Nonlinear Filtering and Robust Learning *

Lars Peter Hansen
email:lhansen@uchicago

Nick Polson
email:ngp@chicagogsb.edu

Thomas J. Sargent
email:thomas.sargent@nyu.edu

October 20, 2010

Abstract

We refine and apply particle filtering algorithms that emphasize portions of the filtered distribution most pertinent to a decision maker. We illustrate our algorithms in an equilibrium model with investors who design robust decision rules by exponentially tilting probabilities of a baseline statistical model.

1 Introduction

This paper develops nonlinear filtering methods for the purpose of building models in which decision makers simultaneously learn and choose, even though they possibly doubt their hidden state Markov model. Particular filtering methods, for example, Kalman filtering and Wonham filtering, have tractable closed-form recursive representations but apply to only a limited range of problems. That situation leads us to explore numerical alternatives that can potentially expand the class of tractable decision problems.

We devote particular attention to decision problems in which some of the hidden states are really parameters and hence invariant over time. We describe, justify, and apply a modified version of particle filtering designed to make parameter learning tractable. Models with parameter learning require special attention because particle filtering simulations typically recurrently revisit the same parameter value. We suggest a less wasteful way to use particles. Our approach is to introduce artificial movements in parameters into the simulations to be used for filtering. In particular, in a simulated time path associated with a particular particle, we make parameters evolve probabilistically over time, but in a way that does not distort the densities that are the targets that we seek to approximate. Our approach partly follows those of Whittle (1969), Storvik (2002), and Fearnhead (2002) by exploiting the existence of sufficient statistics conditioned on data on signals and hidden states. Thus, our definition of a particle will include the vector of hidden states as well as a vector of sufficient statistics for the parameter of interest, conditioned on the hidden states. We differ from Storvik (2002)

*Preliminary and incomplete.

and Fearnhead (2002) in how we make the particle evolve more efficiently. Johannes and Polson (2007) and Carvalho et al. (2010) apply this method in a variety of settings. Here we provide extensions that allow us to explore applications in which individuals learn over time while recognizing their doubts about their statistical specification.

We apply our methods to a model in which investors make decisions that are robust to a set of model misspecification. Such investors exponentially twist continuation values to form a likelihood ratio that they use to adjust probabilities conservatively. Hansen and Sargent (2007) provide a recursive formulation that explicitly incorporates learning by confronting decision makers with a hidden state Markov process in which the state can include unknown parameters. Hansen and Sargent (2010) and Hansen (2007) apply this approach, but only in models for which there are quasi-analytic formulas for the filtering problem. Extending things beyond those special models requires that the numerical filtering algorithm be adapted to produce accurate approximations in regions of the space of parameters and other hidden states that most concern the decision maker. Details of the decision problem, including whether and how much the decision maker doubts his statistical specification, matter in designing a good numerical approximation algorithm. Concerns about robustness inspire our agents to explore tails of distributions, a fact that frames numerical challenges for filtering algorithms. Thurn et al. (2002) suggest using risk functions in conjunction with particle filtering methods in order to focus numerical accuracy on the most important components of uncertainty. We use a similar idea but modify it to suit the economic examples that most interest us.

DeJong et al. (2010) provide both a comprehensive survey of particle filtering methods and develop their own refinements of such methods. Because we focus on the decision problems of private economic agents engaged in learning rather than the problem of econometricians who seek to maximize a likelihood for a given data set, we want to explore different questions than DeJong et al. (2010).

2 Hidden state Markov chain

Let X_t denote an underlying Markov state at date t and let Y_t be a vector of current period signals. We allow a decision maker to observe a component D_t of the state vector. We partition the state vector:

$$X_t = \begin{bmatrix} D_t \\ Z_t \end{bmatrix}.$$

We capture the observability of D_t through a functional relation $D_{t+1} = \phi(Y_{t+1}, D_t)$. The decision-maker does not observe Z_t, nor does he observe a parameter vector θ that governs the dynamic evolution of states and signals.

