\, y, \beta, a)$ marginally via an MH step, followed by $p(h \,|\, y, \beta, a, \theta)$ via an ARMH step;

c. $p(\omega_a \,|\, y, \beta, h)$ marginally via an MH step, followed by $p(a \,|\, y, \beta, h, \theta)$ via a Gibb step.

The details for the collapsed sampler are given in Appendix B.

4 Empirical Results

We now present empirical results based on a set of U.S. macroeconomic variables commonly used in the study of the evolution of monetary policy transmission. We have an interest rate

to capture effects of monetary conditions, a real growth rate variable to capture the state of the economy, and inflation. The dataset is obtained from the U.S. Federal Reserve Bank at St. Louis website that consists quarterly observations from 1947Q1 to the 2011Q2 on the following $n = 3$ U.S. macroeconomic series: U.S. 3-month Treasury bill rate, CPI inflation rate, and real GDP growth. Both the CPI inflation rate and real GDP growth are computed via the formula $400(\log(z_t) - \log(z_{t-1}))$, where z_t is the original quarterly CPI or GDP figures. The inclusion of the interest rate variable, which is bounded below at zero, provides a useful example to demonstrate our methods. The plot of the evolution of the interest rate, given in Figure 3, shows that since the start of the quantitative easing in late 2008, the 3-month Tbill rate has become essentially zero.
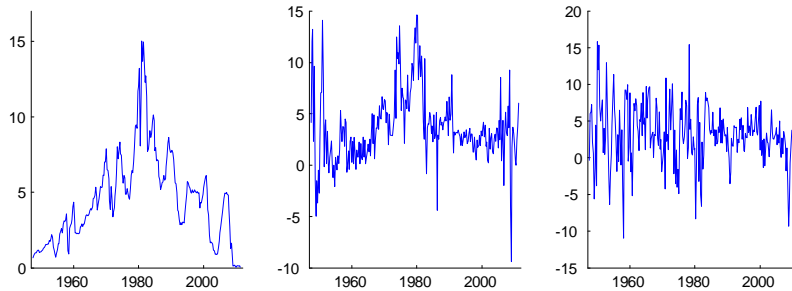

Figure 3: The U.S. 3-month Tbill rate (left), CPI inflation rate (middle), and GDP growth rate (right) from 1947 Q1 to 2011 Q2.

We allow one lag in the VAR as this seems sufficient to capture much of the dynamics. We begin with a comparison of performance of the three sampling schemes. To this end, we estimate the *restricted* model using the three different sampling schemes outlined in the previous section: the Metropolis-Hastings sampler (S1), the Accept-Reject Metropolis-Hastings sampler (S2), and the collapsed sampler with cross-entropy method (S3). We also include results from the *unrestricted* model (U), for which both the transition and measurement equations are linear Gaussian. As such, this unrestricted model can be estimated using standard algorithms.

One popular measure of MCMC efficiency is the inefficiency factor, defined as:

$$1 + 2\sum_{j=1}^{J} \phi_j,$$

where ϕ_j is the sample autocorrelation at lag length j, and J is chosen large enough so that the autocorrelation tapers off. This statistic approximates the ratio of the numerical variance of the posterior mean from the MCMC output relative to that from hypothetical iid draws. As the posterior draws from the Markov chain become less serially correlated, the ratio will approach the ideal minimum value of 1. In the presence of inefficiency due to serial correlation in the draws, the ratio will be larger than 1. Figure 4 presents the boxplots of the inefficiency factors for the four sampling schemes.
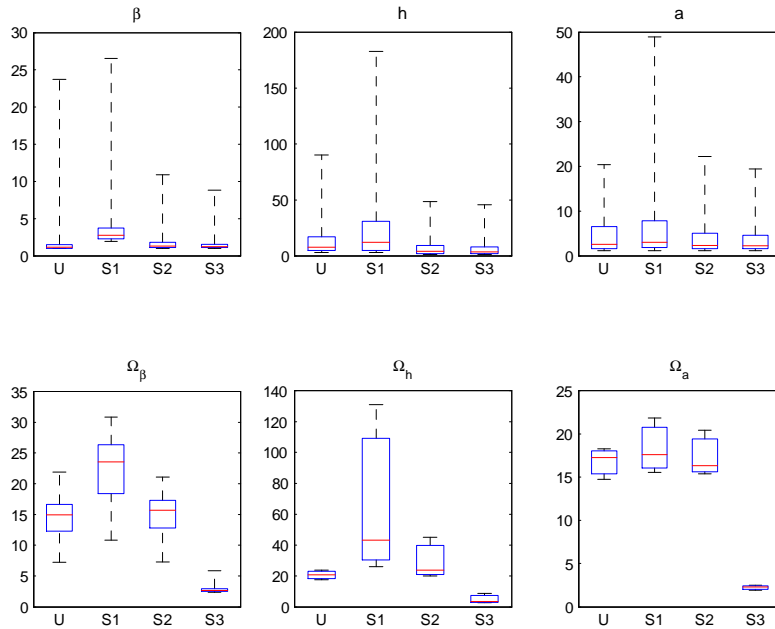
16

