Bayesian Computation in Finance

Satadru Hore[1], Michael Johannes[2]
Hedibert Lopes[3] ,Robert McCulloch[4], and Nicholas Polson[5]

Abstract

In this paper we describe the challenges of Bayesian computation in Finance. We show that empirical asset pricing leads to a nonlinear non-Gaussian state space model for the evolutions of asset returns and derivative prices. Bayesian methods extract latent state variables and estimate parameters by calculating the posterior distributions of interest. We describe the use of direct estimation methods such as Markov chain Monte Carlo (MCMC) and sequential Monte Carlo (SMC) methods based on particle filtering (PF). Our approach also allows for an on-line model assessment via sequential Bayes factors. We illustrate our approach in two examples. First, sequential inference for extracting latent stochastic volatility and jump states from daily data throughout the credit crisis of 2007-2008 and secondly, an equilibrium-based asset pricing model for SP500 put options.

Keywords: Asset pricing, Stochastic Volatility, Financial Econometrics, Markov chain Monte Carlo, Particle filtering, Options.

1 Introduction

Modern-day finance uses arbitrage and equilibrium arguments to derive asset prices as a function of state variables and parameters of the underlying dynamics of the economy. Many applications require extracting information from asset returns and derivative prices such as options or to understand macro-finance models such as consumption-based asset pricing models. To do this the researcher needs to combine information from different sources, asset returns on the one hand and derivative prices on the other. A natural approach to provide inference is Bayesian (Berger, 1985, Bernardo and Smith, 1994 and Gamerman and Lopes, 2007).

Our computational challenges arise from the inherent nonlinearities that arise in the pricing equation, in particular through the dependence on parameters. Duffie (1996) and Johannes and Polson (2009) show that empirical asset pricing problems can be viewed as a nonlinear state space models. These so-called affine models provide a natural framework for addressing the problem as well. Whilst affine pricing models in continuous time so a long way to describe the evolution of derivative prices, empirically extracting the latent state variables and parameters that drive prices has up until now received less attention due to computational challenges. In this paper, we address these challenges by using simulation-based methods, such as Markov chain Monte Carlo (MCMC), Forward filtering backward sampling (FFBS) and particle filter (PF). Hence we solve the inverse problem of filtering state variables and estimating parameters given empirical realizations on returns and derivative prices.

The statistical tools that we describe include MCMC methods, with particular emphasis on the FFBS algorithm of Carter and Kohn (1994) and Frühwirth-Schnatter (1994). For sequential methods we describe PF algorithms, with particular emphasis to the sequential importance sampling with resampling (SISR) filter of Gordon et al. (1993) and the particle learning (PL) algorithm of Lopes et al. (2010).

[1]Federal Reserve Bank of Boston, 600 Atlantic Avenue, Boston, MA 02210.
[2]Graduate School of Business, Columbia University, 3022 Broadway, Uris Hall 424, New York, NY 10027.
[3]The University of Chicago Booth School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637.
[4]McCombs School of Business, University of Texas at Austin, 1 University Station, B6500, Austin, TX 78712.
[5]The University of Chicago Booth School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637.

This current research shows how to also estimate parameters such as agents' preferences from empirical data. In many cases the agents will be given the underlying parameters and the problem becomes one of filtering the hidden states as conditioning information arrives.

The rest of the paper is outlined as follows. Section 2 describes asset pricing problems and Bayesian inference for these problems. Section 3 describes sequential Bayesian computation. Section 4 provides an illustration to equilibrium-based stochastic volatility models. Finally, Section 5 concludes.

2 Empirical Bayesian Asset Pricing

In order to solve the inference problem, namely calculating the joint posterior $p(\theta, X_t|Y^t)$ where $Y^t = (Y_1, \ldots, Y_t)$ is a set of discretely observed returns, we have to pick a suitable time-discretisation of the continuous-time model and then perform Bayesian inference on the ensuing state space model. We use an Euler discretisation.

In order to make our discretiaztion an accurate representation of the continuous time model, we may discretize at a higher fequency than that of the oberved data. In this case, we simulate additional state variables between the observations via the Euler discretisation scheme. This introduces the concept of missing data. These missing data are drawn via a Gibbs step.

One novel feature of Bayesian methods is that they allow data in the form of observations of derivative prices to aid in the estimation problem. For example, suppose that we observe a call price $C(X_t, \theta)$ or a variance swap price. This information can be combined with the current posterior distribution on states and price, namely $p(X_T, \theta|Y^T)$, to obtain sharper parameter estimates.

We now provide the relevant Bayesian calculations. The conditional likelihood can be written as

$$
\begin{aligned}
p(R_{t+1}, P_{t+1}|X_{t+1}, \theta) &= p(R_{t+1}|X_{t+1})p(P_{t+1}|X_{t+1}, \theta) \\
p(P_{t+1}|X_{t+1}, \theta) &\sim N(F(X_t, \theta), \sigma_D^2)
\end{aligned}
$$

where σ_D is a pricing error of say $(1, 5)\%$ and $F(X_t, \theta)$ describes the pricing formula obtained from equilibrium arguments. In the discretized system, log-returns are given by $R_{t+1} = Y_{t+1} - Y_t = \ln(S_{t+1}/S_t)$ and $p(R_{t+1}|X_{t+1})$ describes the evolution of asset returns.

