Inequality Constrained State Space Models

Qian, Hang

3 September 2015

Inequality Constrained State Space Models

Hang Qian

September 2015

Summary: The standard Kalman filter cannot handle inequality constraints imposed on the state variables, as state truncation induces a non-linear and non-Gaussian model. We propose a Rao-Blackwellised particle filter with the optimal importance function for forward filtering and the likelihood function evaluation. The particle filter effectively enforces the state constraints when the Kalman filter violates them. We find substantial Monte Carlo variance reduction by using the optimal importance function and Rao-Blackwellisation, in which the Gaussian linear sub-structure is exploited at both the cross-sectional and temporal levels.

Keywords: Rao-Blackwellisation, Kalman filter, Particle filter, Sequential Monte Carlo
JEL Classification: C32, C53

E-mail: matlabist@gmail.com

1. Introduction

For economic applications of the state space models, the state variables often represent latent economic processes, some of which are inherently nonnegative or bounded. A leading example is the Gaussian short rate models such as Vasicek (1977) and Hull and White (1990). The conventional wisdom is that the nominal interest rate should be nonnegative (Black, 1995).[1] In an era of low interest rates, the standard Kalman filter results are likely to violate the constraints.

Figure 1 demonstrates the binding inequality constraints in the Vasicek model, in which the instantaneous interest rate is the state variable and the entire term structure is a linear function of the state. Refer to Hull (2003, p. 539) for the model specification. The estimation data are monthly U.S. treasury rates of maturities from three months to ten years, 2003 - 2015. Since there are no negative observations in our sample, it is desirable to have nonnegative short rates as well. However, the upper panel of Figure 1 shows that the Kalman filter produces negative state estimation after the year 2009. The middle panel plots five posterior draws of the state series by the standard simulation smoothing algorithm (Durbin and Koopman, 2002). All of them contains negative values. We generated millions of posterior series, but could not obtain one that satisfies the nonnegative constraints.

State space applications subject to inequality constraints are common. In the local level model using the Nile river data (see Commandeur et al., 2011), the river flow volume is necessarily a nonnegative state variable. In Stock and Watson (2007), the latent inflation rate is a bounded sequence if the central bank sets inflation targets. In the Diebold et al. (2006) yield curve applications, the sign of the level, slope and curvature factors might be known if there is prior information on the shape of the yield curve. In the time-varying parameter (TVP) vector autoregressions (VAR) (see Cogley and Sargent, 2001), it is advisable to restrict the eigenvalues of the VAR process within the unit circle.

State constraints can be in the form of equalities and inequalities. Equality constraints are equivalent to perfect measurement equations. Doran (1992) shows that equality constraints

[1] Recent observations on the negative deposit rate set by the European Central Bank were deemed as a new lower bound by some market participants. In addition, cash storage costs could set a natural lower bound for the negative rate.

can be incorporated in the state estimation by augmentation of measurement equations. Pizzinga (2012) provides a proof of the equality restricted Kalman filtering based on the Hilbert space geometry and demonstrates statistical efficiency of restricted filtering. Koop et al. (2010) consider a Bayesian application in which the states are subject to time-varying equality constraints.

Imposing inequality constraints on simulation smoothing has received attention in the literature. Cogley and Sargent (2005) simulate the unrestricted posterior draws and rule out outcomes that violate the constraints by rejection sampling. This multi-move algorithm is valid, but the acceptance rate of the rejection sampling could be low. In our simulation of the Vasicek model, it is difficult to obtain a nonnegative path. Koop and Potter (2011) develop a single-move algorithm, which works well in their application. The single-move algorithm might produce highly correlated draws, as demonstrated in Carter and Kohn (1994). The prior state distribution in Koop and Potter (2011) differs from Cogley and Sargent (2005) due to a prior integrating constant, which yields an analytically intractable posterior distribution and they resort to the Metropolis-Hasting sampler.

Imposing inequality constraints on forward filtering receives little attention. To the best of our knowledge, there are no rigorous approaches addressing constrained filtering in the economics and statistics literature. However, there are reasonable ways to add constraints to the Kalman filter. In engineering applications, Simon and Simon (2005) and Gupta and Hauser (2008) adapt the Kalman filter by treating an active set of inequality constraints as equality constraints. Simon and Simon (2010) truncates univariate normal densities for an adapted Kalman filter.

This paper provides a rigorous treatment of the inequality constrained state filtering and the likelihood function evaluation. Our main contribution is a Rao-Blackwellised particle filter with the optimal importance function, which effectively enforces the inequality constraints when the Kalman filter violates them. Our algorithm departs from the Kalman filter, but analytic integration by the Kalman filter is utilized by Rao-Blackwellisation at both cross-sectional and temporal levels. Our algorithm is based on the particle filter, but not as computationally intensive, since marginalization reduces the state dimensions for particle filtering, and muffles

Monte Carlo noises. Variance reduction is also significant in the likelihood function evaluation, which facilitates numerical search for the maximum likelihood estimator. Our algorithm is rigorous because the standard convergence results for sequential Monte Carlo methods apply. Our sampling method will restore the true constrained filtering distribution asymptotically and the estimated likelihood function will converge to the true likelihood value.

The reminder of the paper is organized as follows. Section 2 specifies the transition and observation distributions of the inequality constrained model, based on which a particle filter is proposed in Section 3. Section 4 and 5 discuss the cross-sectional and temporal Rao-Blackwellisation. An application in Section 6 demonstrates the effects of constraint enforcement and variance reduction by the optimal importance function and Rao-Blackwellisation. Sections 7 extends the model by an alternative type of state constraints, which is computationally faster and numerically stable. In Section 8, we suggest a practical workflow of parsimoniously imposing inequality constraints.

2. The Model

Let $x_t, t = 1, \dots, T$ be a $m \times 1$ state vector, and y_t be a $n \times 1$ observation vector. We define a probabilistic model by the joint density $p(x_{1:T}, y_{1:T})$, where $x_{1:T} = (x_1', \dots, x_T')'$ and $y_{1:T} = (y_1', \dots, y_T')'$. The joint density, decomposed as $\prod_{t=1}^{T} p(x_t|x_{1:t-1}, y_{1:t-1})p(y_t|x_{1:t}, y_{1:t-1})$, is said to be an inequality constrained state space model (ICSSM) if

$$p(x_t|x_{1:t-1}, y_{1:t-1}) = \frac{\phi(x_t; A_t x_{t-1}, Q_t)}{F(A_t x_{t-1}, Q_t, \mathcal{X}_t)} \cdot 1(x_t \in \mathcal{X}_t), \tag{1}$$

$$p(y_t|x_{1:t}, y_{1:t-1}) = \phi(y_t; C_t x_t, R_t), \tag{2}$$

where the matrices A_t, C_t, Q_t, R_t are time-varying coefficients, which could be functions of past observations $y_{1:t-1}$ in economic applications (e.g., autoregressive terms in C_t). The set $\mathcal{X}_t \subset \mathbb{R}^m$ represents the state constraints and the function $1(x_t \in \mathcal{X}_t)$ is a binary indicator for the event $\{x_t | x_t \in \mathcal{X}_t\}$. Also, the density $\phi(x_t; A_t x_{t-1}, Q_t)$ denotes the multivariate normal $N(A_t x_{t-1}, Q_t)$ density evaluated at x_t, and the normalisation term $F(A_t x_{t-1}, Q_t, \mathcal{X}_t)$ denotes the probability of $N(A_t x_{t-1}, Q_t)$ in the region $\mathcal{X}_t$. Note that the normalisation term is a function of the past state x_{t-1}, hence a non-linear model. We assume that $F(\cdot) > 0$, for we aim at inequality constraints. Equality constraints can be cast as perfect measurement equations and put in Eq (2) instead. As an example of inequality constraints, nonnegative states are represented by $\mathcal{X}_t = \{x_t | x_t \geq 0\}$ with $F(\cdot)$ as the upper cumulative distribution function (c.d.f.). Inequality constraints can be a non-linear function of the states, say $\mathcal{X}_t =$ $\left\{(x_{1t}, x_{2t}, x_{3t}, x_{4t}) \middle| \text{eigenvalues for} \begin{pmatrix} x_{1t} & x_{2t} \\ x_{3t} & x_{4t} \end{pmatrix} \text{in unit circle}\right\}$.

ICSSM is conformable to the state space paradigm. First, Markovian transition: $p(x_t | x_{1:t-1}, y_{1:t-1}) = p(x_t | x_{t-1}, y_{1:t-1})$, which is a truncated normal distribution denoted by $TN(A_t x_{t-1}, Q_t, \mathcal{X}_t)$. Second, contemporaneous observations: $p(y_t | x_{1:t}, y_{1:t-1}) =$ $p(y_t | x_t, y_{1:t-1})$. If $\mathcal{X}_t = \mathbb{R}^m$, then Eq (1) and (2) reduces to a Gaussian linear state space form:

$$x_t = A_t x_{t-1} + \varepsilon_t, \tag{3}$$

$$y_t = C_t x_t + v_t, \tag{4}$$

where $\varepsilon_t \sim N(0, Q_t)$, $v_t \sim N(0, R_t)$.

We assume that the initial state vector x_0 is deterministic, without loss of generality because the time-varying coefficient matrices can replicate a non-deterministic initial state distribution. For example, suppose that we require $x_0 \sim TN(\mu_0, \Sigma_0, \mathcal{X}_0)$. Then we may put $x_{-1} = \mu_0$ with $A_0 = I$, $Q_0 = \Sigma_0$, $C_0 = 0$, $R_0 = 0$, $y_0 = 0$. Forward-shifting the time script for all variables in the model by one period (i.e., rewrite x_{-1} as x_0, A_0 as A_1, etc.), we obtain an equivalent state space model with deterministic initial states.

The posterior state distribution takes the form

$$p(x_{1:t} | y_{1:t}) \propto \prod_{\tau=1}^{t} \left[\frac{\phi(x_\tau; A_\tau x_{\tau-1}, Q_\tau) \phi(y_\tau; C_\tau x_\tau, R_\tau)}{F(A_\tau x_{\tau-1}, Q_\tau, \mathcal{X}_\tau)} \cdot 1(x_\tau \in \mathcal{X}_\tau) \right].$$

Due to the normalisation term $F(A_\tau x_{\tau-1}, Q_\tau, \mathcal{X}_t)$ in the denominator, the posterior state distribution does not have a closed form for $t > 1$. However, the single-period filtering distribution conditional on the past states, namely $p(x_t | x_{t-1}, y_{1:t})$, has an analytic form, which remains to be a truncated normal distribution since the normalisation term can be treated as a constant. Also, the unknown proportionality constant is of interest. Integrating the right hand side with respect to $x_{1:t}$ yields the likelihood function $p(y_{1:t})$, which is crucial for maximum likelihood estimation of the unknown parameters in the coefficient matrices A_t, C_t, Q_t, R_t.

3. The Particle Filter

Introduced by Gordon et al. (1993), the particle filter is a powerful tool for characterizing a series of target distributions of increasing dimensions: $p(x_{1:t}|y_{1:t}), t = 1, \dots, T$. The target density is proportional to $p(x_{1:t}, y_{1:t})$, which can be evaluated pointwise. The proportionality constant is the likelihood function $p(y_{1:t})$.

Particle filtering is developed in the importance sampling framework. Particles are generated from a well-chosen proposal density $f_t(x_{1:t})$, and assigned the unnormalised importance weights $w_t(x_{1:t}) = \frac{p(x_{1:t}, y_{1:t})}{f_t(x_{1:t})}$. The weighted particles represent a categorical distribution that approximates the target distribution, as the empirical c.d.f. uniformly converges to the target c.d.f..[2] In addition, the sample average of the unnormalised weights approximates the likelihood function, as the average weight is an unbiased and strongly consistent estimator for the likelihood value. Refer to Liu and Chen (1998), Chopin (2004), Doucet and Johansen (2009), and others.