Figure 4: Boxplots of the inefficiency factors for the unrestricted linear Gaussian model (U), and the three sampling schemes: MH (S1), ARMH (S2) and the collapsed sampler (S3). The central mark of each box is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the maximum and minimum.

Remember that the unrestricted model is linear Gaussian and can be estimated via a standard Gibbs sampler. In contrast, the restricted model that incorporates the ZLB is non-linear, and the conditional densities of the states are non-standard. Since the proposed samplers need to approximate these conditional densities, we would generally expect that they would not perform as well compared to the standard Gibbs sampler used to estimate the *unrestricted* model. As evidenced by the plots in Figure 4, the proposed samplers do not perform substantially worse at estimating the non-linear model than the standard precision sampler for fitting the linear Gaussian model. The collapsed sampler with cross-entropy, S3, has inefficiency factors as small as or smaller than the other samplers for most parameters, including the standard sampler for the unrestricted model. The improvement in efficiency for S3 is most substantial in the precision of the state parameters, which is significant as these are hyperparameters and are typically not as well estimated as parameters that appear in the measurement equation. These hyperparameters are also important as they play an important role in the estimation of the states. We see that S1 (the MH sampler) is generally not as efficient compared to the other samplers, although its performance is not greatly worse than the others. The efficiency of S2 (the ARMH sampler) is as good as S3 for the states, but it is worse than S3 for the estimation of the state precisions.

17

Table 1: Acceptance rate (in %) and the computing time (in minutes) of the three sampling schemes: MH (S1), ARMH (S2) and the collapsed sampler with CE (S3).

	β	$h_{1.}$	$h_{2.}$	$h_{3.}$	Ω_β	Ω_h	Ω_a	Time
S1	68	28	35	59	–	–	–	23
S2	95	71	79	97	–	–	–	27
S3	98	69	79	97	62	58	76	182

In Table 1 we present the acceptance rates of draws from candidates, as well as the computation time for obtaining 50,000 draws. On the whole the three samplers are relatively fast and have reasonable acceptance rates. These results are more significant given our high-dimensional model: β has more than 3,000 elements and h has more than 750. To compare among the three sampling schemes, we see that although S1 is relatively fast, it can have low acceptance rates particularly for the log volatilities. By contrast, S2 has higher acceptance rates at the expense of only a little more computation time. Although S3 is more efficient relative to S2 in terms of lower inefficiency factors, its computation time is almost seven times compared to that of S2. For our model and dataset, it would seem that S2 is the best among the three.

We now present empirical results for the restricted model estimated using the ARMH sampler (S2). For comparison, we also report the corresponding results for the unrestricted model. This comparison is provided to demonstrate the implications for inference of neglecting the restriction of the ZLB. We begin with a discussion of the implications of the restriction for parameter estimation and then show the effect on impulse responses of not correctly accounting for the ZLB. These differences are significant and justify the new estimation methods presented in this paper.