The goal of empirical asset pricing is to learn about the risk neutral and objective parameters $(\theta^{\mathrm{Q}}, \theta^{\mathrm{P}})$, respectively. Moreover, one can also recover filtered estimates of the state variables X_{t+1}, namely volatility, V_{t+1}, jump times, J_{t+1} and jump sizes, and the model specification from the observed asset returns and derivative prices. In general, we have a joint system with data $Y_{t+1} = (R_{t+1}, P_{t+1})$ corresponding to returns and derivative prices evolving according to the system

$$
\begin{pmatrix} P_{t+1} \\ R_{t+1} \end{pmatrix} = \begin{pmatrix} A(\theta^{\mathrm{P}}, \theta^{\mathrm{Q}})X_{t+1} + B(\theta^{\mathrm{P}}, \theta^{\mathrm{Q}}) \\ \mu^{\mathrm{P}} \end{pmatrix} + \begin{pmatrix} \sigma_D \epsilon_{t+1}^D \\ \sqrt{V_{t+1}} \epsilon_{t+1}^R \end{pmatrix}.
$$

Here σ_D is a pricing error and there exist pricing formulas for $A(\theta^{\mathrm{P}}, \theta^{\mathrm{Q}})$ and $B(\theta^{\mathrm{P}}, \theta^{\mathrm{Q}})$. See Chernov and Ghysels (2000) and Polson and Stroud (2003) for further discussion. Traditional pure inversion methods infer parameters by first taking the derivative price as given and then try and match state and parameter values to the given price, P_t, by inversion, namely $(\hat{X}_t, \hat{\theta}) = C^{-1}(P_t)$. However, taking the time series of implied states $\hat{X}_t$ can also lead to parameter estimates that are inconsistent.

Posterior Distribution. The Bayesian posterior distribution now uses both returns and derivative pricing information to estimate $(\theta^{\mathrm{P}}, X_t)$ and implicitly determine θ^{Q}. This is given by

$$p(\theta^{\mathrm{P}}, \theta^{\mathrm{Q}}, X_t|Y^t) \propto p(R^t|\theta^{\mathrm{P}}, X_t)p(P^t|\theta^{\mathrm{P}}, \theta^{\mathrm{Q}}, X_t)p(X_t|\theta^{\mathrm{P}})p(\theta^{\mathrm{P}}, \theta^{\mathrm{Q}})$$

where $Y_t = (R_t, P_t)$ contains returns and prices from the equilibrium model.

Hence we can sequentially filter $p(\theta^{\mathrm{P}}, \theta^{\mathrm{Q}}, X_{t+1}|Y^{t+1})$. For very small pricing errors the derivative price information will give very precise estimates of X_{t+1} which in turn will precisely estimate the parameters of the system. However, too restrictive will lead to noisy co-variance estimates and the derivative prices will be in conflict with the physical model evolution which is maybe a sign of model misspecification.

3 Bayesian Inference via SMC

Here we review the particle methods that are developed for state filtering and sequential parameter learning with and without derivative price information. Let X_t denote a latent state variable, θ underlying parameters and observed data Y_t at time t. This might just be returns R_t or a combination of returns and derivative prices (R_t, D_t). Then, the models considered in the paper are instances of the following general state-space model

$$(Y_t|X_{t-1}, \theta) \sim p(Y_t|X_{t-1}, \theta) \quad (1)$$

$$(X_t|X_{t-1}, \theta) \sim p(X_t|X_{t-1}, \theta) \quad (2)$$

for $t = 1, \ldots, T$ and $X_0 \sim p(X_0|\theta)$. The parameter θ is kept fixed (and omitted) for the moment. There are three filtering and learning posterior distributions:

1. *Filtering:* computation of or sampling from $p(X_t|Y^t)$ on-line for $t = 1, \ldots, T$;
2. *Filtering and learning:* computation of and/or sample from the posterior $p(\theta, X_t|Y^t)$ from which we can calculate marginals $p(X_t|Y^t)$ and $p(\theta|Y^t)$ for $t = 1, \ldots, T$; and
3. *Smoothing:* computation of or sampling from the full joint distribution $p(\theta, X^T|Y^T)$.

Predictive distributions are also straightforward to compute as forward functionals of the process taking into account this posterior uncertainty about parameters. The optimal Bayesian nonlinear filters (under squared error loss) are $\hat{X}_t = \mathbb{E}\left(X_t|Y^t\right)$ and $\hat{\theta} = \mathbb{E}\left(\theta|Y^t\right)$. Given these filtered posterior estimates of the states and the parameters we can then provide estimates of the following dynamics: i) *physical and risk-neutral:* $\mathbb{P}$ and $\mathbb{Q}$ dynamics; ii) *market price of risk:* given a specification of market prices of risk we can calculate the posterior $p(\lambda|Y^T)$ given a panel of derivative prices (calls, variance swaps).

Particle filters. Unfortunately, closed-form solution for the filtering and smoothing distributions is only available for simple cases. For the general state-space model, the filtered distributions are temporally connected via the following recursive propagation/update equations:

$$p(X_{t+1}|Y^t) = \int p(X_{t+1}|X_t)p(X_t|Y^t)dX_t \quad (3)$$

$$p(X_{t+1}|Y^{t+1}) \propto p(Y_{t+1})p(X_{t+1}|Y^t) \quad (4)$$

i.e. the propagation rule emulates the prior distribution of X_{t+1} (a high dimensional integral) to be combined with the likelihood $p(X_{t+1}|Y^{t+1})$ via Bayes' theorem (also a function of a high-dimensional integral).

We use particle filter algorithms to sequentially update the particle set $\{X_t^{(1)}, \ldots, X_t^{(N)}\}$ to the particle set $\{X_{t+1}^{(1)}, \ldots, X_{t+1}^{(N)}\}$ once Y_{t+1} become available. Particle filters provide a natural alternative to MCMC methods which are computationally intensive for the sequential inference problem (see Johannes and Polson, 2009). The posterior distribution $p(X_t|Y^t)$ for the filtering and learning distribution on states and parameters is approximated by

$$p^N(X_t|Y^t) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_t^{(i)}}(X_t)$$

based on the particle set $\{X_t^{(i)}, i = 1, \ldots, N\}$, where $\delta_z(\cdot)$ denotes the delta-Dirac mass located in z . We need to show how to efficiently update particles. More precisely, after observing Y_{t+1}, we need to efficiently produce a new particle set $\{X_{t+1}^{(i)}, i = 1, \ldots, N\}$ that approximates $p(X_{t+1}|Y^{t+1})$. We review the seminal sample-importance resample or *bootstrap filter* of Gordon et al. (1993) and the particle learning schemes of Lopes et al. (2010). A recent and thorough literature review of existing methods and recent advances in sequential Monte Carlo is provided by Cappé et al. (2007).