In the sequential importance sampling, the proposal density is formulated recursively such that $f_t(x_{1:t}) = \ f_{t-1}(x_{1:t-1}) \cdot g(x_t|x_{1:t-1})$, where $g(x_t|x_{1:t-1})$ is a well-chosen transition kernel. We wish that the proposal density is close to the target density and the particle importance weights have a small variance.[3] For state space models, if we choose $p(x_t|x_{t-1}, y_{1:t})$ as the transition kernel, the weights have a minimum variance conditional on $x_{1:t-1}$. For that reason, $p(x_t|x_{t-1}, y_{1:t})$ is termed as the *optimal importance function* (See Doucet et al., 2000, p. 199). Under that optimal choice, the weights can be recursively computed as $w_t(x_{1:t}) = w_{t-1}(x_{1:t-1}) \cdot p(y_t|x_{t-1}, y_{1:t-1})$, where $p(y_t|x_{t-1}, y_{1:t-1})$ is termed as the *incremental importance weights*.[4]

[2] The realized particles are the outcomes of the categorical distribution, but the particles are random draws. The empirical c.d.f. evaluated at an arbitrary point is a random variable, which converges almost surely to the c.d.f. of $p(x_{1:t}|y_{1:t})$ evaluated at the same point. Then by the Polya Theorem and Glivenko-Cantelli Theorem, it is also the uniform convergence (see Athreya and Lahiri, 2006).

[3] Since we are interested in multiple characteristics (mean, variance, and so on) of the posterior state distributions, we interpret particle filtering as a Monte Carlo sampling method instead of a variance reduction technique for numerical integration. Therefore, we wish that the proposal density is as close to the target density as possible.

[4] For the target distribution $p(x_{1:t}|y_{1:t})$, the general form of the optimal importance function is $p(x_t|x_{1:t-1}, y_{1:t})$ with incremental weights $p(y_t|x_{1:t-1}, y_{1:t-1})$. For our state space model, $p(x_t|x_{1:t-1}, y_{1:t}) = p(x_t|x_{t-1}, y_{1:t})$, and $p(y_t|x_{1:t-1}, y_{1:t-1}) = p(y_t|x_{t-1}, y_{1:t-1})$. If the coefficient matrices are not functions of past observations, the

Proposition 1: The optimal importance function for ICSSM particle filtering is given by:

$$p(x_t|x_{t-1}, y_{1:t}) = \frac{\phi(x_t;\mu_t,\Sigma_t)}{F(\mu_t,\Sigma_t,\mathcal{X}_t)} \cdot 1(x_t \in \mathcal{X}_t), \tag{5}$$

where

$$\mu_t = A_t x_{t-1} + Q_t C_t'(C_t Q_t C_t' + R_t)^{-1}(y_t - C_t A_t x_{t-1}),$$
$$\Sigma_t = Q_t - Q_t C_t'(C_t Q_t C_t' + R_t)^{-1} C_t Q_t.$$

The incremental importance weights can be calculated as

$$p(y_t|x_{t-1}, y_{1:t-1}) = \phi(y_t; C_t A_t x_{t-1}, C_t Q_t C_t' + R_t) \cdot \frac{F(\mu_t,\Sigma_t,\mathcal{X}_t)}{F(A_t x_{t-1},Q_t,\mathcal{X}_t)}. \tag{6}$$

A proof of Proposition 1 is in the appendix. Eq (5) indicates that $p(x_t|x_{t-1}, y_{1:t})$ follows a truncated normal distribution $TN(\mu_t, \Sigma_t, \mathcal{X}_t)$. Meanwhile, y_t is subject to incidental truncation (refer to sample selection econometric models; Greene, 2008, p. 883) and $p(y_t|x_{t-1}, y_{1:t-1})$ follows an extended skewed normal distribution, whose density is given by Eq (6). The optimal importance function has a closed form mainly because the normalisation term $F(A_t x_{t-1}, Q_t, \mathcal{X}_t)$ can be viewed as a constant conditional on the past states, and μ_t, Σ_t in Eq (5) are the single-period Kalman filter outputs.

An alternative importance function, known as the bootstrap filter, only relies on the state transition. Period-t particles are generated from $TN(A_t x_{t-1}, Q_t, \mathcal{X}_t)$, with the incremental importance weights $\phi(y_t; C_t x_t, R_t)$. Despite its simplicity, the bootstrap filter could induce large Monte Carlo variations, as it ignores the current-period observation y_t in the proposal distribution.

To implement the particle filter, we generate period-t particles by Eq (5), and assign them weights by multiplying the previous weights by Eq (6). To evaluate the likelihood function, we take the sample average of the unnormalised weights. New particles are drawn from a univariate or low-dimensional truncated normal distribution, which is feasible by inversion

optimal importance function can be further simplified as $p(x_t|x_{t-1}, y_t)$. Meanwhile, the incremental weights reduce to $p(y_t|x_{t-1})$. However, for the Rao-Blackwellised particle filter, we only condition on a subset of the state variables. In that case, the entire history of the past constrained states matters. The optimal importance function $p(\xi_t|\xi_{1:t-1}, y_{1:t})$ and the increment weights $p(y_t|\xi_{1:t-1}, y_{1:t-1})$ cannot be further simplified.

sampling or rejection sampling. In practice, it is necessary to resample the particles when the weights are dispersed (or resample in every period).[5] Under the optimal importance function, the weights are not functions of the period-t particles. It is legitimate to reverse the order of sampling and resampling so as to preserve the diversity of the particles.

The bottom panel of Figure 1 illustrates the particle filtering results with nonnegative constraints imposed on the short rate series. In contrast with the Kalman filter that apparently yields negative state estimators, the constraints are honored for all outcomes of the particle-filtered distributions. Both the posterior means (the solid line) and the 95% intervals (the dotted lines) of the short rate series are positive.

4. Cross-sectional Rao-Blackwellisation

In some ICSSM applications, not all the state variables are subject to inequality constraints; some states might be free. It is desirable to decompose the filtering distribution into the analytically tractable and intractable components. The former has a conditionally linear sub-structure and thus can be marginalized by the Kalman filter. We only apply the particle filter to the latter so as to reduce Monte Carlo variations. That technique is known as Rao-Blackwellisation (see Doucet et al., 2001), or termed as mixture Kalman filters (Chen and Liu, 2000) or marginalized particle filtering (Schon et al., 2005).

Partition the state variables and let $x_t = (\xi_t', \eta_t')'$, where the $m_1 \times 1$ constrained states ξ_t must fall into the set $\Xi_t \subset \mathbb{R}^{m_1}$, while $m_2 \times 1$ states η_t are unconstrained. For notational convenience, we consider a diagonal model in which the state transition of ξ_t and η_t has no interactions (this assumption can be relaxed, see below), so that the transition matrix takes a block diagonal form $A_t = diag(A_{1t}, A_{2t})$, $Q_t = diag(Q_{1t}, Q_{2t})$, $C_t = (C_{1t}, C_{2t})$. The transition and observation densities for the state space model can be written as

[5] If we resample particles in period $t-1$, the particles are approximate draws from $p(x_{1:t-1}|y_{1:t-1})$. Then the next-period proposal density becomes $f_t(x_{1:t}) = p(x_{1:t-1}|y_{1:t-1}) \cdot p(x_t|x_{t-1}, y_{1:t})$. Provided that we reset the unnormalised weights to the likelihood function value immediately after resampling (i.e., assign $w_{t-1}(x_{1:t-1}) = p(y_{1:t-1})$ for all the resampled particles), the incremental importance weights for period-t particles still take the form $p(y_t|x_{t-1}, y_{1:t-1})$. No matter whether we resample or not, Proportion 1 specifies the two major steps, namely generating particles and assigning weights, for particle filtering.

$$p(\xi_t|\xi_{t-1}) = \frac{\phi(\xi_t; A_{1t}\xi_{t-1}, Q_{1t})}{F(A_{1t}\xi_{t-1}, Q_{1t}, \Xi_t)} \cdot 1(\xi_t \in \Xi_t),$$

$$p(\eta_t|\eta_{t-1}) = \phi(\eta_t; A_{2t}\eta_{t-1}, Q_{2t}),$$

$$p(y_t|\xi_t, \eta_t) = \phi(y_t; C_{1t}\xi_t + C_{2t}\eta_t, R_t).$$

The target distributions for particle filtering are $p(\xi_{1:t}, \eta_{1:t}|y_{1:t}), t = 1, \dots, T$, which can be decomposed as

$$p(\xi_{1:t}, \eta_{1:t}|y_{1:t}) = p(\xi_{1:t}|y_{1:t}) \cdot p(\eta_{1:t}|\xi_{1:t}, y_{1:t}).$$

On the one hand, $p(\eta_{1:t}|\xi_{1:t}, y_{1:t})$ is analytically tractable. Conditional on $\xi_{1:t}$, the system reduces to a Gaussian linear sub-model (GLSM), in which η_t is the state vector:

$$\eta_t = A_{2t}\eta_{t-1} + \varepsilon_{2t},$$

$$\tilde{y}_t = C_{2t}\eta_t + v_t,$$

where $\tilde{y}_t = y_t - C_{1t}\xi_t$, and $\varepsilon_{2t} \sim N(0, Q_{2t})$, $v_t \sim N(0, R_t)$.

In the GLSM, $p(\eta_{1:t}|\xi_{1:t}, y_{1:t})$ is a multivariate normal density, whose mean and variance can be obtained from the standard Kalman filter. The law of iterated expectations is useful for characterizing the unconditional mean $E(\eta_{1:t}|y_{1:t})$. Suppose we have obtained the weighted particles that represent $p(\xi_{1:t}|y_{1:t})$ (see below), then we can plug each particle into the Kalman filter, and the weighted average of the conditional means approximates the unconditional mean. Similarly, the unconditional variance equals the sum of the expected conditional variance and the variance of the conditional means.

On the other hand, the intractable component $p(\xi_{1:t}|y_{1:t})$ requires particle filtering. The optimal importance function $p(\xi_t|\xi_{1:t-1}, y_{1:t})$ and the incremental importance weights $p(y_t|\xi_{1:t-1}, y_{1:t-1})$ are summarized in the following proposition.