The effect of neglecting the ZLB restriction shows up in all blocks of parameters: the variances; the correlations; and mean equation coefficients. Figure 5 shows the estimated log-volatilities and correlations for the restricted model and the unrestricted model. The figure for the log-volatilities of the monetary shock, $h_{1,t}$, shows that ignoring the restriction would lead to a significant underestimation of this parameter in the period since 2005. The monetary shock volatility is much higher than the unrestricted model suggests. Similarly the volatility of real activity, $h_{3,t}$, is over estimated when the ZLB is ignored. The volatility and correlations of the nominal variable shock does not show much influence from the ZLB. However, the correlation of the error from the interest rate equation with the error from the growth equation is strongly affected, this correlation would be estimated as being near zero rather than very negative. This effect has important implications for the impulse responses.

The plots in Figure 6 show the effect of the ZLB on the impulse response functions of the three variables to a monetary shock. In the introduction we made the point that in a state of low output and prices growth, when interest rates are low, shocks to the interest rate equatoin, such as credit shocks, would be stronger have a more immediate effect. For this reason we identify the credit shock by allowing it to have a contemporaneous effect on all three variables.[4] We produce

---

[4]Note that the ordering used in the estimation does not dictate the identification. We can use estimates to construct the precision matrix as $\Sigma_t = L_t' D_t^{-1} L_t$, and then take whatever appropriate decomposition of the
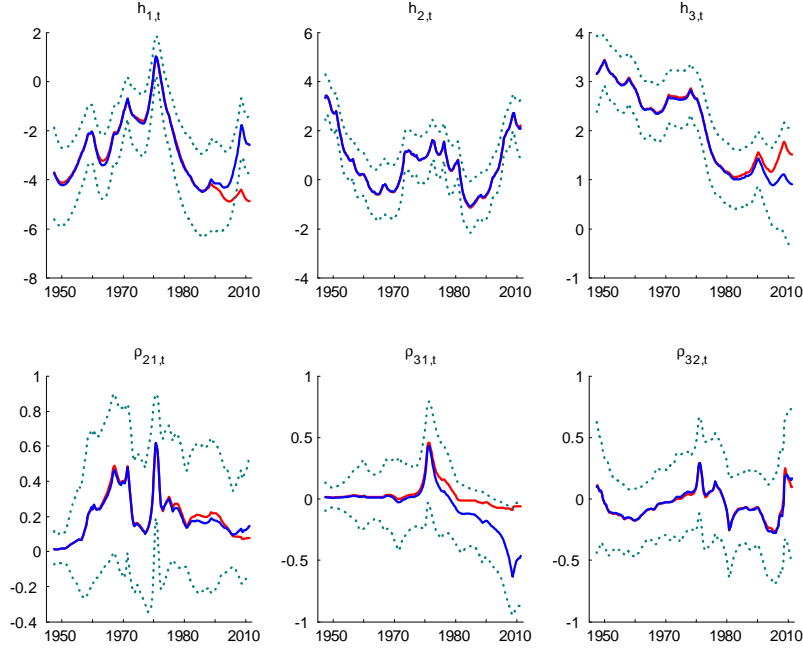
Figure 5: Evolution of the log-volatilities and correlations. The solid red line is the estimated posterior mean under the unrestricted model. The solid blue line is the estimated posterior mean and the dotted green lines are the 5%-tile and 95%-tile, respectively, under the model with the inequality restrictions imposed.

the impulse responses as the differences in forecasts. However, due to the non-linear form of the model, these impulses are not the standard ones derived from the VMA representation in linear models. We forecast the variables using the parameter values at 2011 Q2, assuming the parameters cease evolving, but taking into account the ZLB into these forecasts. We then forecast again, but increase the error in the interest rate equation, the monetary shock, by 0.25%. Our impulse responses are the differences between these two forecasts. We see that there is a faster response of interest rates to the monetary shock, but the form of the response is similar.