3.1 The Sample-Importance Resample (SIR) Filter

The classical bootstrap filter is also well known as the sequential importance sample with resampling (SISR) filter, and can be described in two steps mimicking the above propagation/update rule (equations (3) and (4)).

Let the particle set $\{X_t^{(i)}, i = 1, \ldots, N\}$ approximate $p(X_t|Y^t)$.

1. *Propagate:* Draw $\tilde{X}_{t+1}^{(i)} \sim p(X_{t+1}|X_t^{(i)})$ and compute associated (unnormalized) weights $w_{t+1}^{(i)} \propto p(Y_{t+1}|X_{t+1}^{(i)})$, for $i = 1, \ldots, N$;

2. *Resample:* $X_{t+1}^{(i)} = \tilde{X}_{t+1}^{(k^i)}$, where $k^i \sim \text{Multinomial}(w_{t+1}^{(1)}, \ldots, w_{t+1}^{(N)})$, for $i = 1, \ldots, N$.

Now, the particle set $\{X_{t+1}^{(i)}, i = 1, \ldots, N\}$ approximate $p(X_{t+1}|Y^{t+1})$. In words, the propagation step generates particles that approximate the prior distribution at time $t+1$, i.e. $p(X_{t+1}|Y^t)$ (equation (3)). Then, a simple SIR argument (reweighing prior draws with their likelihoods) transforms prior particles into posterior particles that approximate $p(X_{t+1}|Y^{t+1})$ (equation (4)).

Despite (or due to) its attractive simplicity and generality, the SISR algorithm suffers from *particle degeneracy* and is bounded to break down after a few hundred observations, even in the simplest scenarios, such as the local level model. Additionally, SIR filters tend to become even more unstable when sequential parameter learning is dealt with. We propose below a particle filter that overcome these obstacles in a large class of dynamic models.

3.2 Particle Learning

Bayes' rule links these to the next filtering distribution through Kalman-type updating. This takes the form of a smoothing and a prediction step that reverses the standard order of the propagation/update rule of equations (3) and (4). More precisely,

$$\begin{aligned} p(X_t|Y^{t+1},\theta) &\propto p(Y_{t+1}|X_t,\theta)p(X_t|Y^t,\theta) \\ p(X_{t+1}|Y^{t+1},\theta) &= \int p(X_{t+1}|X_t,\theta)p(X_t|Y^{t+1},\theta)dX_t \\ p(\theta|X^{t+1},Y^{t+1},\theta) &= p(\theta|Z_{t+1}) \end{aligned}$$

where $Z_{t+1} = \mathcal{Z}(Z_t, X_{t+1}, Y_{t+1})$ is a vector of conditional sufficient statistics for θ. This leads us to the following particle simulation algorithm. As before, let the particle set $\{(X_t, Z_t, \theta)^{(i)}, i = 1, \ldots, N\}$ approximate $p(X_t, \theta, Z_t|Y^t)$.

1. *Resample:* $(\tilde{X}_t, \tilde{Z}_t, \tilde{\theta})^{(i)} = (X_t, Z_t, \theta)^{(k^i)}$, where $k^i \sim$ Multinomial$(w_t^{(1)}, \ldots, w_t^{(N)})$ and (unnormalized) weights $w_t^{(j)} \propto p(Y_{t+1}|X_t^{(j)}, \theta^{(j)})$, for $j = 1, \ldots, N$;
2. *Propagate:* Draw $X_{t+1}^{(i)}$ from $p(X_{t+1}|\tilde{X}_t^{(i)}, \tilde{\theta}^{(i)}, Y_{t+1})$, for $i = 1, \ldots, N$;
3. *Update sufficient statistics:* $Z_{t+1}^{(i)} = \mathcal{Z}(\tilde{Z}_t^{(i)}, X_{t+1}^{(i)}, Y_{t+1})$;
4. *Parameter learning:* $\theta|Z_{t+1}^{(i)} \sim p(\theta|Z_{t+1}^{(i)})$.

Central to PL algorithms is the possibility of directly sampling from the joint posterior distribution of state (augmented or not) and parameter conditional sufficient statistics. There are a number of advantages to using particle learning: *i)* optimal filtering distributions in sequential parameter learning cases (Pitt and Shephard, 1999); *ii)* parameter sufficient statistics for Gaussian and conditionally Gaussian state-space models (Storvik, 2002); *iii)* straightforward particle smoothing (Godsill et al., 2004); *v)* state sufficient statistics reduces Monte Carlo error (Chen and Liu, 2000); and *iv)* alternative to standard MCMC methods in state-space models (Carvalho et al., 2009). Lopes et al. (2010) introduces PL as a framework for (sequential) posterior inference for a large class of dynamic and static models.