Proposition 2: The optimal importance function for particle filtering $p(\xi_{1:t}|y_{1:t})$, $t = 1, \dots, T$, takes the form:

$$p(\xi_t|\xi_{1:t-1}, y_{1:t}) = \frac{\phi(\xi_t; \mu_{\xi t}, \Sigma_{\xi t})}{F(\mu_{\xi t}, \Sigma_{\xi t}, \Xi_t)} \cdot 1(\xi_t \in \Xi_t), \quad (7)$$

where

$$\mu_{\xi t} = A_{1t}\xi_{t-1} + Q_{1t}C_{1t}'\Sigma_{yt}^{-1}(y_t - \mu_{yt}),$$

$$\Sigma_{\xi t} = Q_{1t} - Q_{1t}C_{1t}'\Sigma_{yt}^{-1}C_{1t}Q_{1t},$$

$$\mu_{yt} = C_{1t}A_{1t}\xi_{t-1} + C_{2t}\underline{\mu}_{\eta t},$$

$$\Sigma_{yt} = C_{1t}Q_{1t}C_{1t}' + C_{2t}\underline{\Sigma}_{\eta t}C_{2t}' + R_t.$$

The predictive moments $\underline{\mu}_{\eta t}$ and $\underline{\Sigma}_{\eta t}$ are functions of $\xi_{1:t-1}$, and can be recursively computed by the Kalman filter using the GLSM. To be specific,

$$\underline{\mu}_{\eta t} = A_{2,t}\bar{\mu}_{\eta,t-1},$$

$$\underline{\Sigma}_{\eta t} = A_{2,t}\bar{\Sigma}_{\eta,t-1}A_{2,t}' + Q_{2,t},$$

$$\bar{\mu}_{\eta,t} = \underline{\mu}_{\eta t} + \underline{\Sigma}_{\eta t}C_{2t}'\left(C_{2t}\underline{\Sigma}_{\eta t}C_{2t}' + R_t\right)^{-1}\left(y_t - C_{1t}\xi_t - C_{2t}\underline{\mu}_{\eta t}\right),$$

$$\bar{\Sigma}_{\eta,t} = \underline{\Sigma}_{\eta t} - \underline{\Sigma}_{\eta t}C_{2t}'\left(C_{2t}\underline{\Sigma}_{\eta t}C_{2t}' + R_t\right)^{-1}C_{2t}\underline{\Sigma}_{\eta t}.$$

The incremental importance weights under the optimal importance function are given by

$$p(y_t|\xi_{1:t-1}, y_{1:t-1}) = \phi\left(y_t; \mu_{yt}, \Sigma_{yt}\right) \cdot \frac{F\left(\mu_{\xi t}, \Sigma_{\xi t}, \Xi_t\right)}{F\left(A_{1t}\xi_{t-1}, Q_{1t}, \Xi_t\right)}. \tag{8}$$

A proof is in the appendix. The main reason that Proposition 2 holds is that the normalisation term $F(A_{1t}\xi_{t-1}, Q_{1t}, \Xi_t)$ can be treated as a constant conditioning on $\xi_{1:t-1}$. In the Rao-Blackwellised filter, each particle has a Kalman filter, which has contemporaneous interactions with importance sampling, as the Kalman filter "waits for" the realizations of particles before it updates the state distributions. Specifically, upon receiving the particles ξ_{t-1}, the Kalman filter calculates the filtered state distribution $\left(\bar{\mu}_{\eta,t-1}, \bar{\Sigma}_{\eta,t-1}\right)$ and predicts $\left(\underline{\mu}_{\eta t}, \underline{\Sigma}_{\eta t}\right)$ based on the GLSM. Then the Kalman filter pauses. The particle filter generates new particles ξ_t from $TN\left(\mu_{\xi t}, \Sigma_{\xi t}, \Xi_t\right)$ and assigns importance weights. Taking the particles for ξ_t as given, the Kalman filter updates $\left(\bar{\mu}_{\eta,t}, \bar{\Sigma}_{\eta,t}\right)$ and proceeds to period $t+1$ for $\left(\underline{\mu}_{\eta,t+1}, \underline{\Sigma}_{\eta,t+1}\right)$, and so on.

The assumption on the diagonal model can be relaxed, and A_t, Q_t are not necessarily block diagonal matrices. Cross-sectional Rao-Blackwellisation is applicable provided that the normalisation term $F(A_t x_{t-1}, Q_t, \mathcal{X}_t)$ in Eq (1) is not a function of the past unconstrained states. For example, when A_t is a block lower-triangular matrix and Q_t is a full matrix, the normalisation term only depends on the past constrained states, and thus can be treated as a constant term conditional on $\xi_{1:t-1}$. It follows that

$$p(x_{1:t-1}|\xi_{1:t-1}, y_{1:t-1}) \propto \prod_{\tau=1}^{t-1} \phi(x_\tau; A_\tau x_{\tau-1}, Q_\tau) \cdot \phi(y_\tau; C_\tau x_\tau, R_\tau),$$

which is a Gaussian density whose means and variances are outputs of the Kalman filter using an expanded linear state space model for $\tau = 1, \dots, t-1$:

$$x_\tau = A_\tau x_{\tau-1} + \varepsilon_\tau, \tag{9}$$

$$y_\tau = C_\tau x_\tau + v_\tau, \tag{10}$$

$$\xi_\tau = \left(I_{m_1 \times m_1}, 0_{m_1 \times m_2}\right) \cdot x_\tau, \tag{11}$$

where $x_\tau = (\xi_\tau', \eta_\tau')'$, and Eq (11) is a perfect measurement as the state itself is observed.[6]

Given that $p(x_{1:t-1}|\xi_{1:t-1}, y_{1:t-1})$ is a Gaussian density, the state constraints take effects only in period t. As a result, the optimal important function $p(\xi_t|\xi_{1:t-1}, y_{1:t})$ remains to be a tractable low-dimensional truncated normal distribution, and $p(y_t|\xi_{1:t-1}, y_{1:t-1})$ is still the extended skewed normal distribution.

We may interpret the cross-sectional Rao-Blackwellised particle filter as a two-step Kalman filter for each particle. Denote $x_t|\xi_{1:t-1}, y_{1:t} \sim TN\left(\underline{\mu}_{x,t}, \underline{\Sigma}_{x,t}, \mathcal{X}_t\right)$ and $x_t|\xi_{1:t}, y_{1:t} \sim N\left(\bar{\mu}_{x,t}, \bar{\Sigma}_{x,t}\right)$, where $\underline{\mu}_{x,t}, \underline{\Sigma}_{x,t}, \bar{\mu}_{x,t}, \bar{\Sigma}_{x,t}$ can be recursively computed by a two-step process. In the first step, given $\left(\bar{\mu}_{x,t-1}, \bar{\Sigma}_{x,t-1}\right)$, we employ a single-period Kalman filter based on Eq (9) and (10) to calculate $\left(\underline{\mu}_{x,t}, \underline{\Sigma}_{x,t}\right)$, which will be used for generating period-t particles. In the second step, given $\left(\bar{\mu}_{x,t-1}, \bar{\Sigma}_{x,t-1}\right)$ and the new particles, we use a single-period Kalman filter based on Eq (9), (10) and (11) to compute $\left(\bar{\mu}_{x,t}, \bar{\Sigma}_{x,t}\right)$.

This algorithm is valid provided that the normalisation term does not interfere with the conditional distributions, which requires that the past unconstrained states have no impact on the normalisation term. If such requirement cannot be satisfied, there is a remedy. Note that an unconstrained state can be classified as a constrained one with infinity bounds. Therefore, cross-sectional Rao-Blackwellisation is applicable if a subset of the unconstrained states have no influence on the normalisation term.

[6] Alternatively, we can plug the perfect measurement into the state equation, and arrive at an observation equation and a transition equation for η_t. For example, suppose $x_t = A_t x_{t-1} + \varepsilon_t$ can be decomposed as $\xi_t = A_{11t}\xi_{t-1} + A_{12t}\eta_{t-1} + \varepsilon_{1t}$ and $\eta_t = A_{21t}\xi_{t-1} + A_{22t}\eta_{t-1} + \varepsilon_{2t}$. The former is an observation equation with $\xi_t - A_{11t}\xi_{t-1}$ being observed, while the latter is a transition equation.

5. Temporal Rao-Blackwellisation

In the era of high interest rates, few practitioners concerned about the negative rates. Not until recent years when the interest rates plummeted did such concern loom large. Though an inequality constraint always binds the posterior state distribution, the restriction can be tight or loose, depending on the probability that the unrestricted state distribution violates the constraint. In Section 6 we demonstrate that a loosely constrained state behaviors virtually the same as an unconstrained one. It is sensible to impose a constraint only if there is a substantial probability that the constraint is violated. We design a particle filter that can switch to the Kalman filter for analytic results whenever the constraints are absent. This is in accordance with Rao-Blackwellisation, which exploits the Gaussian linear sub-structure for analytic integration. In contrast with cross-sectional marginalization that employs the Kalman filter on a subset of the states, temporal Rao-Blackwellisation resorts to the Kalman filter in a subsample.

Consider Eq (1) – (3) with time-varying constraints. Suppose that $\mathcal{X}_t = \mathbb{R}^m$ for $t = S + 1, \dots, V$, where $1 < S < V < T$. That is, ICSSM reduces to a linear system Eq (3) and (4) in the subsample from period $S + 1$ to V. We are interested in the filtering distribution $p(x_t | y_{1:t}), t = 1, \dots, T$ as well as the likelihood function $p(y_{1:t})$. Suppose that we have employed the particle filter in the first S periods and the filtering distribution $p(x_{1:S} | y_{1:S})$ are represented by K particles $x_{1:S}^{(i)}$ with the unnormalised weights $\underline{w}_S^{(i)}, i = 1, \dots, K$. In practice, we may only store $x_S^{(i)}$ instead of the entire series.

The question is how to switch to the Kalman filter. It is tempting to initialize the Kalman filter by computing $E(x_S | y_{1:S})$ and $Var(x_S | y_{1:S})$ using the weighted particles, and then apply the standard Kalman filter for $t = S + 1, \dots, V$. The Kalman filter can produce the best linear state estimator, but cannot characterize the non-Gaussian filtering distribution and cannot represent the likelihood function for ICSSM. It is also tempting to apply the Kalman filter under each of the deterministic initial state $x_S^{(i)}$, and then uses the weight $\underline{w}_S^{(i)}$ to average the Kalman filter outputs. As is shown in the following proposition, such method is flawed because the correct weight should incorporate the information contents of $y_{S+1:V}$.

Proposition 3: Assume that $p(x_S|y_{1:S})$ is a categorical distribution represented by the K particles $x_S^{(i)}$ with the unnormalised weights $\underline{w}_S^{(i)}, i = 1, \dots, K$. Then for the unconstrained periods $t = S + 1, \dots, V$, we have

$$E(x_t|y_{1:t}) = X_{t|t} \cdot \begin{bmatrix} 1 \\ E(x_S|y_{1:t}) \end{bmatrix},$$

$$Var(x_t|y_{1:t}) = P_{t|t} + X_{t|t} \cdot \begin{bmatrix} 0 & 0_{1\times m} \\ 0_{m\times 1} & Var(x_S|y_{1:t}) \end{bmatrix} \cdot X'_{t|t},$$

where $E(x_S|y_{1:t})$ and $Var(x_S|y_{1:t})$ are the mean and variance for the smoothed distribution defined by the same particles $x_S^{(i)}$ with the updated weights $\overline{w}_S^{(i)}, i = 1, \dots, K$:

$$\overline{w}_S^{(i)} \propto \underline{w}_S^{(i)} \cdot \prod_{\tau=S+1}^{t} \phi\left[V_\tau \cdot \begin{pmatrix} 1 \\ x_S^{(i)} \end{pmatrix}; 0, O_{\tau|\tau-1}\right], \tag{12}$$

where the proportionality constant equals the sum of the right hand side of the equation.

To evaluate the likelihood function,

$$\hat{p}(y_{1:t}) = \frac{1}{K} \sum_{i=1}^{K} \left\{\underline{w}_S^{(i)} \cdot \prod_{\tau=S+1}^{t} \phi\left[V_\tau \cdot \begin{pmatrix} 1 \\ x_S^{(i)} \end{pmatrix}; 0, O_{\tau|\tau-1}\right]\right\} \tag{13}$$

is a consistent estimator for the likelihood value $p(y_{1:t})$.

The matrices $X_{t|t}, P_{t|t}, V_t, O_{t|t-1}$ are obtained from the augmented Kalman filter (see Durbin and Koopman, 2012, p. 141). The forward recursion starts from the matrices $X_{S|S} = (0_{m\times 1}, I_{m\times m})$, $P_{S|S} = 0_{m\times m}$. For period $t = S + 1, \dots, V$, we sequentially compute the following variables:

$$X_{t|t-1} = A_t X_{t-1|t-1},$$

$$P_{t|t-1} = A_t P_{t-1|t-1} A'_t + Q_t,$$

$$Y_{t|t-1} = C_t X_{t|t-1},$$

$$O_{t|t-1} = C_t P_{t|t-1} C'_t + R_t,$$

$$V_t = [y_t, 0_{n\times m}] - Y_{t|t-1},$$

$$X_{t|t} = X_{t|t-1} + P_{t|t-1} C'_t \left(O_{t|t-1}\right)^{-1} V_t,$$

$$P_{t|t} = P_{t|t-1} - P_{t|t-1} C'_t \left(O_{t|t-1}\right)^{-1} C_t P_{t|t-1}.$$

A proof is in the appendix. Proposition 3 shows that $E(x_t|y_{1:t})$ can be computed by the law of iterated expectations. Given a deterministic initial state x_S, the conditional mean

$E(x_t|x_S, y_{1:t})$ is a Kalman filter output. Since each particle represents a different initial state, it is legitimate to take the weighted average of the Kalman filter outputs. However, the correct weights come from the smoothing distribution $p(x_S|y_{1:t})$.