The responses of inflation and growth to this shock are very different with the ZLB imposed compared to that without the restriction. As these results come from a time varying parameter model, it is important to interpret these responses in the context of the economic environment at the time. In 2011 Q2 inflation and inflation volatility were increasing, while growth and volatility of the error in the growth equation were both low. The response of inflation with the ZLB is initially positive but then falls by a larger amount than the initial response. This pattern contrasts with the steady decline toward zero from the initial shock we would conclude

covariance matrix Σ_t^{-1} that accords with our chosen identification scheme.
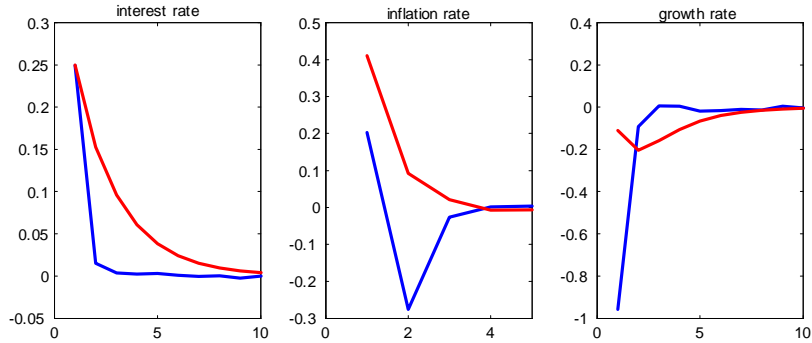
Figure 6: Impulse response of a 0.25% increase in interest rate under the unrestricted model (red solid line) and the model with the inequality restrictions imposed (blue solid line).

was the response from the unrestricted model. The difference in the response of growth to this shock under the two specifications is even more stark. While the shock is negative but small for the unrestricted model, when we account for the ZLB we see the initial response is negative and very large. The differences of these responses would lead to very different assessments of the risks of unanticipated monetary shocks.

5   Concluding Remarks and Further Research

In this paper we have proposed a new approach to efficiently estimate high-dimensional non-linear non-Gaussian state space models. Due to the general applicability of the proposed approach, it will prove useful in a wide range of applications. We extend the recently developed precision-based samplers (Chan and Jeliazkov, 2009b and McCausland, Millera, and Pelletier, 2011) and sparse matrix procedures to build fast, efficient samplers for these non-linear models. We develop a practical way to sample the model parameters θ and the states η jointly to circumvent the problem of high autocorrelations in high-dimensional settings. This approach uses the cross-entropy method (Rubinstein and Kroese, 2004) to obtain the optimal candidate densities $q(\theta \,|\, y)$. We show via an empirical example that the efficiency of the sampling scheme is substantially improved by drawing (θ, y) jointly. Three samplers are presented each with virtues in different circumstances. Finally, we apply these techniques in a TVP-VAR in which one of the variables is restricted to be strictly positive. Using this framework, we investigate the implications for transmission of monetary shocks of accounting for the zero lower bound (ZLB) on interest rates.

Another advantage of the proposed method is that it can be applied to non-Markovian state equations, which arise in, e.g., various non-linear DSGE models, and they are more difficult to handle under other approaches. Therefore in future work we will apply this approach to the estimation of non-linear DSGE models. Another direction will be in models with measurement equations that involving more than current, past or even future states.

20

Appendix A: Efficient Simulation of β and h

In this appendix we provide the details of the independence-chain Metropolis-Hastings step for sampling from $p(\beta \mid y, h, a, \theta)$ and $p(h \mid y, a, \beta, \theta)$. We use the decomposition $\Sigma_t = L_t' D_t^{-1} L_t$, where

$$
D_t = \begin{pmatrix} e^{h_{1t}} & 0 & \cdots & 0 \\ 0 & e^{h_{2t}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{h_{nt}} \end{pmatrix}, \quad
L_t = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{21,t} & 1 & 0 & \cdots & 0 \\ a_{31,t} & a_{32,t} & 1 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1,t} & a_{n2,t} & \cdots & a_{n(n-1),t} & 1 \end{pmatrix}.
$$

Recall that h_{it} is the i-th diagonal element in D_t, $h_{\cdot t} = (h_{1t}, \ldots, h_{nt})'$ and $h_{i\cdot} = (h_{i1}, \ldots, h_{iT})'$. That is, $h_{\cdot t}$ is the $n \times 1$ vector obtained by stacking h_{it} by the first subscript, whereas $h_{i\cdot}$ is the $T \times 1$ vector obtained by stacking h_{it} by the second subscript. Also, a_t denotes the free elements in L_t ordered by rows, i.e., $a_t = (a_{21,t}, a_{31,t}, a_{32,t}, \ldots, a_{n(n-1),t})'$. In what follows we use the two parameterizations Σ_t and $(h_{\cdot t}, a_t)$ interchangeably. Then the log-density for y_t given (β_t, Σ_t) is