3.3 The 2007-2008 Credit Crisis: Extracting Volatility and Jumps

Lopes and Polson (2010) used particle filtering methods to estimate volatility and examine volatility dynamics for three financial time series (S&P500, NDX100 and XLF) during the early part of the credit crisis. Standard and Poor's SP500 stock index and the Nasdaq NDX 100 index are well known. The XLF index is an equity index for the prices for US financial firms. They compared pure stochastic volatility models to stochastic volatility models with jumps. More specifically, the stochastic volatility jump model includes the possibility of jumps to asset prices and one possible model is

$$\begin{aligned} \frac{dS_t}{S_t} &= \mu + \sqrt{V_t}dB_t^{\mathbb{P}} + d\left(\sum_{s=N_t}^{N_{t+1}} Z_s\right) \\ d\log V_t &= \kappa_v(\theta_v - \log V_t) + \sigma_v dB_t^V \end{aligned}$$

where the additional term in the equity price evolution describes the jump process and is absent in the pure stochastic volatility model. The parameter μ is an expected rate of return and the parameters governing the volatility evolution are κ_v, θ_v and σ_v. The Brownian motions $(B_t^{\mathbb{P}}, B_t^V)$ are possibly correlated giving rise to a leverage effect. The probabilistic evolution $\mathbb{P}$ describes what is known as the physical dynamics as opposed to the risk-neutral dynamics $\mathbb{Q}$ which is used for pricing. Sequential model choice shows how the evidence in support of the stochastic volatility jump model accumulates over time as market turbulence increases (Figure 1).

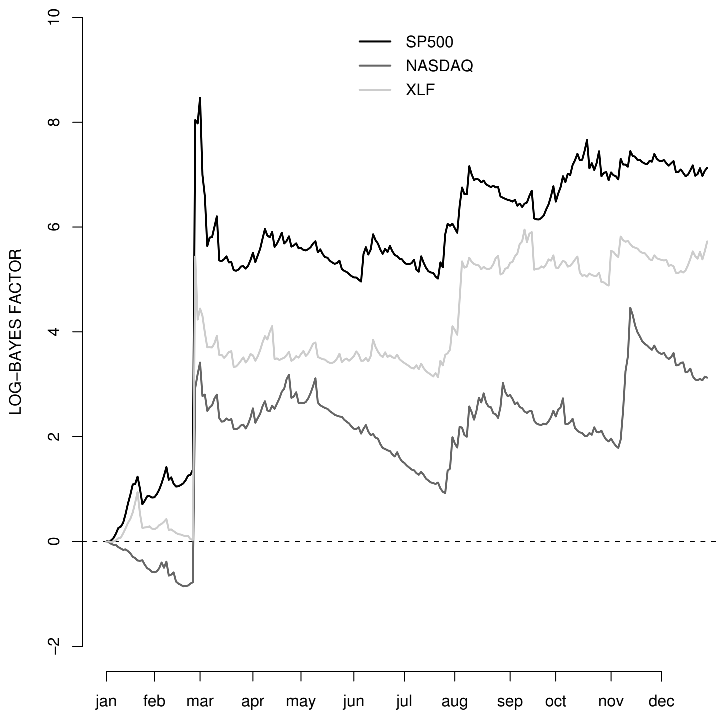

Figure 1: Sequential (log) Bayes factor for the SV with jumps model versus the pure SV model for the year 2007.

4 Bayesian Inference via MCMC

In this section we outline a simple and robust approach to the estimation of non-linear state-space models derived from discretized asset pricing models. We illustrate the approach on asset pricing derived from the equilibrium conditions of a fully specified economic model. Our approach is to discretize thes states and then apply the FFBS sampling algorithm (forward-filtering, backward-sampling, Frühwirth-Schnatter, 1994, Carter and Kohn, 1994) as an alternative to the general MCMC scheme of Carlin, Polson and Stoffer (1992).

Our exposition and examples will assume a one-dimensional state space model. In higher dimensional cases, we can update the components of the state one at a time in a Gibbs sampler. The parameter θ is kept fixed during the FFBS step and is usually sampled jointly. Hore et al. (2009), for instance, used a mixture of various Metropolis-Hastings chains to sample from $p(\theta|X^T, Y^T)$.

The rest of this section is largely based on Hore et al. (2009). They derive option prices in a general equilibrium setting under recursive preferences and time-varying growth rates. The key distinction from the option prices from the previous section is that the market prices of risks are determined endogenously from the solution of the agent's utility maximization problem under recursive preferences and stochastic growth rates. This highlights the strength of our estimation procedure. We can apply our Bayesian methodology to estimate deep parameters of a dynamic general equilibrium model and obtain full inference on the underlying states (and parameters as well) given the data on prices (or return) and/or quantity dynamics that are implied by the economic system. It takes a valuable step forward in giving us inference on economic parameters that guide us to understand dynamic rational expectations models that form the building block of structural asset pricing problems.

4.1 Forward Filtering, Backward Sampling

A general, nonlinear state-space model is specified by the distributions

$$p(X_0), \;\; p(X_t \mid X_{t-1}), \;\; p(Y_t \mid X_t, X_{t-1}).$$

The first two distributions determine the distribution of the latent states $\{X_t\}$ and the third distribution gives the distribution of the the observable $\{Y_t\}$ given the current and previous states. Often the observation equation is given as $p(Y_t \mid X_t)$ so that the current observation only depends on the current state. We will need the generalization given above for our financial application.

Given the three distributions above, the joint distribution of (X_0, X^T, Y^T), where $Z^t = (Z_1, \ldots, Z_t)$, is given by

$$p(X_0, X^T, Y^T) = p(X_0) \prod_{t=1}^{T} p(X_t \mid X_{t-1}) \prod_{t=1}^{T} p(Y_t \mid X_t, X_{t-1}). \tag{5}$$

We now review the basic steps involved in FFBS. After we have finished our review we will discuss the attractive simplicity of the state discretization strategy. FFBS consists of first forward filtering (FF) and then backward sampling (BS). The forward filtering step recursively updates

$$p(X_{t-1}, X_t \mid Y^t) \quad \Rightarrow \quad p(X_t, X_{t+1} \mid Y^{t+1}).$$

The backward sampling draws from the joint distribution of the states given the data using

$$p(X_1, X_2, \ldots, X_T \mid Y^T) = p(X_T, X_{T-1} \mid Y^T) \prod_{t=T-2}^{1} p(X_t \mid X_{t+1}, X_{t+2}, \ldots, X_T, Y^T).$$

That is, we first draw the last two states given all the data and then work backwards in time drawing each state conditional on all the subsequent ones.