We present Proposition 3 in terms of the extended Kalman filter because of its computational efficiency for multi-period unconstrained filtering with varied initial state values. A more computationally intensive version is a Kalman filter for each and every particle.

Random samples from $p(x_t|y_{1:t})$ can be generated using the following identity:

$$p(x_s, x_t|y_{1:t}) = p(x_s|y_{1:t}) \cdot p(x_t|x_s, y_{1:t}). \tag{14}$$

We can first generate a draw from the smoothing distribution $p(x_S|y_{1:t})$, which is essentially a resample of the particles with weights given by Eq (12). Conditional on that draw, we sample from $p(x_t|x_S, y_{S+1:t})$ using the Kalman filter. Those equally-weighted samples fully characterize the filtering distribution $p(x_t|y_{1:t})$. Alternatively, weighted draws can also represent that distribution. Note that

$$p(x_t|y_{1:t}) = \int p(x_s, x_t|y_{1:t}) dx_s = \sum_{i=1}^{K} \left\{ \overline{w}_S^{(i)} \cdot p\left(x_t \middle| x_S^{(i)}, y_{1:t}\right) \right\}.$$

We may use the Kalman filter to generate a draw from $p\left(x_t \middle| x_S^{(i)}, y_{1:t}\right)$ based on the original particles, then assign it with the smoothing weight $\overline{w}_S^{(i)}$.

In practice, we only need to generate random samples or weighted samples in period V, as the state constraints will be in effect again and we switch to the particle filter for $t = V + 1, \dots, T$. If we treat the random samples generated from $p(x_V|y_{1:V})$ as the period-V particles, we assign them the unnormalised weights $p(y_{1:V})$, an estimator of which is given by Eq (13). If we treat the weighted samples as the period-V particles, we assign them the unnormalised weights given by the right hand side of Eq (12), namely the unnormalised version of $\overline{w}_S^{(i)}$. Both methods ensure that the mean of the unnormalised weights approximates the likelihood function.

Temporal Rao-Blackwellisation can be carefully interpreted as a special case of cross-sectional Rao-Blackwellisation, if we allow time-varying state dimensions, support linear algebra with empty matrices, and handle judiciously the last unconstrained period. In Section 4, we partition the state vector x_t into the constrained ξ_t and unconstrained η_t, and then simulate $p(\xi_{1:t}|y_{1:t})$ by the particle filter. To apply that algorithm under time-varying constraints, we put $\xi_t = x_t$ for $t = 1, \dots, S, V, V+1, \dots, S$. For period $t = S+1, \dots, V-1$, all the state variables are classified as the unconstrained states. There are no new particles to generate. However, the incremental importance weights must be computed in order to update the unnormalised weights.[7] This is exactly the smoothing procedure given by Eq (12). Most importantly, we give a special treatment by artificially labelling the period-V states as constrained states with infinity bounds, because cross-sectional Rao-Blackwellisation requires that the normalisation term $F(\cdot)$ cannot be a function of the past unconstrained states.[8] We generate period-V particles from $p(x_V|x_S, y_{1:V})$ with the importance weights given by Eq (12). If we resample particles, and reverse the order of sampling and resample under the optimal importance function, this is exactly the sampling procedure given by Eq (14).

6. An Application

Time-varying parameter regression is a popular economic application of state space modelling. Some well-known studies include Cogley and Sargent (2005), Primiceri (2005), and Stock and Watson (2007). Parameter uncertainty and instability are addressed by random coefficients, which are assumed to follow the random walk or autoregressive processes. For example, an AR(2) model with random-walk coefficients can be specified as

$$y_t = \phi_0 + \phi_{1t} y_{t-1} + \phi_{2t} y_{t-2} + \varepsilon_t,$$
$$\phi_{1t} = \phi_{1,t-1} + u_{1t},$$
$$\phi_{2t} = \phi_{2,t-1} + u_{2t},$$

where the independent noises satisfy $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, $u_{1t} \sim N(0, \sigma_1^2)$, $u_{2t} \sim N(0, \sigma_2^2)$.

[7] Recall that the target distributions are $p(\xi_{1:t}|y_{1:t}), t = 1, \dots, T$. For period $t = S+1, \dots, V-1$, the unconstrained states ξ_t become an empty set, hence no new particles. However, the target distribution still evolves as $y_{1:t}$ expand. The importance weights must be updated accordingly.

[8] Unconstrained states in period $V-1, V-2$, etc. do not need a special treatment because the normalisation terms in those periods equal to one, which is not a function of any variables.

It is desirable to impose restrictions on ϕ_{1t} and ϕ_{2t} so that the AR(2) process is non-explosive at each point in time. Triangular conditions ensure that under any realizations of the random coefficients, the eigenvalues of the AR(2) process never fall outside the unit circle.

$\phi_{2t} + \phi_{1t} \leq 1,$

$\phi_{2t} - \phi_{1t} \leq 1,$

$\phi_{2t} \geq -1.$

Our data are quarterly U.S. civilian unemployment rates 1969 – 2015. Observations range from 3.4 to 10.8, with the mean 6.3 and standard deviation 1.6. This time series exhibits geometrically decaying autocorrelations and a clear truncation after two lags for the partial autocorrelations. By the Box and Jenkins (1970) approach, this series is ideal for an AR(2) model. However, the ten-year subsample rolling window AR(2) regressions reveal that the coefficients ϕ_{1t}, ϕ_{2t} have fluctuations, and the persistency of the time series, measured by $\phi_{2t} + \phi_{1t}$, has counter-cyclical movements. The persistency becomes higher when the economy is in recession. Based on the rolling window estimation results, we informally calibrate the model parameters: $\phi_0 = 0.404$, $\sigma_\varepsilon = 0.286$, $\sigma_1 = 0.047$, $\sigma_2 = 0.044$.[9] Then we run the unconstrained Kalman filter, with the initial states obtained from a ten-year presample AR(2) regression. As seen in the upper panel of Figure 2, the filtered series for $\phi_{2t} + \phi_{1t}$ exhibit large spikes that exceed the unity upper bound. However, the other two inequality constraints are unlikely to tightly bind the states because the estimated $\phi_{2t} - \phi_{1t}$ range from -1.47 to -0.57, and the estimated ϕ_{2t} ranges between -0.28 and 0.24.

To apply the ICSSM particle filter, we could impose all constraints in all periods. However, as we will see shortly, particle filtering results under loose constraints are nearly identical to the unconstrained filtering results. It is sensible to impose a constraint only if there is a substantial probability that the unrestricted filter violates the constraint. Parsimony keeps Monte Carlo variations to the minimum, and enhances reliability of particle filtering. After some trials, we decide to impose the inequality $\phi_{2t} + \phi_{1t} \leq 1$ in periods when the unconstrained Kalman

[9] The calibrated ϕ_0 is obtained from the average intercept estimates of the rolling window AR(2), and σ_ε from the average disturbance standard deviation, and σ_1, σ_2 from the standard deviations of the first differenced rolling window AR(2) coefficient series.

estimate on $\phi_{2t} + \phi_{1t}$ is larger than 0.95. That is, 31 out of 186 periods are subject to the constraint and they are marked on the horizontal axis of the middle panel of Figure 2.

Since the unnormalised weights of the particles can approximate the likelihood function, we use the maximum simulated likelihood method to estimate the unknown parameters: $\phi_0 = 0.643$, $\sigma_\varepsilon = 0.254$, $\sigma_1 = 0.021$, $\sigma_2 = 0.002$. Though the estimated parameters are different from the informally calibrated ones, the unconstrained state estimators have similar patterns, as we compare the dashed lines in the upper and middle panels of Figure 2. The spikes of $\phi_{2t} + \phi_{1t}$ still exceed one.

Figure 2 demonstrates that the particle filter effectively suppresses all the spikes that violate the constraint. For instance, the unconstrained estimator in the first quarter of 2009 is 1.030, while the constrained estimator is 0.987. When the estimated $\phi_{2t} + \phi_{1t}$ are relatively low, the Kalman and particle filters yield nearly identical results. For example, in the years around 1991 and 2001, the constraint is in effect but the state estimators are away from the upper bound. The constrained and unconstrained curves overlap with each other.

The bottom panel of Figure 2 shows the filtering results by including an additional constraint $\phi_{2t} \geq -1$, which we claimed to be a loose constraint. We put $(\phi_{2t} + \phi_{1t}, -\phi_{2t})'$, instead of $(\phi_{1t}, \phi_{2t})'$, as the state vector, so that the normalisation term $F(\cdot)$ in Eq (1) reduces to the bivariate normal c.d.f.. The filtered series with and without this constraint almost overlap, and the maximum discrepancy between them is 0.0008. However, the bivariate normal c.d.f. is more computationally expensive than the univariate one. The computing time is about 10 times longer.

Taking this application as a numerical exercise, we show the efficiency of particle filtering by adopting the optimal importance function, cross-sectional and temporal Rao-Blackwellisation. Table 1 compares the filtering results using alternative particle filtering algorithms, including 1) the bootstrap particle filter (BT) that generates new particles solely based on the state transition; 2) the particle filter with the optimal importance function but without Rao-Blackwellisation (PF); 3) the temporal Rao-Blackwellised particle filter with the optimal importance function (TS); and 4) the optimal filter with cross-sectional and temporal Rao-Blackwellisation under the optimal importance function (OP). For each algorithm, we use 500 particles to approximate the filtering distributions. We repeat the exercise by 500 times using different sets of random particles. The standard deviations of the results among the 500 experiments reflect the Monte Carlo variations of the algorithms. We also report the average filtering results, the approximated likelihood function values, as well as the computing time. The codes are written in MATLAB and run on a personal computer.

Table 1 reports the average state estimators on selected dates. 1969:Q3 (the third quarter of the year 1969) is an unconstrained period (the first constrained date 1970:Q1). Both TS and OP automatically resort to the standard Kalman filter, and the results are free from Monte Carlo errors with zero standard deviations (up to floating-point numeric errors). However, BT and PF treat no constraints as infinity bounds and implement the particle filter. The filtering results are similar to the analytic filter but polluted by simulation noises. By using the optimal importance function, PF reduces Monte Carlo variations by 50%, as the standard deviation drops from 3.6e-3 to 1.8e-3. The unconstrained Kalman filter violates the upper bound in 1974:Q4, while all versions of the particle filters obey the constraint, not only for the posterior means but also for all outcomes of the filtering distribution. Though the four methods offer similar state estimators (about 0.975), the standard deviations of BT, PF, TS and OP are 4.3e-3, 2.0e-3, 1.8e-3 and 1.7e-3, respectively. PF is better than BT because of the optimal importance function; TS is better than PF because of temporal Rao-Blackwellisation; OP is better than TS because of cross-sectional Rao-Blackwellisation. Similar patterns of variance reduction can also be seen from the filtering results in other periods.

Monte Carlo errors plague maximum simulated likelihood estimation. A fundamental way to improve simulation-based estimation is to reduce the Monte Carlo variations in evaluating the likelihood function. Figure 3 illustrates a rudimentary grid search for the optimal σ_ε (with other parameters being fixed). We put 100 discrete points evenly spaced between 0.22 and 0.29. For each σ_ε point, we run the particle filter once with 300 particles. So each σ_ε corresponds to a noisy likelihood value. As seen in Figure 3, the BT results are so noisy that we can hardly see the trend of the curve. The PF and TS results are less volatile, and we can roughly see a hump shaped function. The OP results are of highest quality, with the simulated likelihood values concentrating around an inverted-U function. The maximizer is near 0.25. We found that OP can reliably approximate the likelihood value even with a small number of particles.