$$
\log p(y_t \mid \beta_t, \Sigma_t) \propto -\frac{1}{2}(y_t - x_t \beta_t)' \Sigma_t (y_t - x_t \beta_t) - \log \Phi(\alpha_t),
$$

where $\alpha_t = x_{1t} \beta_t e^{-\frac{1}{2} h_{1t}}$. Using the notation in Section 3.1, we have

$$
f_t \equiv \left. \frac{\partial}{\partial \beta_t} \log p(y_t \mid \beta_t, \Sigma_t) \right|_{\beta_t = \widetilde{\beta}_t}, \quad
G_t \equiv \left. -\frac{\partial^2}{\partial \beta_t \beta_t'} \log p(y_t \mid \beta_t, \Sigma_t) \right|_{\beta_t = \widetilde{\beta}_t},
$$

where

$$
\frac{\partial}{\partial \beta_t} \log p(y_t \mid \beta_t, \Sigma_t) = x_t' \Sigma_t (y_t - x_t \beta_t) - \frac{\phi(\alpha_t)}{\Phi(\alpha_t)} e^{-\frac{1}{2} h_{1t}} x_{1t}',
$$

$$
\frac{\partial^2}{\partial \beta_t \beta_t'} \log p(y_t \mid \beta_t, \Sigma_t) = -x_t' \Sigma_t x_t + \frac{\phi(\alpha_t)}{\Phi(\alpha_t)} e^{-h_{1t}} \left(\alpha_t + \frac{\phi(\alpha_t)}{\Phi(\alpha_t)} \right) x_{1t}' x_{1t},
$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the standard Gaussian probability density function and cumulative distribution function respectively. Given f_t and G_t, we can then use the Gaussian or t approximations in Section 3.1 as a proposal density.

We now discuss sampling from the conditional density $p(h \mid y, a, \beta, \theta)$. We first show that

$$
\log p(h \mid y, a, \beta, \theta) = \sum_{i=1}^{n} \log p(h_{i\cdot} \mid y, a, \beta, \theta).
$$

Put differently, to obtain a draw from $p(h \mid y, a, \beta, \theta)$, we can instead sample from $p(h_{i\cdot} \mid y, a, \beta, \theta)$ sequentially without adversely affecting the efficiency of the sampler. To this end, decompose $\Sigma_t = L_t' D_t^{-1} L_t$ as before. Since $\log |\Sigma_t| = \log |D_t| = \sum_{i=1}^{n} h_{it}$ and $\sigma_{t,11}^2 = e^{h_{1t}}$, it follows that

the log-likelihood is given by

$$\log p(y \mid \beta, h, a, \theta) \propto \sum_{t=1}^{T} \left[-\frac{1}{2} \sum_{i=1}^{n} h_{it} - \frac{1}{2}(L_t \epsilon_t)' D_t^{-1} L_t \epsilon_t - \log \Phi \left(e^{-h_{1t}/2} x_{1t} \beta_t \right) \right],$$

$$= \sum_{t=1}^{T} \left[-\frac{1}{2} \sum_{i=1}^{n} h_{it} - \frac{1}{2} \sum_{i=1}^{n} e^{-h_{it}} s_{it}^2 - \log \Phi \left(e^{-h_{1t}/2} x_{1t} \beta_t \right) \right], \quad (13)$$

where $\epsilon_t = y_t - x_t \beta_t$ and s_{it}^2 is the i-th diagonal element of $(L_t \epsilon_t)(L_t \epsilon_t)'$. On the other hand, the state equation (2) implies that each $h_{i\cdot}$ follows independently a Gaussian distribution. In fact, we have

$$h_{it} = h_{i,t-1} + \xi_{it}, \quad \xi_{it} \sim \mathsf{N}(0, \omega_{hi}). \quad (14)$$

Hence, it follows from (13) and (14) that $h_{i\cdot}, i = 1, \ldots, n$ are conditionally independent given the data and other parameters.