Let $\mathbb{Z}$ denote a locally compact metric space of admissible hidden states, $\mathcal{B}(\mathbb{Z})$ the Borel sigma algebra of $\mathbb{Z}$, and λ a measure on the measurable space of hidden states $(\mathbb{Z}, \mathcal{B}(\mathbb{Z}))$. Let $\Gamma(dz^*|y^*, x, \theta)$ denote the conditional distribution of Z_{t+1} conditioned on Y_{t+1}, X_t, and θ, where θ is a parameter that resides in a measurable space $(\Theta, \mathcal{B}(\Theta))$. We will sometimes find it convenient to assume that the conditional probability measure Γ is absolutely continuous with respect to λ with density γ:

$$\Gamma(dz^*|y^*, x, \theta) = \gamma(z^*|y^*, x, \theta)\lambda(dz^*).$$

Let π be a measure on $(\Theta, \mathcal{B}(\Theta))$; in particular, we can think of it as a *prior* probability measure for θ. While θ could be viewed as a time invariant hidden component of the state z, we will have cause to treat it separately. As we shall see, in designing simulation algorithms to use in approximation, the distinction between a parameter that is invariant over time and an evolving hidden state is important. Finally, let $\mathbb{Y}$ denote a locally compact metric space of admissible signals, $\mathcal{B}(\mathbb{Y})$ the corresponding Borel sigma algebra, and η a measure on the measurable space $(\mathbb{Y}, \mathcal{B}(\mathbb{Y}))$ of signals. Let $\nu(y^*|x, \theta)\eta(dy^*)$ denote the distribution of Y_{t+1} conditioned on X_t and θ. Notice that we specify the hidden state Markov chain using a convenient factorization of the joint distribution for X_{t+1}, Y_{t+1} conditioned on (X_t, θ). If next period signals are conditionally independent of the next period state vector, then $\Gamma(dz^*|y^*, x, \theta)d\lambda(z^*)$ does not depend on y^*.

Let $\mathcal{Y}_t$ denote the signal history including initial conditions for the observable state vector. Thus $\mathcal{Y}_t$ is generated by $D_0, Y_1, ..., Y_t$. Evidently, D_t can be constructed as a function of $\mathcal{Y}_t$, and Y_{t+1} combined with $\mathcal{Y}_t$ generates the same sigma algebra as $\mathcal{Y}_{t+1}$. Let $\mathcal{X}_t$ denote the combined history of the states and signals; $\mathcal{X}_t$ generates a larger sigma algebra than $\mathcal{Y}_t$. The aim of filtering is to compute a sequence of probability measures:

$$q_t(z, \theta)\lambda(dz)\pi(d\theta)$$

for the date t hidden state, Z_t, and the unknown parameter vector θ, conditioned on $\mathcal{Y}_t$. The predictive density for Y_{t+1} given $\mathcal{Y}_t$ is:

$$\ell_t(y^*|\mathcal{Y}_t) = \int \nu(y^*|d, z, \theta)q_t(z, \theta)\lambda(dz)\pi(d\theta).$$

3 Partial smoothing

Filtering means estimating a hidden state or parameter at time t using an information set $\mathcal{Y}_t$ observed at time t. *Smoothing* means estimating a hidden state or parameter at time t

using an information set available at a later date $s > t$. When feasible, it is advantageous at least partially to smooth our estimates of the states prior to simulation, where 'partially' means that the information used is for a date s satisfying $t < s < +\infty$.

For example, we construct a new state vector

$$\hat{X}_t \doteq \begin{bmatrix} Y_t \\ X_{t-1} \end{bmatrix}.$$

The unobservable component for the new state vector is Z_{t-1}, whereas earlier it had been Z_t. The signal vector remains as it was before.

Construct the distribution of the signal conditioned on $\hat{X}_t$:

$$\hat{\eta}(y^*|\hat{X}_t, \theta) = \int \eta(y^*|D_t, z, \theta)\Gamma(dz|Y_t, X_{t-1}, \theta)$$

where we integrate out Z_t because it is not part of the current value of the newly constructed state $\hat{X}_t$. In this new construction, Z_t becomes the hidden component of the next period state vector $\hat{X}_{t+1}$. The new state evolution conditioned on the signal Y_{t+1} is

$$\hat{\Gamma}(dz|Y_{t+1}, \hat{X}_t, \theta) = \frac{\eta(Y_{t+1}|D_t, z, \theta)\Gamma(dz|Y_t, X_{t-1}, \theta)}{\int \eta(Y_{t+1}|D_t, \tilde{z}, \theta)\Gamma(d\tilde{z}|Y_t, X_{t-1}, \theta)}$$

in conjunction with the functional relation $D_t = \phi(Y_t, D_{t-1})$. Since Y_{t+1} is both a signal and a component of the next period state, part of the evolution of the next period state given the next period signal is degenerate by construction.