7. An Alternative View on State Constraints

Having presented our ICSSM and its particle filtering algorithms, we discuss an alternative model for inequality constraints. State space models use observations to update the prior state distributions. The constraints can be viewed as additional observations for learning the latent states, whose posterior distributions conditional on such observations satisfy the inequality constraints. A model is said to be a posterior constrained state space model (PCSSM) if

$$x_t = A_t x_{t-1} + \varepsilon_t, \tag{15}$$

$$y_t = C_t x_t + v_t, \tag{16}$$

$$z_t = 1(x_t \in \mathcal{X}_t), \tag{17}$$

where the independent disturbances satisfy $\varepsilon_t \sim N(0, Q_t)$, $v_t \sim N(0, R_t)$. We introduce an auxiliary measurement variable z_t to represent the inequality constraints on the states. In addition to the regular observations y_t, we also observe $z_t = 1, \forall t$.

To make the concept of the prior and posterior state distributions transparent, we will ignore the fact that coefficient matrices could be functions of past observations.

ICSSM and PCSSM are two different models that describe the inequality constraints on the states. ICSSM imposes constraints at the prior stage, while the constraints in PCSSM are honored in the posterior distributions. The TVP-VAR model in Cogley and Sargent (2005) can be interpreted as a PCSSM, while that in Koop and Potter (2011) can be viewed as an ICSSM.[10] A key difference is that the ICSSM normalisation term $F(A_t x_{t-1}, Q_t, \mathcal{X}_t)$ is a function of the past states, while that of PCSSM is truly a constant. Consequently, PCSSM has analytic properties that ICSSM does not have, which translates to computational advantages in favor of PCSSM.

To derive the analytical results for PCSSM, we resort to the matrix formulation of the state space model. Eq (15) and (16) can be written as:

[10] Refer to Eq (6) in Cogley and Sargent(2005, p. 266). The prior state distribution for the constrained model is proportional to that for the unconstrained model, hence PCSSM. Also refer to Eq (4) in Koop and Potter (2011, p. 1129). The prior state distribution includes a prior integrating constant in the denominator, hence ICSSM.

$Ax_{1:t} = c_0 + \varepsilon_{1:t}$,

$y_{1:t} = Cx_{1:t} + v_{1:t}$,

where $A = \begin{pmatrix} I & 0 & \cdots & 0 & 0 \\ -A_2 & I & \cdots & 0 & 0 \\ \cdots & \cdots & \ddots & \cdots & \cdots \\ 0 & 0 & \cdots & -A_t & I \end{pmatrix}$, $c_0 = \begin{pmatrix} A_1 x_0 \\ 0 \\ \cdots \\ 0 \end{pmatrix}$, and $C = diag(C_1, \dots, C_t)$. Note that A is lower-triangular with all nonzero entries on the diagonal, hence an invertible matrix.

Proposition 4: Conditional on $z_{1:t} = 1$, the state filtering distribution for PCSSM is given by:[11]

$p(x_{1:t}|z_{1:t}) = \frac{\phi\left(x_{1:t};\underline{\mu}_{1:t},\underline{\Sigma}_{1:t}\right)}{F\left(\underline{\mu}_{1:t},\underline{\Sigma}_{1:t},\mathcal{X}_{1:t}\right)} \cdot 1(x_{1:t} \in \mathcal{X}_{1:t})$,

$p(x_{1:t}|y_{1:t}, z_{1:t}) = \frac{\phi(x_{1:t};\overline{\mu}_{1:t},\overline{\Sigma}_{1:t})}{F(\overline{\mu}_{1:t},\overline{\Sigma}_{1:t},\mathcal{X}_{1:t})} \cdot 1(x_{1:t} \in \mathcal{X}_{1:t})$,

where

$\underline{\mu}_{1:t} = A^{-1}c_0$,

$\underline{\Sigma}_{1:t} = A^{-1}QA'^{-1}$,

$\overline{\mu}_{1:t} = \underline{\mu}_{1:t} + \underline{\Sigma}_{1:t}C'\left(C\underline{\Sigma}_{1:t}C' + R\right)^{-1}\left(y_{1:t} - C\underline{\mu}_{1:t}\right)$,

$\overline{\Sigma}_{1:t} = \underline{\Sigma}_{1:t} - \underline{\Sigma}_{1:t}C'\left(C\underline{\Sigma}_{1:t}C' + R\right)^{-1}C\underline{\Sigma}_{1:t}$,

$Q = diag(Q_1, \dots, Q_t)$, $R = diag(R_1, \dots, R_t)$, $\mathcal{X}_{1:t} = \{x_{1:t}|x_\tau \in \mathcal{X}_\tau, \forall\tau = 1, \dots, t\}$.

Proposition 4 suggests that the constrained state distribution for PCSSM, prior to the observations $y_{1:t}$, is a multivariate truncated normal distribution $TN\left(\underline{\mu}_{1:t}, \underline{\Sigma}_{1:t}, \mathcal{X}_{1:t}\right)$. Conditional on $y_{1:t}$, the posterior state distribution is updated to $TN\left(\overline{\mu}_{1:t}, \overline{\Sigma}_{1:t}, \mathcal{X}_{1:t}\right)$. In contrast, the ICSSM prior and posterior distributions are of unknown form.

The closed-form PCSSM optimal filter exists, as the moment generating function for a multivariate truncated normal distribution has an analytic expression. Suppose that $\mathcal{X}_{1:t}$ is a

[11] We slightly abused the notation, as the right hand side of the equation actually represents $p(x_{1:t}|z_{1:t} = 1)$. Since the auxiliary observations $z_{1:t}$ are always 1, we simply write it as $p(x_{1:t}|z_{1:t})$. However, if we interpret Eq (15) – (17) as a Probit state space model, in which z_t is a binary response that could be either 0 or 1, then we must replace the indicator function $1(x_{1:t} \in \mathcal{X}_{1:t})$ by $\prod_{\tau=1}^{t}[1(x_\tau \in \mathcal{X}_\tau) \cdot 1(z_\tau = 1) + 1(x_\tau \notin \mathcal{X}_\tau) \cdot 1(z_\tau = 0)]$.

rectangular set such that $\mathcal{X}_{1:t} = \{x_{1:t} | a_{1:t} \leq x_{1:t} \leq b_{1:t}\}$, where $a_{1:t}$ and $b_{1:t}$ are the lower and upper bounds of the states. Using the result of Tallis (1961), the moment generating function for $TN(\overline{\mu}_{1:t}, \overline{\Sigma}_{1:t}, \mathcal{X}_{1:t})$ is given by

$$E\left(e^{r' x_{1:t}} | y_{1:t}, z_{1:t}\right) = e^{r'\overline{\mu}_{1:t} + \frac{1}{2} r' \overline{\Sigma}_{1:t} r} \cdot \frac{F(\overline{\Sigma}_{1:t} r,\ \overline{\Sigma}_{1:t},\ a_{1:t} - \overline{\mu}_{1:t},\ b_{1:t} - \overline{\mu}_{1:t})}{F(0,\ \overline{\Sigma}_{1:t},\ a_{1:t} - \overline{\mu}_{1:t},\ b_{1:t} - \overline{\mu}_{1:t})}.$$

The PCSSM particle filter with the optimal importance function is still tractable. The following result is the PCSSM version of Proposition 1.

Proposition 5: Given $z_t = 1, \forall t$, the optimal importance function for particle filtering $p(x_{1:t} | y_{1:t}, z_{1:t}), t = 1, \dots, T$ takes the following form:

$$p(x_t | x_{t-1}, y_{1:t}, z_{1:t}) = \frac{\phi(x_t; \mu_t, \Sigma_t)}{F(\mu_t, \Sigma_t, a_t, b_t)} \cdot 1(x_t \in \mathcal{X}_t),$$

where μ_t, Σ_t are defined below Eq (5). Given $z_t = 1$, the incremental importance weights are

$$p(y_t, z_t | x_{t-1}, y_{1:t-1}, z_{1:t-1}) = \phi(y_t; C_t A_t x_{t-1}, C_t Q_t C_t' + R_t) \cdot F(\mu_t, \Sigma_t, \mathcal{X}_t).$$

A proof is in the appendix. Note that the optimal importance functional form is nearly identical to Eq (5). This is because conditional on x_{t-1}, the normalisation term is a constant for both ICSSM and PCSSM. PCSSM has computational advantages over ICSSM in two aspects. First, to calculate the incremental importance weights for PCSSM, we compute only one multivariate normal probability, while ICSSM requires the ratio of two probabilities. When there are multiple constrained states, calculating or simulating the truncation probabilities could be computationally intensive. PCSSM reduces computation roughly by an half. Second, for an unlucky draw of a particle, the two probabilities could be close to zero. Their ratio is unpredictable due to numerical errors, which could undermine particle filtering as some particles with unreasonably large weights may propagate to the next-period particle filtering. However, PCSSM is immune to that problem because of a self-stabilizing mechanism: the incremental weights assigned to the unlucky particles are also close to zero when $F(\mu_t, \Sigma_t, \mathcal{X}_t)$ is small. The bad particles are likely to be discarded by resampling, and thus numerical inaccuracy in calculating the probability and generating truncated draws will not propagate.

The temporal Rao-Blackwellised particle filter can be immediately applied to PCSSM. Proposition 3 only utilizes the augmented Kalman filter for the unconstrained periods, given the initial particles and weights obtained from either ICSSM or PCSSM. For cross-sectional Rao-Blackwellisation, PCSSM has an advantage to ICSSM. Since the normalisation term is always a constant in PCSSM, Rao-Blackwellisation is applicable no matter how ξ_t and η_t interact in the state transition. The PCSSM counterpart of Proposition 2 is stated below.

Proposition 6: Given $z_t = 1, \forall t$, the optimal importance function for particle filtering $p(\xi_{1:t}|y_{1:t}, z_{1:t}),\ t = 1, \dots, T$ in the diagonal model takes the form:

$$p(\xi_t|\xi_{1:t-1}, y_{1:t}, z_{1:t}) = \frac{\phi(\xi_t; \mu_{\xi t}, \Sigma_{\xi t})}{F(\mu_{\xi t}, \Sigma_{\xi t}, \Xi_t)} \cdot 1(\xi_t \in \Xi_t),$$

where $\mu_{\xi t}, \Sigma_{\xi t}$ and their offspring variables $\mu_{yt}, \Sigma_{yt}, \underline{\mu}_{\eta t}, \underline{\Sigma}_{\eta t}, \bar{\mu}_{\eta,t}, \bar{\bar{\Sigma}}_{\eta,t}$ are the same as those defined in Proposition 2. Given $z_t = 1$, the incremental importance weights associated with the optimal importance function are given by

$$p(y_t, z_t|\xi_{1:t-1}, y_{1:t-1}) = \phi(y_t; \mu_{yt}, \Sigma_{yt}) \cdot F(\mu_{\xi t}, \Sigma_{\xi t}, \Xi_t).$$

Last but not the least, PCSSM particle filtering is also helpful for the simulation smoother. The multi-move simulation smoother for PCSSM is a rejection sampling algorithm. Using the unconstrained model for candidate draws from $\phi(x_{1:T}; \bar{\mu}, \bar{\bar{\Sigma}})$, we reject a candidate draw if a state at any point of time violates the inequality constraints. The acceptance probability equals the normalisation term, namely $F(\overline{\mu}_{1:T}, \overline{\Sigma}_{1:T}, \mathcal{X}_{1:T})$, which is a non-increasing function of T. As T grows, it becomes increasingly difficult to accept a candidate draw. A rough idea on the magnitude of the acceptance probability is helpful for evaluating the feasibility of the algorithm. For the ICSSM multi-move simulation smoothing, Koop and Potter (2011) use a Metropolis-Hastings sampler with proposal draws generated from the unconstrained model. An estimate on the acceptance probability is still useful.