4.1.1 Forward Filtering

We do the forward filtering (FF) iteration in two steps: an evolution step and an update step.

FF Evolution. In the first step, we extend our state knowledge at time t to include the future state using the state equation:

$$\begin{aligned} p(X_{t-1}, X_t, X_{t+1} \mid Y^t) &= p(X_{t-1}, X_t \mid Y^t)\, p(X_{t+1} \mid X_t, X_{t-1}, Y^t) \\ &= p(X_{t-1}, X_t \mid Y^t)\, p(X_{t+1} \mid X_t). \end{aligned}$$

The first term on the right hand side is what we assume we know from the previous iteration, i.e. the time t posterior distribution of (X_{t-1}, X_t), while the second term is the state equation of the state-space model. We then margin out X_{t-1} to obtain $p(X_t, X_{t+1} \mid Y^t)$:

$$p(X_{t-1}, X_t, X_{t+1} \mid Y^t) \Rightarrow p(X_t, X_{t+1} \mid Y^t).$$

Alternatively, we can first margin out X_{t-1} from $p(X_{t-1}, X_t \mid Y^t)$ and then use the state equation:

$$p(X_t, X_{t+1} \mid Y^t) = p(X_t \mid Y^t)\; p(X_{t+1} \mid X_t).$$

This first step may be viewed as computing our prior knowledge of (X_t, X_{t+1}) given Y^t, our observations up to time t.

FF Update. In the second step, we implement the Bayes' theorem to update our distribution of (X_t, X_{t+1}) to incorporate the additional information in Y_{t+1}. Keeping in mind that $Y^{t+1} = (Y^t, Y_{t+1})$,

$$\begin{aligned} p(X_{t+1}, X_t \mid Y^{t+1}) &\propto p(X_{t+1}, X_t, Y_{t+1} \mid Y^t) \\ &= p(X_t, X_{t+1} \mid Y^t)\; p(Y_{t+1} \mid X_t, X_{t+1}, Y^t) \\ &= p(X_t, X_{t+1} \mid Y^t)\; p(Y_{t+1} \mid X_t, X_{t+1}). \end{aligned}$$

The first term on the right hand side is available from the above FF evolution step, while the second term is the observation equation of the state-space model. The FF step is really just Bayes theorem repeated over time.

4.1.2 Backward Sampling

The backward sampling step depends on the observation that

$$\begin{aligned} p(X_t \,|\, X_{t+1}, X_{t+2}, \ldots, X_T, Y^T) &= p(X_t \mid Y^t, X_{t+1}, Y_{t+1}) \\ &= p(X_t \mid X_{t+1}, Y^{t+1}). \end{aligned} \tag{6}$$

Let $\tilde{Z}^t = (Z_t, Z_{t+1}, \ldots, Z_T)$. So superscript t means everything up to and including t and superscript combined with a $\sim$ on top means from t on including t. Using this notation equation (6) becomes

$$p(X_t \mid X_{t+1}, \tilde{X}^{t+2}, Y^t, Y_{t+1}, \tilde{Y}^{t+2}) = p(X_t \mid X_{t+1}, Y^t, Y_{t+1}).$$

Stated in terms of conditional independence, this is equivalent to say that

$$X_t \perp (\tilde{X}^{t+2}, \tilde{Y}^{t+2}) \mid X_{t+1}, Y^t, Y_{t+1},$$

where $\perp$ indicates independence. If the state space model is written as a DAG (directed acyclic graph) and then converted to an undirected graph, this conditional independence statement is obvious. We can also see the property given by equation (6) directly from the joint distribution of (X^T, Y^T) by first noting that

$$p(X^{t-1}, X_t, Y^t \mid X_{t+1}, Y_{t+1}, \tilde{X}^{t+2}, \tilde{Y}^{t+2}) = p(X^{t-1}, X_t, Y^t \mid X_{t+1}, Y_{t+1}). \tag{7}$$

Since the seven quantities in both sides of the above comprise all of (X^T, Y^T), this relation may be easily seen simply by looking at the full joint given in equation (5) and remembering that the conditional is proportional to the joint. That is, we can rewrite equation (5) as

$$\begin{aligned} p(X_0, X^T, Y^T) &= p(X_0)\, p(X^{t-1} \mid X_0)\, p(X_t \mid X_{t-1})\, p(X_{t+1} \mid X_t)\, p(\tilde{X}^{t+2} \mid X_{t+1}) \\ &\times\; p(Y^t \mid X^t)\, p(Y_{t+1} \mid X_{t+1}, X_t)\, p(\tilde{Y}^{t+2} \mid X_{t+1}, \tilde{X}^{t+2}) \end{aligned}$$

and then equation (7) becomes apparent. Equation (6) then is obtained by first further conditioning on Y^t and the margining out X^{t-1}.

4.1.3 Discretized FFBS

Now that the essentials of FFBS have been laid out, we can review the necessary computations and see that they are easily performed given a state discretization. Let each X takes on values in the grid $(x_1, x_2, \ldots, x_M)$. Then $p(X_{t-1}, X_t \mid Y^t)$ may be represented by an $M \times M$ matrix with rows corresponding to X_{t-1} and columns corresponding to X_t. We then obtain the marginal for X_t by summing the rows. We then create the joint distribution of (X_t, X_{t+1}) by constructing the matrix whose (i, j) element is $p(X_t = x_i \mid Y^t)\, p(X_{t+1} = x_j \mid X_t = x_i)$. To update to conditioning on Y^{t+1}, we multiply each element of this matrix by $p(Y_{t+1} \mid X_t = x_i, X_{t+1} = x_j)$ and then renormalize so that the sum over all elements of the matrix is equal to one. For the BS step we can easily compute $p(X_t = x_i \mid X_{t+1} = x_j, Y^{t+1})$ from the matrices representing the joints $p(X_t, X_{t+1} \mid Y^{t+1})$ which we must store while doing the FF step.