We note that although one can apply the auxiliary variable approach in Kim, Shepherd, and Chib (1998) to sample from $p(h_{i\cdot} \mid y, a, \beta, \theta)$ for $i = 2, \ldots, n$, it cannot be used to draw from $p(h_{1\cdot} \mid y, a, \beta, \theta)$ due to the extra term $\log \Phi \left(e^{-h_{1t}/2} x_{1t} \beta_t \right)$ in the log-likelihood (13) that depends on h_{1t}. Instead, we sample each $h_{i\cdot}$ sequentially via an independence-chain Metropolis-Hastings step. As before, we first derive an expression for a second order Taylor expansion of the log-likelihood (13) around the posterior mode $\widehat{h}_{i\cdot} = (\widehat{h}_{i1}, \ldots, \widehat{h}_{iT})'$. Define $\gamma_t = e^{-h_{1t}/2} x_{1t} \beta_t$

$$q_{it} = \left. \frac{\partial}{\partial h_{it}} \log p(y \mid \beta, h, a, \theta) \right|_{h_{it} = \widehat{h}_{it}}, \quad r_{it} = \left. -\frac{\partial^2}{\partial h_{it}^2} \log p(y \mid \beta, h, a, \theta) \right|_{h_{it} = \widehat{h}_{it}},$$

$q_i = (q_{i1}, \ldots, q_{iT})'$ and $R_i = \mathrm{diag}(r_{i1}, \ldots, r_{iT})$, where

$$\frac{\partial}{\partial h_{it}} \log p(y \mid \beta, h, a, \theta) = \frac{1}{2} \left[e^{-h_{it}} s_{it}^2 - 1 + \gamma_t \frac{\phi(\gamma_t)}{\Phi(\gamma_t)} \mathbb{1}(i = 1) \right],$$

and

$$\frac{\partial^2}{\partial h_{it}^2} \log p(y \mid \beta, h, \theta) = -\frac{1}{2} e^{-h_{it}} s_{it}^2 + \frac{1}{4} \gamma_t \frac{\phi(\gamma_t)}{\Phi(\gamma_t)} \left[\gamma_t^2 + \gamma_t \frac{\phi(\gamma_t)}{\Phi(\gamma_t)} - 1 \right] \mathbb{1}(i = 1).$$

If we expand the log-likelihood (13) around the mode $\widehat{h}_{i\cdot}$, we have

$$\log p(y \mid \beta, h, a, \theta) \approx -\frac{1}{2} \left[\widehat{h}_{i\cdot}' R_i \widehat{h}_{i\cdot} - 2 \widehat{h}_{i\cdot}'(q_i + R_i \widehat{h}_{i\cdot}) \right] + c_3,$$

where c_3 is some unimportant constant independent of $\widehat{h}_{i\cdot}$. We consider the proposal density $\mathsf{N}(\widehat{h}_{i\cdot}, (q_i + R_i \widehat{h}_{i\cdot})^{-1})$, and everything follows as before.

Appendix B: Collapsed Sampler with the Cross-Entropy Method

In this appendix we provide the details on the collapsed sampler used in the third sampling scheme. In a nutshell, we first use a small posterior sample from a pre-run and the cross-entropy

method to locate an optimal proposal density within a given parametric family. Then given a candidate draw from the proposal, we implement a Metropolis-Hastings step to decide whether or not to accept the candidate, where the acceptance probability is computed using the importance sampling estimator for the integrated likelihood proposed in Section 3.2. We focus on discussing the approximation to $p(\omega_\beta \,|\, y, h, a)$, where $\omega_\beta = (\omega_{\beta 1}, \ldots, \omega_{\beta k})'$. The approximations to the other two marginal densities follow similarly. Recall that the elements of ω_β have an independent gamma prior: $\omega_{\beta i} \sim \mathsf{Gamma}(r_{\beta i}, s_{\beta i})$ for $i = 1, \ldots, k$. Therefore, a natural parametric family within which to locate the proposal density is the gamma family:

$$\mathcal{G} = \left\{ \prod_{i=1}^{k} f_{\mathsf{G}}(\omega_{\beta i}; c_{\beta i}, d_{\beta i}) \right\},$$