If feasible, these "look forward" calculations are valuable to exploit in simulation.

4 A simulator based on sufficient statistics

Following Johannes and Polson (2007) and Carvalho et al. (2010), we explore an alternative way to construct particles in a particle filtering algorithm that uses sufficient statistics. It is convenient to construct these sufficient statistics by using both the hidden and observed states. In the data generation for the actual time series, θ does not change over time. In standard approaches to particle filtering, this invariance is imposed; but we now alter the implicit time series evolution that underlies the particle filtering algorithm while preserving the filtering outcome. In this section, we construct a new Markov process that uses a vector of sufficient statistics as a state vector. While different from the original process, this new process has the same distribution of the hidden states and parameter vector θ conditional on the observation history $\mathcal{Y}_t$.

Consider a state vector S_t that, given an initial condition S_0, is constructed recursively via:

$$S_{t+1} = \Phi(\theta, Y_{t+1}, Z_{t+1}, X_t, S_t).$$

This recursion will be useful in forming simulations provided that

Assumption 4.1. *The distribution of θ conditioned on S_t, Z_t, and the signal history $\mathcal{Y}_t$ satisfies*

$$\rho_t(d\theta|S_t, Z_t, \mathcal{Y}_t) = \psi(d\theta|S_t).$$

for some prespecified ψ.

Some special cases are revealing. For example,

Condition 4.2.

$$S_t = \theta$$

In this case $\psi(d\tilde{\theta}|S_t)$ assigns probability one to the single value θ. This can lead to a very inefficient algorithm.

Alternatively, the following condition describes a situation in which there is a set of sufficient statistics S_t for the distribution of θ, given the state vector and signal history as well as an appropriately restricted prior. When feasible, this make the use of particle more efficient.

Condition 4.3. *There exists a vector S_t such that*

$$\rho_t(d\theta|S_t, Z_t, \mathcal{Y}_t) = \psi(d\theta|S_t)$$

where S has a recursive representation

$$S_{t+1} = \Phi(Y_{t+1}, Z_{t+1}, X_t, S_t)$$

for $t = 0, 1, ...$ for some choice of S_0 such that the prior $\pi(d\theta) = \psi(d\theta|S_0)$.

Remark 4.4. *Evidently, the entire state vector, including its hidden components, can be used to form sufficient statistics. The distribution ρ_{t+1} can always be represented recursively using a fictitious posterior for θ conditioned on $\mathcal{X}_t$, denoted $\varrho_t(d\theta|\mathcal{X}_t)$. In particular, suppose that $\Gamma(dz^*|y^*, x, \theta) = \gamma(z^*|y^*, x, \theta)\lambda(dz^*)$. Then*

$$\varrho_{t+1}(d\theta|\mathcal{X}_{t+1}) = \frac{\gamma(Z_{t+1}|Y_{t+1}, X_t, \theta)\nu(Y_{t+1}|X_t, \theta)\varrho_t(d\theta|\mathcal{X}_t)}{\int \gamma(Z_{t+1}|Y_{t+1}, X_t, \theta)\nu(Y_{t+1}|X_t, \theta)\varrho_t(d\theta|\mathcal{X}_t)}$$

so that $\varrho_t(d\theta|\mathcal{X}_t)$ could be used as S_t. Typically, S_t constructed in this way would be infinite dimensional, rendering it impractical for our purposes. What we want are finite-dimensional *sufficient statistics.*

Sometimes it will be enough that we can partition the parameter vector θ into two sets of components and then to have sufficient statistics for one only set of components but not for the other. It may still be possible to satisfy assumption 4.1 with a recursive updating equation, albeit possibly one that has degeneracies in the sense that a component of S_t will be the component of θ for which we do not use a sufficient statistic. This component of S_t will be invariant by construction.