With the aid of the PCSSM particle filter, the acceptance probability can be estimated:

$$F(\overline{\mu}_{1:T}, \overline{\Sigma}_{1:T}, \mathcal{X}_{1:T}) = \frac{p(y_{1:T}, z_{1:T})}{p(y_{1:T})}.$$

The nominator is the PCSSM likelihood function value, which can be consistently estimated by the sample mean of the unnormalised weights of the particles. The denominator is the likelihood function value for a Gaussian linear state space model comprising Eq (15) and (16). The standard Kalman filter provides an exact evaluation of $p(y_{1:T})$ in the prediction error decomposition form.

Despite computational virtues, PCSSM is awkward in its data generating process, as it is difficult to draw the triple $(x_{1:T}, y_{1:T}, z_{1:T})$ by Eq (15) – (17). We may first generate a candidate $x_{1:T}$ by Eq (15), but have to reject it if any state violate the constraints. The observations $y_{1:T}$, as seen in a real-world data set, are the result of a lucky state sequence that survives all the constraints.

8. Conclusion

When a state space model is subject to inequality constraints, analytic tractability offered by the Kalman filter is lost. A crude approximation of a high-dimensional unknown distribution could yield poor filtering results. Our method is based on the standard particle filter but not as computationally expensive, since we exploit the Gaussian linear sub-structure for analytic integration whenever possible. The three major features of our particle filter are the optimal importance function, cross-sectional and temporal Rao-Blackwellisation. The optimal importance function is a single-period analytic filter. Cross-sectional Rao-Blackwellisation marginalizes the unconstrained states by the Kalman filter. Temporal Rao-Blackwellisation skips particle filtering in the unconstrained periods and bridges particles of two disjoint periods by the augmented Kalman filter.

A practical workflow of imposing state constraints begins with examining the filtering distributions obtained from the unconstrained Kalman filter. We advocate parsimony and suggest imposing the constraints that tightly bind the states. A constraint is tight when its violation probability is large. For example, if the mean of the unconstrained filtering distribution has exceeded the bound (that is, the Kalman filter has violated the constraint), the violation probability can be deemed large heuristically. Tight constraints can be effectively enforced by the particle filter, which ensures that all outcomes of the filtering distributions honor the constraints. Meanwhile, parsimony is a virtue because the Kalman and particle filters yield similar results under loose constraints. Parsimony translates to smaller Monte Carlo errors and faster filtering. Given the constraints judiciously chosen by the users, our MATLAB program can identify the most suitable cross-sectional and temporal Rao-Blackwellisation scheme, and implement the particle filter with the optimal importance function.

References

Athreya, K. B., Lahiri, S. N., 2006. Measure Theory and Probability Theory. New York: Springer.

Black, F., 1995. Interest rates as options. The Journal of Finance 50 (5), 1371–1376.

Box, G., Jenkins, G., 1970. Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day.

Carter, C. K., Kohn, R., 1994. On Gibbs sampling for state space models. Biometrika 81 (3), 541–553.

Chen, R., Liu, J. S., 2000. Mixture Kalman filters. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62 (3), 493-508.

Chopin, N., 2004. Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. The Annals of Statistics 32 (6), 2385–2411.

Cogley, T., Sargent, T. J., 2001. Evolving post-World War II U.S. inflation dynamics. In: NBER Macroeconomics Annual 2001, Volume 16. Cambridge: MIT Press, 331–388.

Cogley, T., Sargent, T. J., 2005. Drift and volatilities: Monetary policies and outcomes in the post WWII U.S. Review of Economic Dynamics 8 (2), 262–302.

Commandeur, J. J. F., Koopman, S. J., Ooms, M., 2011. Statistical software for state space methods. Journal of Statistical Software 41 (1), 1–18.

De Jong, P., 1991. The diffuse Kalman filter. The Annals of Statistics 19 (2), 1073–1083.

Diebold, F. X., Rudebusch, G. D., Aruoba, B. S., 2006. The macroeconomy and the yield curve: a dynamic latent factor approach. Journal of Econometrics 131 (1-2), 309–338.

Doran, H. E., 1992. Constraining Kalman filter and smoothing estimates to satisfy time-varying restrictions. The Review of Economics and Statistics 74 (3), 568–72.
Doucet, A., Godsill, S., Andrieu, C., 2000. On sequential Monte Carlo sampling methods for Bayesian filtering. Statistics and Computing 10 (3), 197– 208.
Doucet, A., Gordon, N., Krishnamurthy, V., 2001. Particle filters for state estimation of jump Markov linear systems. Signal Processing, IEEE Trans-actions on 49 (3), 613–624.
Doucet, A., Johansen, A. M., 2009. A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later. Oxford: Oxford University Press.
Durbin, J., Koopman, S. J., 2002. A simple and efficient simulation smoother for state space time series analysis. Biometrika 89 (3), 603-615.
Durbin, J., Koopman, S. J., 2012. Time Series Analysis by State Space Methods: Second Edition. Oxford: Oxford University Press.
Gordon, N., Salmond, D., Smith, A., 1993. Novel approach to nonlinear/nongaussian Bayesian state estimation. Radar and Signal Processing, IEE Proceedings F 140 (2), 107-113.
Greene, W. H., 2008. Econometric Analysis, Sixth Edition. New Jersey: Prentice Hall.
Gupta, N., Hauser, R., 2008. Kalman filtering with equality and inequality state constraints, Manuscript: http://arxiv.org/pdf/0709.2791.pdf.
Hull, J., 2003. Options, Futures and Other Derivatives, Fifth Edition. New Jersey: Prentice Hall.
Hull, J., White, A., 1990. Pricing interest-rate-derivative securities. Review of Financial Studies 3(4), 573–592.
Koop, G., Leon-Gonzalez, R., Strachan, R. W., 2010. Dynamic probabilities of restrictions in state space models: An application to the Phillips curve. Journal of Business & Economic Statistics 28 (3), 370–379.
Koop, G., Potter, S. M., 2011. Time varying VARs with inequality restrictions. Journal of Economic Dynamics and Control 35 (7), 1126–1138.
Liu, J. S., Chen, R., 1998. Sequential Monte Carlo methods for dynamic systems. Journal of the American Statistical Association 93 (443), 1032– 1044.
Pizzinga, A., 2012. Restricted Kalman Filtering, Theory, Methods, and Application. Springer: New York.

Primiceri, G. E., 2005. Time varying structural vector autoregressions and monetary policy. Review of Economic Studies 72 (3), 821–852.
Schon, T., Gustafsson, F., Nordlund, P., 2005. Marginalized particle filters for mixed linear/nonlinear state-space models. Signal Processing, IEEE Transactions on 53 (7), 2279–2289.
Simon, D. J., Simon, D. L., 2005. Aircraft turbofan engine health estimation using constrained Kalman filtering. Journal of Engineering for Gas Turbines and Power 127(2), 323–328.
Simon, D. J., Simon, D. L., 2010. Constrained Kalman filtering via density function truncation for turbofan engine health estimation. International Journal of Systems Science 41(2), 159–171.
Stock, J. H., Watson, M. W., 2007. Why has U.S. inflation become harder to forecast? Journal of Money, Credit and Banking 39 (s1), 3–33.
Tallis, G. M., 1961. The moment generating function of the truncated multi-normal distribution. Journal of the Royal Statistical Society. Series B (Methodological) 23 (1), 223-229.
Vasicek, O., 1977. An equilibrium characterization of the term structure. Journal of Financial Economics 5 (2), 177 – 188.

Appendix

Proof of Proposition 1

Conditional on x_{t-1}, the normalisation term $F(A_t x_{t-1}, Q_t, \mathcal{X}_t)$ can be treated as a constant. Thus

$$p(x_t|x_{t-1}, y_{1:t}) \propto \phi(x_t; A_t x_{t-1}, Q_t) \cdot \phi(y_t; C_t x_t, R_t) \cdot 1(x_t \in \mathcal{X}_t),$$

Since $\phi(x_t; A_t x_{t-1}, Q_t) \cdot \phi(y_t; C_t x_t, R_t) = \phi\left[\begin{pmatrix} x_t \\ y_t \end{pmatrix}; \begin{pmatrix} A_t x_{t-1} \\ C_t A_t x_{t-1} \end{pmatrix}, \begin{pmatrix} Q_t & Q_t C_t' \\ C_t Q_t & C_t Q_t C_t' + R_t \end{pmatrix}\right]$, we recognize that $x_t|x_{t-1}, y_{1:t} \sim TN(\mu_t, \Sigma_t, \mathcal{X}_t)$.

To compute the incremental weights $p(y_t|x_{t-1}, y_{1:t-1})$, we use the Bayes formula:

$$\begin{aligned} p(y_t|x_{t-1}, y_{1:t-1}) &= \frac{p(x_t, y_t|x_{t-1}, y_{1:t-1})}{p(x_t|x_{t-1}, y_{1:t})} \\ &= \frac{\phi(x_t; A_t x_{t-1}, Q_t) \cdot \phi(y_t; C_t x_t, R_t) \cdot \left[F\left(A_t x_{t-1}, B_t B_t', \mathcal{X}_t\right)\right]^{-1}}{\phi(x_t; \mu_t, \Sigma_t) \cdot [F(\mu_t, \Sigma_t, \mathcal{X}_t)]^{-1}} \\ &= \phi(y_t; C_t A_t x_{t-1}, C_t Q_t C_t' + R_t) \cdot \frac{F(\mu_t, \Sigma_t, \mathcal{X}_t)}{F\left(A_t x_{t-1}, B_t B_t', \mathcal{X}_t\right)}. \end{aligned}$$

Proof of Proposition 2

In order to prove Proposition 2, we will use a statistical property of the multivariate truncated normal distribution. Though the marginal distributions do not have analytic forms in general, an exception is the case in which a variable block is unrestricted. The results are summarized in the following lemma:

Lemma: suppose that $\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim TN\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}, \begin{pmatrix} \mathcal{Z}_1 \\ \mathbb{R}^n \end{pmatrix}\right]$. Then the marginal distribution for z_1 remains truncated normal: $z_1 \sim TN(\mu_1, \Omega_{11}, \mathcal{Z}_1)$, while the marginal distribution for z_2 is the extended skewed normal with the density $p(z_2) = \phi(z_2; \mu_2, \Omega_{22}) \cdot \frac{F(\bar{\mu}_1, \bar{\Omega}_{11}, \mathcal{Z}_1)}{F(\mu_1, \Omega_{11}, \mathcal{Z}_1)}$, where $\bar{\mu}_1 = \mu_1 + \Omega_{12}\Omega_{22}^{-1}(z_2 - \mu_2)$, $\bar{\Omega}_{11} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$.

Proof:

The normalisation term for the joint distribution of z_1 and z_2 satisfies:

$$F\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}, \begin{pmatrix} \mathcal{Z}_1 \\ \mathbb{R}^n \end{pmatrix}\right] = \int_{z_1 \in \mathcal{Z}_1} \int_{-\infty}^{\infty} \phi(z_1; \mu_1, \Omega_{11}) \phi(z_2; \bar{\mu}_2, \bar{\Omega}_{22}) dz_2 dz_1 = F(\mu_1, \Omega_{11}, \mathcal{Z}_1)$$

where $\bar{\mu}_2 = \mu_2 + \Omega_{21}\Omega_{11}^{-1}(z_1 - \mu_1)$, $\bar{\Omega}_{22} = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}$.