4.2 Equilibrium Put Option Pricing

In this section we sketch for the reader the continuous time general equilibrium model in Hore et al. (2009). The model simultaneously determines consumption dynamics and option prices given preferences and capital accumulation dynamics. The above application of FFBS is used to estimate the discretized version of the model.

At a very high level, the full state space dynamics of equilibrium quantities and the underlying latent variable is given by

$$
\begin{aligned}
\frac{dC_t}{C_t} &= \mu_C(X_t)\, dt + a_{11}(X_t)\, dB_k + a_{12}(X_t)\, dB_x && (8)\\
p_{ti} &= f(X_t, S_i, \tau_{ti}, R_t) && (9)\\
dX_t &= \delta(\bar{X} - X_t)\, dt + \sigma_x dB_x && (10)
\end{aligned}
$$

where the Brownian motion terms B_k and B_x are correlated. There is a single state variable X_t (here t denotes continuous time) which is the expected return on production technology. C_t denotes the path of consumption in the economy under equilibrium and p_{ti} be the option prices consistent with equilibrium consumption dynamics. The i^{th} option has strike price S_i, time till expiration τ_{it}, and R_t is the equilibrium wealth upon which the option is written.

We orthogonalize the Brownian motion terms and express the shocks driven by independent Brownian motions. We can then write the model as

$$
\begin{aligned}
\frac{dC_t}{C_t} &= \mu_C(X_t)\, dt + a_{11}(X_t)\, dB_k + a_{12}(X_t)\, dB_u\\
p_{ti} &= f(X_t, S_i, \tau_{ti}, R_t) \qquad i = 1, \ldots, I\\
dX_t &= \delta(\bar{X} - X_t)\, dt + a_{21}(X_t)\, dB_k + a_{22}(X_t)\, dB_u
\end{aligned}
$$

where B_k and B_u are independent Brownian motions and p_{it} is the price of the i^{th} put option having strike price S_i and expiration time τ_{it}, for $i = 1, \ldots, I$, while R_t is the price of the underlying asset upon which the option is written. The parameters of the model are δ and $\bar{X}$. The functions μ_C, a_{ij}, and f are complex. They also depend on other model parameters, but this dependence is suppressed in order to highlight the state-space nature of the model. If we let θ denote these suppressed parameters then θ includes, for example, utility parameters that capture the risk aversion, elasticity of inter-temporal substitution, and time discount factor. The functions that describe the consumption process are derived by maximizing the expected utility of the consumption time path subject to constraints driven by the evolution of the expected return on production. The option is written on equilibrium wealth R_t determined in equilibrium by capital growth and current consumption level.

To form a discrete time version of the model over time interval Δt, we discretize the consumption and state equations and then add independent error to the option pricing equations. Let g_{t+1} be the consumption growth from t to $t+1$. Our discretized model is

$$\begin{aligned} g_{t+1} &= \mu_g(X_t) + a_{11}(X_t)\sqrt{\Delta t}\, Z_{t1} + a_{12}(X_t)\sqrt{\Delta t}\, Z_{t2} \\ p_{ti} &= f(X_t, S_i, \tau_{ti}, R_t) + \sigma\, \epsilon_{ti} \qquad i = 1, \ldots, I \\ X_{t+1} &= \alpha + \rho X_t + a_{21}(X_t)\sqrt{\Delta t}\, Z_{t1} + a_{22}(X_t)\sqrt{\Delta t}\, Z_{t2} \end{aligned}$$

where $\mu_g(X_t) = \mu_C(X_t)\Delta_t$, $\rho = 1 - \delta\bar{X}$, $\alpha = (1-\rho)\Delta_t$ and all Z_{ti} and ϵ_{ti} independent and identically distributed standard normal shocks. To put this model in our general form we let $W_t = a_{11}(X_t)\sqrt{\Delta t}\, Z_{t1} + a_{12}(X_t)\sqrt{\Delta t}\, Z_{t2}$ and $V_t = a_{21}(X_t)\sqrt{\Delta t}\, Z_{t1} + a_{22}(X_t)\sqrt{\Delta t}\, Z_{t2}$. We then draw V_t from its marginal distribution and W_t from its conditional distribution given V_t, or equivalently, its conditional distribution given X_{t+1} and X_t. This gives rise to the following nonlinear state space model

$$\begin{aligned} g_{t+1} &= \mu_g(X_t) + W_t \\ p_{ti} &= f(X_t, S_i, \tau_{ti}, R_t) + \sigma\, \epsilon_{ti} \qquad i = 1, \ldots, I \\ X_{t+1} &= \alpha + \rho X_t + V_t. \end{aligned}$$

where $W_t \sim p(W_t|X_{t+1}, X_t)$. The state equation could not be simpler, it is just an AR(1). The observation equations relate the observed relative put prices and consumption growth to the current and previous state as in our general prescription.

Inference for this complex model is now conceptually straightforward. At the top level, we have Gibbs sampler that alternates between $p(\theta|X^T)$ and $p(X^T|\theta)$. Drawing the states given θ using the discretized FFBS is quite simple. The only drawback is that since the function f is very expensive to compute it is necessary to precompute all possible $p(y_t|X_t, X_{t-1})$ where y_t varies over all the observed values and both X_{t-1} and X_t vary over the grid values. The more difficult step is to draw from $p(\theta|X^T)$. Hore et al. (2009), for instance, used a mixture of various Metropolis-Hastings chains.

Figure 2 shows the option data and state inference. Time is discretized to be months. Data on four different options corresponding to different strikes were used. The top panel plot the time series of relative option prices. The four options differ in their strike prices The other two panels show the inference for the states X_t with t now denoting the month. The marginal distribution of each state is indicated by the vertical dashed line. Such inference is straightforwardly obtained from our FFBS draws. The difference between the two plots is the prior on the smoothness of the state evolution.