where $f_{\mathsf{G}}(\cdot; c, d)$ is the density of $\mathsf{Gamma}(c, d)$. Given the R posterior draws $\{\omega_{\beta 1}^{(j)}, \ldots, \omega_{\beta k}^{(j)}\}$, $j = 1, \ldots, R$, we solve the CE optimization problem in (12) to obtain $\widehat{v}_{\mathrm{ce}} = (\widehat{c}_{\beta 1}, \widehat{d}_{\beta 1}, \ldots, \widehat{c}_{\beta k}, \widehat{d}_{\beta k})$. Specifically, the optimal CE reference parameter vector $\widehat{v}_{\mathrm{ce}}$ can be obtained as follows. First note that $\widehat{d}_{\beta i}$ can be solved analytically given $c_{\beta i}$:

$$\widehat{d}_{\beta i} = \frac{R c_{\beta i}}{\sum_{j=1}^{R} \omega_{\beta i}^{(j)}}.$$

Now by substituting $d_{\beta i} = \widehat{d}_{\beta i}$ into the density $f_{\mathsf{G}}(\cdot; c_{\beta i}, d_{\beta i})$, $\widehat{c}_{\beta i}$ can be obtained by any one-dimensional root-finding algorithm (e.g., Newton-Raphson method). Hence, we can obtain $(\widehat{c}_{\beta 1}, \widehat{d}_{\beta 1}, \ldots, \widehat{c}_{\beta k}, \widehat{d}_{\beta k})$ easily. Finally, the proposal density is

$$f_\beta(\omega_\beta) = \prod_{i=1}^{k} f_{\mathsf{G}}(\omega_{\beta i}; \widehat{c}_{\beta i}, \widehat{d}_{\beta i}),$$

which is the member within $\mathcal{G}$ that is the closest in cross-entropy divergence to the marginal density $p(\omega_\beta \,|\, y, h, a)$.

Appendix C: Integrated Likelihood Evaluation

The integrated likelihood $p(y \,|\, \theta)$ is defined as the joint distribution of the data conditional on the parameter vector θ but integrated over the states η. More explicitly,

$$p(y \,|\, \theta) = \int p(y \,|\, \theta, \eta) p(\eta \,|\, \theta) \mathrm{d}\eta. \tag{15}$$

The need to evaluate the integrated likelihood efficiently arises in both frequentist and Bayesian estimation. In classical inference, one needs to maximize the integrated likelihood $p(y \,|\, \theta)$ with respect to θ to obtain the maximum likelihood estimator (MLE). In our context one has to compute the MLE numerically, and the maximization routine typically requires hundreds or even thousands of functional evaluations of $p(y \,|\, \theta)$. Hence, it is crucial to be able to evaluate the integrated likelihood efficiently. For Bayesian estimation, if one can evaluate $p(y \,|\, \theta)$

quickly, more efficient samplers can be developed to obtain draws from the posterior, such as the collapsed sampler that draws η and θ jointly in a single step which we discuss in Section 3.2.

Given the Gaussian approximation proposed in the previous section, one can estimate $p(y \,|\, \theta)$ via importance sampling (see, e.g., Geweke, 1989; Kroese *et al.*, 2011, ch. 9). To do this, sample M independent draws $\eta^1, \ldots, \eta^M$ from the proposal density $q(\eta \,|\, y, \theta)$, and compute the Monte Carlo average

$$\widehat{p}(y \,|\, \theta) = \frac{1}{M} \sum_{i=1}^{M} \frac{p(y \,|\, \theta, \eta^i) p(\eta^i \,|\, \theta)}{q(\eta^i \,|\, y, \theta)}.$$

It is easy to see that the Monte Carlo estimator $\widehat{p}(y \,|\, \theta)$ is an unbiased and consistent estimator for $p(y \,|\, \theta)$. In addition, if the likelihood ratio $p(y \,|\, \theta, \eta) p(\eta \,|\, \theta)/q(\eta \,|\, y, \theta)$ or equivalently $p(\eta \,|\, y, \theta)/q(\eta \,|\, y, \theta)$ is bounded for all η, then the variance of the estimator is also finite (Geweke, 1989). The proposed precision-based algorithms are especially fit for evaluating the integrated likelihood via importance sampling. This is because one needs multiple draws (often hundreds or thousands) from the proposal density $q(\eta \,|\, y, \theta)$ to compute the Monte Carlo average. As discussed earlier, one important and useful feature of the precision-based algorithms is that once we obtained the mean vector and precision matrix, additional draws can be obtained with little marginal cost.