Thus the marginal densities for z_1 and z_2 are given by

$$p(z_1) = \int_{-\infty}^{\infty} \frac{\phi(z_1; \mu_1, \Omega_{11}) \phi(z_2; \bar{\mu}_2, \bar{\Omega}_{22}) \cdot 1(z_1 \in \mathcal{Z}_1)}{F(\mu_1, \Omega_{11}, \mathcal{Z}_1)} dz_2 = \frac{\phi(z_1; \mu_1, \Omega_{11}) \cdot 1(z_1 \in \mathcal{Z}_1)}{F(\mu_1, \Omega_{11}, \mathcal{Z}_1)},$$

$$p(z_2) = \int_{z_1 \in \mathcal{Z}_1} \frac{\phi(z_2; \mu_2, \Omega_{22}) \phi(z_1; \bar{\mu}_1, \bar{\Omega}_{11})}{F(\mu_1, \Omega_{11}, \mathcal{Z}_1)} dz_1 = \phi(z_2; \mu_2, \Omega_{22}) \cdot \frac{F(\bar{\mu}_1, \bar{\Omega}_{11}, \mathcal{Z}_1)}{F(\mu_1, \Omega_{11}, \mathcal{Z}_1)}. \quad \blacksquare$$

Now we start to prove Proposition 2. We will show that the optimal importance function $p(\xi_t | \xi_{1:t-1}, y_{1:t})$ follows a truncated normal distribution, while the incremental weight $p(y_t | \xi_{1:t-1}, y_{1:t-1})$ is an extended skewed normal density. By the preceding lemma, we only need to show $p(\xi_t, \eta_{1:t} | \xi_{1:t-1}, y_{1:t})$ and $p(y_t, \xi_t, \eta_{1:t} | \xi_{1:t-1}, y_{1:t-1})$ are multivariate truncated normal distributions.

$$p(\xi_t, \eta_{1:t} | \xi_{1:t-1}, y_{1:t})$$
$$\propto \prod_{\tau=1}^{t} p(\xi_\tau | \xi_{\tau-1}) p(\eta_\tau | \eta_{\tau-1}) p(y_\tau | \xi_\tau, \eta_\tau)$$

$\propto 1(\xi_t \in \Xi_t) \cdot \prod_{\tau=1}^{t} \phi(\xi_\tau; A_{1\tau}\xi_{\tau-1}, Q_{1\tau})\phi(\eta_\tau; A_{2\tau}\eta_{\tau-1}, Q_{2\tau})\phi(y_\tau; C_{1\tau}\xi_\tau + C_{2\tau}\eta_\tau, R_\tau)$

$\propto 1(\xi_t \in \Xi_t) \cdot \tilde{p}(\xi_t, \eta_{1:t} | \xi_{1:t-1}, y_{1:t})$

We have dropped the normalisation term $F(A_{1\tau}\xi_{\tau-1}, Q_{1\tau}, \Xi_\tau), \tau = 1, \dots, t$, as we are using the symbol $\propto$. The term $\tilde{p}(\xi_t, \eta_{1:t} | \xi_{1:t-1}, y_{1:t})$ denotes the conditional distribution in the context of a Gaussian linear state space model:

$\xi_t = A_{1t}\xi_{t-1} + \varepsilon_{1t}$,

$\eta_t = A_{2t}\eta_{t-1} + \varepsilon_{2t}$,

$y_t = C_{1t}\xi_t + C_{2t}\eta_t + v_t$,

where $\varepsilon_{1t} \sim N(0, Q_{1t})$, $\varepsilon_{2t} \sim N(0, Q_{2t})$, $v_t \sim N(0, R_t)$.

In a Gaussian linear model, all the marginal and conditional distributions are normal. So $\tilde{p}(\xi_t, \eta_{1:t} | \xi_{1:t-1}, y_{1:t})$ is a multivariate normal density. It follows that $p(\xi_t, \eta_{1:t} | \xi_{1:t-1}, y_{1:t})$ is a multivariate truncated normal distribution, in which $\eta_{1:t}$ is free while ξ_t is truncated to the region Ξ_t.

Similarly, we have

$p(y_t, \xi_t, \eta_{1:t} | \xi_{1:t-1}, y_{1:t-1}) \propto 1(\xi_t \in \Xi_t) \cdot \tilde{p}(y_t, \xi_t, , \eta_{1:t} | \xi_{1:t-1}, y_{1:t-1})$,

where $\tilde{p}(y_t, \xi_t, , \eta_{1:t} | \xi_{1:t-1}, y_{1:t-1})$ is the counterpart of $p(y_t, \xi_t, , \eta_{1:t} | \xi_{1:t-1}, y_{1:t-1})$ in the Gaussian linear state space model. Therefore, $p(y_t, \xi_t, \eta_{1:t} | \xi_{1:t-1}, y_{1:t-1})$ is a multivariate truncated normal density. By the lemma, $p(y_t, \xi_t | \xi_{1:t-1}, y_{1:t-1})$ remains multivariate truncated normal. Apply the lemma again, $p(y_t | \xi_{1:t-1}, y_{1:t-1})$ is an extended skewed normal density.

Now we show how to use the Kalman filter to partial out $\eta_{1:t}$ so as to obtain the filtering distributions.

First, conditioning on $\xi_{1:t-1}$, the predictive distribution for η_t are computed by the Kalman filter based on the GLSM. The predictive moments $\underline{\mu}_{\eta t}$ and $\underline{\Sigma}_{\eta t}$ represent the mean and variance for the Gaussian density $p(\eta_t | \xi_{1:t-1}, y_{1:t-1})$.

Second, consider the joint distribution $p(\xi_t, \eta_t, y_t | \xi_{1:t-1}, y_{1:t-1})$, which is a multivariate truncated normal distribution:

$$TN\left[\begin{pmatrix} A_{1t}\xi_{t-1} \\ \underline{\mu}_{\eta t} \\ \underline{\mu}_{yt} \end{pmatrix}, \begin{pmatrix} Q_{1t} & 0 & Q_{1t}C_{1t}' \\ 0 & \underline{\Sigma}_{\eta t} & \underline{\Sigma}_{\eta t}C_{2t}' \\ C_{1t}Q_{1t} & C_{2t}\underline{\Sigma}_{\eta t} & \Sigma_{yt} \end{pmatrix}, \begin{pmatrix} \Xi_t \\ \mathbb{R}^{m_2} \\ \mathbb{R}^n \end{pmatrix}\right],$$

Third, we partial out η_t and obtain $p(\xi_t, y_t | \xi_{1:t-1}, y_{1:t-1})$, which is

$TN\left[\begin{pmatrix} A_{1t}\xi_{t-1} \\ \mu_{yt} \end{pmatrix}, \begin{pmatrix} Q_{1t} & Q_{1t}C'_{1t} \\ C_{1t}Q_{1t} & \Sigma_{yt} \end{pmatrix}, \begin{pmatrix} \Xi_t \\ \mathbb{R}^n \end{pmatrix}\right].$

Conditioning on y_t, we obtain $p(\xi_t|\xi_{1:t-1}, y_{1:t})$, which follows $TN(\mu_{\xi t}, \Sigma_{\xi t}, \Xi_t)$. As for the incremental importance weight $p(y_t|\xi_{1:t-1}, y_{1:t-1})$, we apply the lemma and obtain Eq (8).

Proof of Proposition 3

For a linear state space model with deterministic initial states, the filtered states are linear with respect to the initial states. See De Jong (1991). The augmented Kalman filter has the property:

$E(x_t|x_S, y_{S+1:t}) = X_{t|t} \cdot \begin{pmatrix} 1 \\ x_S \end{pmatrix},$

$Var(x_t|x_S, y_{S+1:t}) = P_{t|t}.$

By the law of iterated expectations, we have

$E(x_t|y_{1:t}) = E[E(x_t|x_S, y_{S+1:t})|y_{1:t}] = X_{t|t} \cdot \begin{bmatrix} 1 \\ E(x_S|y_{1:t}) \end{bmatrix}.$

For the initial state smoothing, we use the fact that the prediction errors of the observations are also linear with respect to the initial states.

$p(x_S|y_{1:t}) \propto p(x_S|y_{1:S}) \cdot \prod_{\tau=S+1}^{t} \phi\left[V_\tau \cdot \begin{pmatrix} 1 \\ x_S \end{pmatrix}; 0, O_{\tau|\tau-1}\right].$

Recall that $p(x_S|y_{1:S})$ is represented by particles $x_S^{(i)}$ with weights $\underline{w}_S^{(i)}, i = 1, \dots, K$. By the Bayes formula, the smoothing distribution $p(x_S|y_{1:t})$ can be represented by the same particles with the updated weights $\overline{w}_S^{(i)}$ such that

$\overline{w}_S^{(i)} \propto \underline{w}_S^{(i)} \cdot \prod_{\tau=S+1}^{t} \phi\left[V_\tau \cdot \begin{pmatrix} 1 \\ x_S \end{pmatrix}; 0, O_{\tau|\tau-1}\right].$

The likelihood function has the following decomposition form:

$p(y_{1:t}) = p(y_{1:S}) \cdot \int p(x_S|y_{1:S}) \cdot p(y_{S+1:t}|x_S) dx_S.$

Note that $\frac{1}{K}\sum_{i=1}^{K} \underline{w}_S^{(i)}$ is a consistent estimator for $p(y_{1:S})$, which is a basic result of particle filtering.

$\int p(x_S|y_{1:S}) \cdot p(y_{S+1:t}|x_S) dx_S = \sum_{i=1}^{K} \left\{ \frac{\underline{w}_S^{(i)}}{\sum_{j=1}^{K} \underline{w}_S^{(i)}} \cdot \prod_{\tau=S+1}^{t} \phi\left[V_\tau \cdot \begin{pmatrix} 1 \\ w_S^{(i)} \end{pmatrix}; 0, O_{\tau|\tau-1}\right] \right\}.$

Hence Eq (13) is a consistent estimator for $p(y_{1:t})$.

Proof of Proposition 4

Invert the state equation, we have $x_{1:t} = A^{-1}c_0 + A^{-1}\varepsilon_{1:t}$, hence $p(x_{1:t}) = \phi\left(x_{1:t}; \underline{\mu}_{1:t}, \underline{\Sigma}_{1:t}\right)$.

Since $z_{1:t} = 1(x_{1:t} \in \mathcal{X}_{1:t})$, we have $p(x_{1:t}|z_{1:t} = 1) \propto \phi\left(x_{1:t}; \underline{\mu}_{1:t}, \underline{\Sigma}_{1:t}\right) \cdot 1(x_{1:t} \in \mathcal{X}_{1:t})$.

Also, $\binom{x_{1:t}}{y_{1:t}} \sim N\left[\begin{pmatrix} \underline{\mu}_{1:t} \\ C\underline{\mu}_{1:t} \end{pmatrix}, \begin{pmatrix} \underline{\Sigma}_{1:t} & \underline{\Sigma}_{1:t}C' \\ C\underline{\Sigma}_{1:t} & C\underline{\Sigma}_{1:t}C' + R \end{pmatrix}\right]$, hence the conditional distribution $p(x_{1:t}|y_{1:t}, z_{1:t}) \propto \phi\left(x_{1:t}; \overline{\mu}_{1:t}, \overline{\Sigma}_{1:t}\right) \cdot 1(x_{1:t} \in \mathcal{X}_{1:t})$.

Proof of Proposition 5

Note that $p(x_t|x_{t-1}, y_{1:t}, z_{1:t})$ is the state filtering distribution corresponding to a single-period PCSSM with the deterministic initial state x_{t-1}. Applying Proposition 4, we obtain $p(x_t|x_{t-1}, y_{1:t}, z_{1:t}) \propto \phi(x_t; \mu_t, \Sigma_t) \cdot 1(x_t \in \mathcal{X}_t)$.

The incremental weights can be decomposed as

$p(y_t, z_t|x_{t-1}, y_{1:t-1}, z_{1:t-1}) = p(y_t|x_{t-1}, y_{1:t-1}, z_{1:t-1}) \cdot p(z_t|x_{t-1}, y_{1:t}, z_{1:t-1})$.