Inference of complex functions of our states and parameters are easily obtained given the MCMC draws. A quantity of economic interest is the risk-premia. The risk-premia in Figure 3 is clearly countercyclical. Higher growth rates imply lower risk-premia and vice- versa. In bad times, the high precautionary savings motivation is consistent with high risk-premia. The agent feels more risk-averse at a time when the probability of a wealth loss is high. Clearly, the agent's demand for insurance is high and he is willing to pay more for put options in these states. Likewise, in good states the precautionary savings motive dissipates and the agents risk-premia is low. In these states, the agent is not fearful of a wealth loss and his willingness to pay for put options to insure his wealth is low. This explains the counter-cyclical risk-premia pattern that we filter out of the time-series of put option prices. For further discussion of modle specification and derivatives pricing and returns see Broadie et al. (2007, 2009).

Note that the difficulty in drawing θ also makes sequential estimation via particle filtering a practically infeasible alternative. While particle filtering might work for the draw of $X^T|\theta$ sequential, i.e. a pure

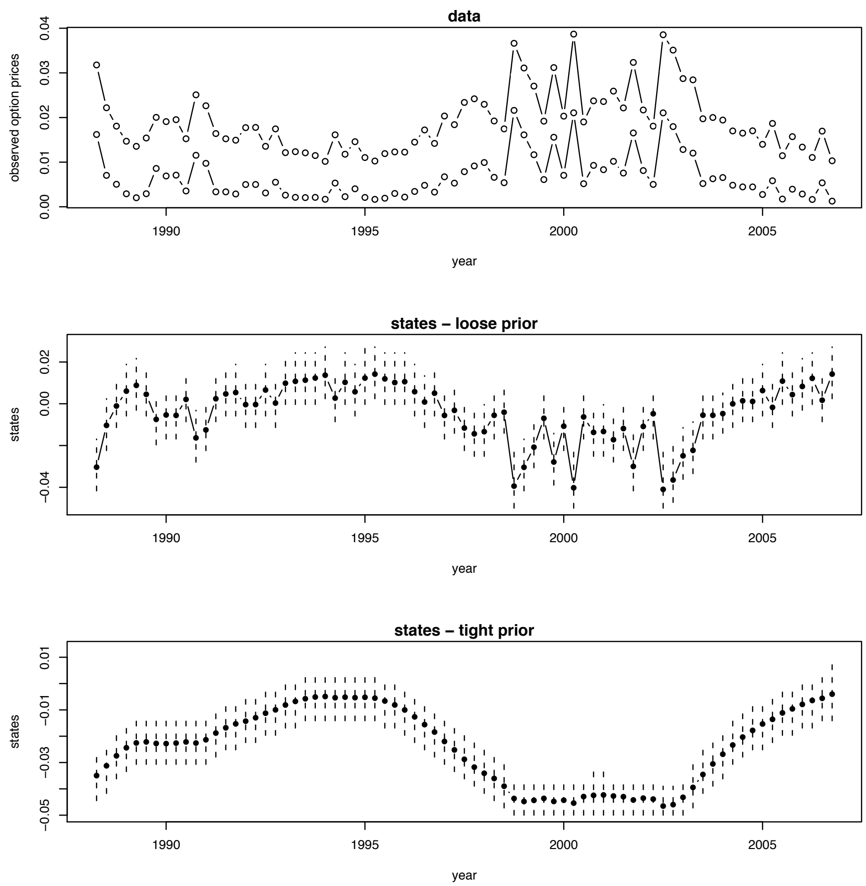

Figure 2: The top panel shows the time-series plot of the four option prices. The next two panels show the posterior distribution of the time-series of the underlying states corresponding to a loose and tight prior setting. The dashed line represents the 95% posterior band around each state.

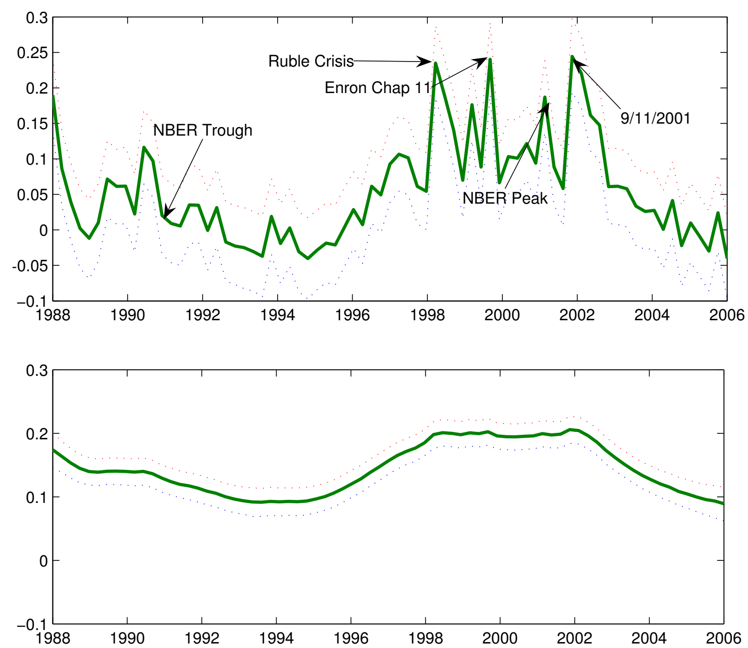

Figure 3: Time-Series of long-run risk estimated from option prices. Using the posterior distribution of parameters and the state, we compute the time-series of long-run risk. The median time-series estimate is presented in the above plot with the dotted line representing the 95% posterior interval at each point. The top time-series plot corresponds to loose prior, whereas the bottom time-series plot corresponds to the tight prior.

state learning case, it is not a straightforward matter to get joint inference for X^T and θ, i.e. a state and parameter learning case. Liu and West (2001), for instance, who sequentially approximates $p(\lambda|y^t)$ by a multivariate normal density, is an unreasonable alternative here given the highly nonlinear nature of θ. Similarly, it would be hard to implement particle filter and particle learning algorithms (see Section 3), since the posterior distribution of X^T and θ can not be represented by a small dimensional set of conditional sufficient statistics. The discretized FFBS scheme seems to be the best available alternative.