References

Andrieu, C. , K. K. Berthelsen, A. Doucet, and G. O. Roberts. The expected auxiliary variable method for Monte Carlo simulation. Technical report, 2007.

Chan, J. C. C., and I. Jeliazkov. MCMC estimation of restricted covariance matrix. Journal of Computational and Graphical Statistics, 18:457–480, 2009a.

Chan, J. C. C., and I. Jeliazkov. Efficient simulation and integrated likelihood estimation in state space models. International Journal of Mathematical modelling and Numerical Optimisation, 1:101–120, 2009b.

Chib, S., and E. Greenberg. Understanding the Metropolis-Hastings algorithm. The American Statistician, 49(4):327–335, 1995.

Chib, S., and I. Jeliazkov. Accept-reject Metropolis-Hastings sampling and marginal likelihood estimation. Statistica Neerlandica, 59:30–44, 2005.

Cogley, T. and T. J. Sargent. Evolving post-World War II inflation dynamics, *NBER Macroeconomic Annual*, 16, 331-373, 2001.

Cogley, T., and T. J. Sargent. Drifts and volatilities: monetary policies and outcomes in the post WWII US. Review of Economic Dynamics, 8(2):262 – 302, 2005.

Flury, T., and N. Shephard. Bayesian inference based only on simulated likelihood: particle

filter analysis of dynamic economic models. Economics SeriesWorking Papers 413, University of Oxford, Department of Economics, 2008.

Geweke, J. Bayesian inference in econometric models using Monte Carlo integration. Econometrica, 57 (6):1317–1339, 1989.

Iwata, S. and S. Wu. Estimating monetary policy effects when interest rates are close to zero. Journal of Monetary Economics, Elsevier, vol. 53(7), pages 1395-1408, October, 2006.

Keith, J. M., D. P. Kroese, and G. Y. Sofronov. Adaptive independence samplers. Statistics and Computing, 18:409–420, 2008.

Kim, S., N. Shepherd, and S. Chib. Stochastic volatility: Likelihood inference and comparison with ARCH models. Review of Economic Studies, 65(3):361–393, 1998.

Koop, G. Bayesian Econometrics. Wiley & Sons, New York, 2003.

Koop G., R. Léon-González and R. W. Strachan. On the Evolution of Monetary Policy. Journal of Economic Dynamics and Control 33, 997-1017, 2009.

Kroese, D. P., T. Taimre, and Z. I. Botev. Handbook of Monte Carlo Methods. John Wiley & Sons, New York, 2011.

Series, 2008-03, 2008.

McCausland, W. J., S. Millera, and D. Pelletier. Simulation smoothing for state-space models: A computational efficiency analysis. Computational Statistics and Data Analysis, 55:199–212, 2011.

Pourahmadi, M. Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. Biometrika, 86:677–690, 1999.

Pourahmadi, M. Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. Biometrika, 87:425–435, 2000.

Primiceri, G. E. Time varying structural vector autoregressions and monetary policy. Review of Economic Studies, 72(3):821–852, 2005.

Reifschneider, D. and J. C. Williams. Three lessons for monetary policy in a low inflation era. Journal of Money, Credit and Banking 32:936-966, 2000.

Roberts, G. O., and J. S. Rosenthal. General state space Markov chains and MCMC algorithms. Probability Surveys, 1:20–71, 2004.

Rubinstein, R. Y., and D. P. Kroese. The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization Monte-Carlo Simulation, and Machine Learning. Springer-Verlag, New York, 2004.

Sims, C. and T. Zha, Were there regime switches in macroeconomic policy? American Economic

Review, 96, 54-81, 2006.

Smith, M., and R. Kohn. Parsimonious covariance matrix estimation for longitudinal data. Journal of the American Statistical Association, 97:1141–1153, 2002.

Tierney, L. Markov chains for exploring posterior distributions. The Annals of Statistics, 22(4):1701– 1728, 1994.