Since $y_t = C_t(A_t x_{t-1} + \varepsilon_t) + v_t$, we have $p(y_t|x_{t-1}, y_{1:t-1}, z_{1:t-1}) = \phi(y_t; C_t A_t x_{t-1}, C_t Q_t C_t' + R_t)$.

Recall that $z_t = 1, \forall t$, hence

$p(z_t|x_{t-1}, y_{1:t}, z_{1:t-1}) = P(x_t \in \mathcal{X}_t|x_{t-1}, y_{1:t}) = F(\mu_t, \Sigma_t, \mathcal{X}_t)$.

Proof of Proposition 6

Conditioning on $\xi_{1:t-1}$, the normalisation term is a constant for both ICSSM and PCSSM. Therefore, the proof for Proposition 2 applies, and the optimal importance function is truncated normal: $p(\xi_t|\xi_{1:t-1}, y_{1:t}, z_{1:t}) \propto \phi\left(\xi_t; \mu_{\xi t}, \Sigma_{\xi t}\right) \cdot 1(\xi_t \in \Xi_t)$. As for the incremental importance weights, it has the decomposition form:

$$p(y_t, z_t | \xi_{1:t-1}, y_{1:t-1}, z_{1:t-1}) = p(y_t | \xi_{1:t-1}, y_{1:t-1}, z_{1:t-1}) \cdot p(z_t | \xi_{1:t-1}, y_{1:t}, z_{1:t-1}),$$

where $p(y_t | \xi_{1:t-1}, y_{1:t-1}, z_{1:t-1}) = \phi(y_t; \mu_{yt}, \Sigma_{yt})$ and $p(z_t | \xi_{1:t-1}, y_{1:t}, z_{1:t-1})$ equals the truncation probability $F(\mu_{\xi t}, \Sigma_{\xi t}, \Xi_t)$, as $z_t = 1$.

Table 1

	BT	PF	TS	OP
1969:Q3	0.9338	0.9338	0.9337	0.9337
	(3.6e-03)	(1.8e-03)	(9.9e-15)	(8.6e-15)
1974:Q4	0.9746	0.9749	0.9751	0.9751
	(4.3e-03)	(2.0e-03)	(1.8e-03)	(1.7e-03)
1980:Q2	0.9748	0.9752	0.9751	0.9753
	(3.7e-03)	(1.5e-03)	(1.8e-03)	(1.7e-03)
2001:Q1	0.8617	0.8617	0.8617	0.8617
	(2.8e-03)	(2.5e-03)	(5.1e-05)	(2.5e-06)
2009:Q1	0.9872	0.9873	0.9872	0.9872
	(1.0e-03)	(6.9e-04)	(7.7e-04)	(7.2e-04)
logL	-54.1367	-53.9149	-53.9329	-53.8892
	(7.7e-01)	(4.8e-01)	(4.0e-01)	(3.7e-01)
CPU Time	0.1005	0.1500	0.1155	0.1168
	(5.9e-03)	(6.2e-03)	(5.7e-03)	(2.8e-03)

Table 1 Particle filter for the time-varying parameter AR(2) model

BT refers to the bootstrap particle filter. PF adopts the optimal importance function without Rao-Blackwellisation. TS utilizes temporal Rao-Blackwellisation and the optimal importance function. OP is the optimal filter that employs all the variance reduction techniques. Each algorithm is experimented 500 times and the average state estimators for $\phi_{2t} + \phi_{1t}$, the log likelihood and the computing time are reported. The standard deviations are in parentheses, which reflect the Monto Carlo variations.

Figure 1

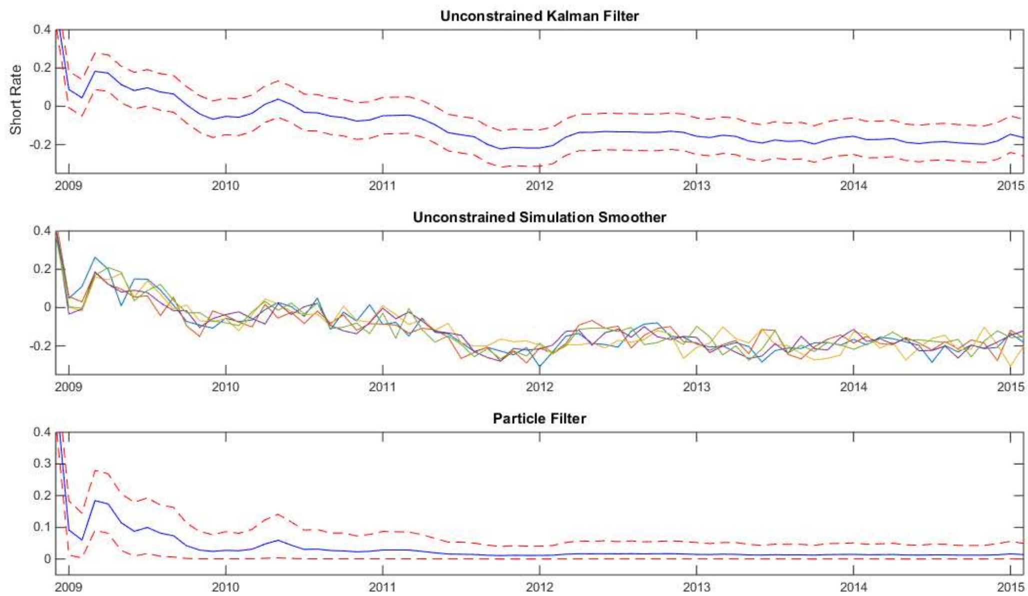

Figure 1 Short rate estimation of the Vasicek Model

The upper panel plots the unconstrained Kalman filter estimation of the instantaneous interest rates. The solid line represents the means of the filtered state series and the two dotted lines are the 95% intervals of the filtering distributions. The middle panel illustrates five paths of the simulated posterior state series using the unconstrained simulation smoother. The bottom panel shows the particle filtering results with nonnegative constraints on the short rate series. The estimation sample is 2003 – 2015. To highlight the low interest rate era, the curves are plotted from 2009 to 2015.

Figure 2

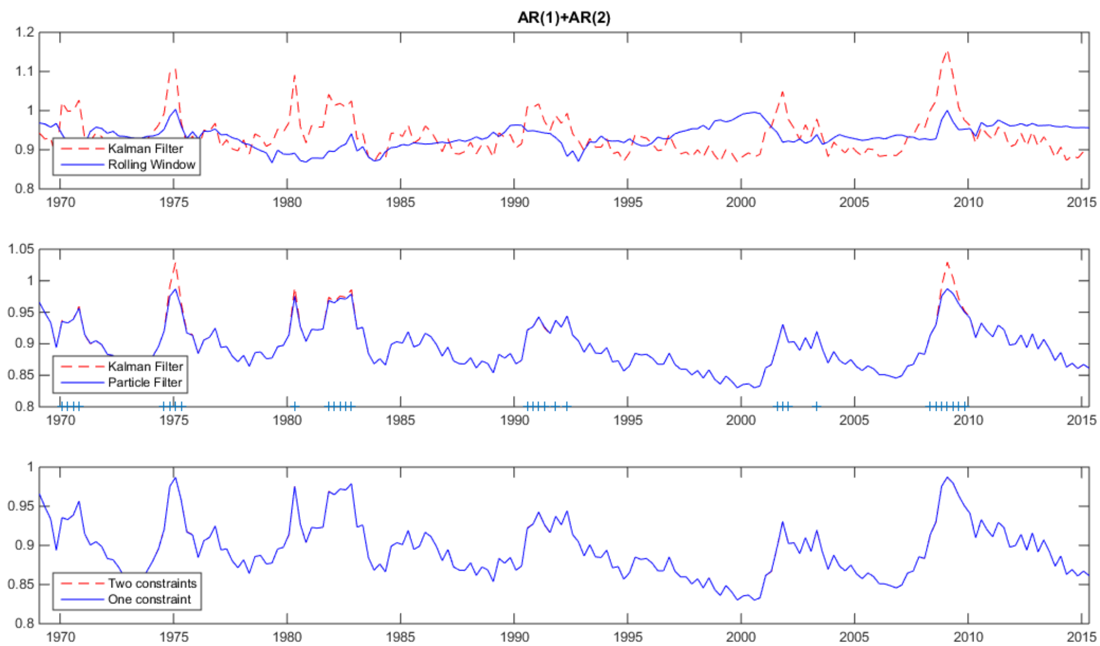

Figure 2 Time-varying parameter AR(2) regression

The estimated series $\phi_{2t} + \phi_{1t}$ using various methods are plotted. The upper panel shows the ten-year rolling window AR(2) regression results (solid line) and the unconstrained Kalman filter (dashed line) calibrated by the rolling window estimators. The middle panel plots the particle filtering results (solid line) under the constraint $\phi_{2t} + \phi_{1t} \leq 1$ imposed on 31 out of 186 periods (marked on the horizontal axis). The model parameters are estimated by maximum simulated likelihood. The unconstrained Kalman filter (dashed line) using those parameters is also plotted. The bottom panel illustrates the case when a loose constraint $\phi_{2t} \geq -1$ is added. The filtered series under one and two constraints overlap with each other, with the maximum discrepancy 0.0008.

Figure 3

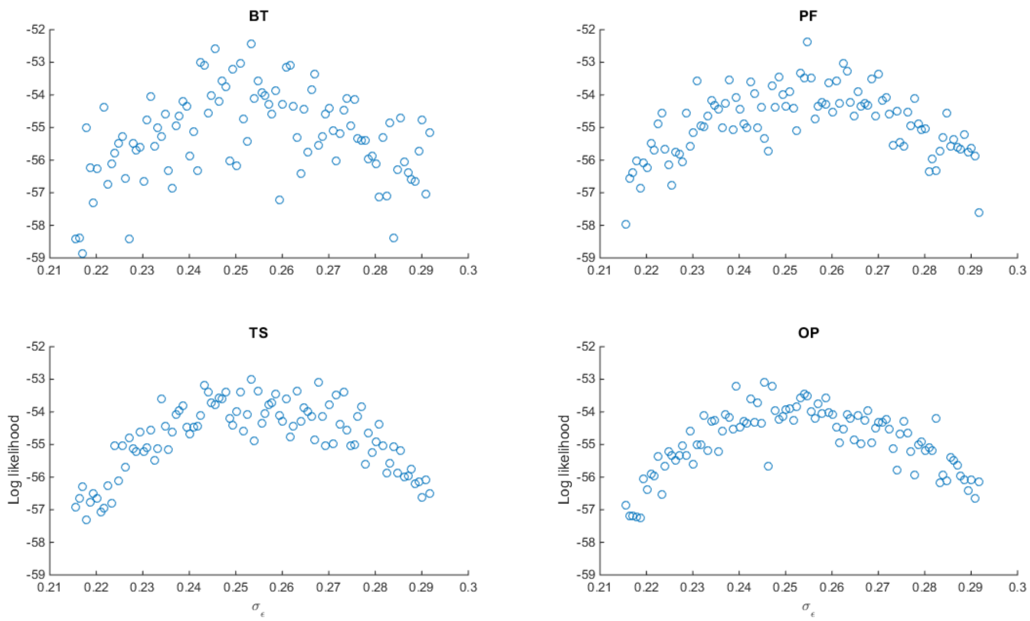

Figure 3 Grid search for the parameter σ_ε that maximizes the noisy likelihood function

We discretize σ_ε by 100 grid points and evaluate the log likelihood using the four particle filtering algorithms. BT refers to the bootstrap particle filter. PF adopts the optimal importance function without Rao-Blackwellisation. TS utilizes temporal Rao-Blackwellisation and the optimal importance function. OP is the optimal filter that employs of all the variance reduction techniques.