5 Conclusion

Bayesian approaches are natural in the analysis of financial models. It has long been recognized that Bayesian thinking is relevant to fundamental questions about risk and uncertainty. Many modern financial moderns have a sequential structure expressed in terms of dynamics of latent variables. In these models, the basic Bayesian advantage in coherent assessment of uncertainty is coupled with powerful computational methods.

In this paper we have reviewed two approaches to the Bayesian analysis of sequential models and attempted to illustrate ways to apply them to models derived from financial and economic theory. Particle filtering methods allow us to quickly and dynamically update our inferences about state and, in some cases, parameters. The suggested use of FFBS is much slower and more suited when complete joint inference is needed for a states over a fixed time period and underlying model parameters. The advantage of this approach is its simplicity and wide applicability without asking the user to make difficult choices about algorithm details.

6 References

Berger, James, 1985, *Statistical Decision Theory and Bayesian Analysis*, New York: SpringerVerlag.

Bernardo, José and Adrian Smith, 1994, *Bayesian Theory*, New York: Wiley.

Broadie, Mark, Mikhail Chernov and Michael Johannes, 2007, Model Specification and Risk premia: Evidence from SP500 futures options. *Journal of Finance* (to appear).

Broadie, Mark, Mikhail Chernov and Michael Johannes, 2009, Understanding index option returns. *Review of Financial Studies* (to appear).

Cappé, Olivier, Simon Godsill and Eric Moulines, 2007, An Overview of Existing Methods and Recent Advances in Sequential Monte Carlo, *Proceedings of the IEEE*, 95, 899-924.

Carlin, Bradley, and Nicholas Polson, and David Stoffer, 1992, A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling, *Journal of the American Statistical Association*, 87, 493-500.

Carter, Chris and Robert Kohn, 1994, On Gibbs sampling for state space models, *Biometrika*, 81, 541-553.

Carvalho, Carlos, Michael Johannes, Hedibert Lopes and Nicholas Polson, 2009, Particle learning and smoothing, Discussion paper, The University of Chicago Booth School of Business.

Chen, Rong and Liu, Jun (2000). Mixture Kalman filters. *Journal of Royal Statistical Society, Series B*, 62, 493-508.

Chernov, Mikhail and Eric Ghysels, 2000, A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. *Journal of Financial Economics*, 56, 407-458.

Duffie, Darrell, 1996, State-Space Models of the Term Structure of Interest Rates, in H.Körezlioglu, B. Øksendal, and A. Üstünel, editors, *Stochastic Analysis and Related Topics V*: The Silivri Workshop, Boston: Birkhauser.

Frühwirth-Schnatter, Sylvia, 1994, Data augmentation and dynamic linear models, *Journal of Time Series Analysis*, 15, 183-202.

Gamerman, Dani and Hedibert Lopes, 2007, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference (2nd Edition)*, Baton Rouge: Chapman & Hall/CRC.

Godsill, Simon, Arnaud Doucet and Mike West, 2004, Monte Carlo smoothing for nonlinear time series, *Journal of the American Statistical Association*, 99, 156-168.

Gordon, Neil, D. Salmond and Adrian Smith, 1993, Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEE Proceedings*, F-140, 107–113.

Heston, Steven, 1993, A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, *Review of Financial Studies*, 6, 327-343.

Hore, Satadru, Hedibert Lopes, and Robert McCulloch, 2009, Put option implied risk-premia in general equilibrium under recursive preferences, Discussion paper, The University of Chicago Booth School of Business.

Jacquier, Eric, Nicholas Polson, and Peter Rossi, 1994, Bayesian analysis of Stochastic Volatility Models, (with discussion), *Journal of Business and Economic Statistics*, 12, 371-471.

Johannes, Michael, Nicholas Polson and Jonathan Stroud, 2009, Optimal Filtering of Jump Diffusions: Extracting Latent States from Asset Prices, *Review of Financial Studies*, 22, 2759-2799.

Johannes, Michael and Nicholas Polson, 2009, MCMC methods for Financial Econometrics, in Y. Ait-Sahalia and L. Hansen, editors, *Handbook of Financial Econometrics, Volume 2*, Oxford: Elsevier, 1-72.

Liu, Jane and Mike West, 2001, Combined parameters and state estimation in simulation-based filtering, in A. Doucet, N. de Freitas and N. Gordon, editors, *Sequential Monte Carlo Methods in Practice*, New York: Springer-Verlag.

Lopes, Hedibert, Carlos Carvalho, Michael Johannes and Nicholas Polson, 2010, Particle Learning for sequential Bayesian computation, in J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West, editors, *Bayesian Statistics 9*, Oxford: Oxford University Press (to appear).

Lopes, Hedibert and Nicholas Polson, 2010, Extracting SP500 and NASDAQ volatility: The credit crisis of 2007-2008, in A. O'Hagan and M. West, editors, *Handbook of Applied Bayesian Analysis*, Oxford: Oxford University Press (to appear).

Pitt, Michael and Neil Shephard, 1999, Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94, 590-599.

Polson, Nicholas and John Stroud, 2003, Bayesian Inference for Derivative Prices, in J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West, editors, *Bayesian Statistics 7*, Oxford: Oxford University Press, 641-650.

Storvik, Geir (2002) Particle filters in state space models with the presence of unknown static parameters. *IEEE Transactions of Signal Processing*, 50, 281-289.