In applications, it is often the case that we are interested in further restricting the parameter space. Suppose that absent such restrictions, Assumption 4.1 is satisfied. Now impose an additional prior restriction on the parameter θ of the form $\theta \in \Theta$ where the set Θ is prespecified. Recall the restriction:

$$\rho_t(d\theta|S_t, Z_t, \mathcal{Y}_t) = \psi(d\theta|S_t)$$

Form

$$\bar{\rho}_t(d\theta|S_t, Z_t, \mathcal{Y}_t) = \frac{\rho_t(d\theta|S_t, Z_t, \mathcal{Y}_t)}{\int_\Theta \rho_t(d\vartheta|S_t, \mathcal{Y}_t)d\pi(\vartheta)} = \frac{\psi(d\theta|S_t)}{\int_\Theta \psi(d\vartheta|S_t)d\pi(\vartheta)} = \bar{\psi}(d\theta|S_t),$$

which asserts the well known result that if S_t is sufficient for θ without the restriction on the parameter space, it remains sufficient given this restriction.

Let κ_t be the distribution of (Z_t, S_t) conditioned on $\mathcal{Y}_t$. We now describe a recursive method of constructing κ_{t+1} conditioned on $\mathcal{Y}_{t+1}$. Note that

$$\begin{aligned}\varsigma_{t+1}(dz^*, dz, ds, d\theta) =& \frac{\Gamma(dz^*|Y_{t+1}, D_t, z, \theta)\nu(Y_{t+1}|D_t, z, \theta)\rho_t(d\theta|z, s, \mathcal{Y}_t)\kappa_t(dz, ds)}{\int \Gamma(dz^*|Y_{t+1}, D_t, z, \theta)\nu(Y_{t+1}|D_t, z, \theta)\rho_t(d\theta|z, s, \mathcal{Y}_t)\kappa_t(dz, ds)} \\ & \frac{\Gamma(dz^*|Y_{t+1}, D_t, z, \theta)\nu(Y_{t+1}|D_t, z, \theta)\rho_t(d\theta|z, s, \mathcal{Y}_t)\lambda(dz^*)\kappa_t(dz, ds)}{\int \nu(Y_{t+1}|D_t, z, \theta)\rho_t(d\theta|z, s, \mathcal{Y}_t)\lambda(dz^*)\kappa_t(dz, ds)}\end{aligned} \tag{1}$$

gives the joint distribution for $Z_{t+1}, Z_t, S_t, \theta$ conditioned on $\mathcal{Y}_{t+1}$. Since S_{t+1} is a known function of $Z_{t+1}, Y_{t+1}, X_t, S_t, \theta$, we can use this distribution to deduce the distribution of Z_{t+1}, S_{t+1} conditioned on $\mathcal{Y}_{t+1}$. To see this, consider any bounded Borel measurable function f of (z, s). Then

$$E\left[f(Z_{t+1}, S_{t+1})|\mathcal{Y}_{t+1}\right] = \int f[z^*, \Phi(\theta, Y_{t+1}, z^*, D_t, z, s)]\varsigma_{t+1}(dz^*, dz, ds, d\theta).$$

Because this formula applies for all such f, it implies a distribution κ_{t+1} for Z_{t+1}, S_{t+1} conditioned on $\mathcal{Y}_{t+1}$. To obtain the joint distribution of $(\theta, Z_{t+1}, S_{t+1})$ conditioned on $\mathcal{Y}_{t+1}$, we form

$$\rho_{t+1}(d\theta|s^*, z^*, \mathcal{Y}_{t+1})\kappa_{t+1}(dz^*, ds^*).$$

For purposes of simulation, we construct a new Markov process with an expanded state vector $(\tilde{X}_t, \tilde{S}_t)$. The $\tilde{\cdot}$ notation is used because $(\tilde{X}_t, \tilde{S}_t)$ has a different stochastic evolution than does (X_t, S_t). Under this new evolution, the parameter vector θ ceases to be time invariant, which allows particles to regenerate. This regeneration is helpful in building a simulation algorithm that works well over time. Since the transition distribution for $\tilde{\theta}_{t+1}$ will degenerate eventually as the parameter is learned with more and more precision as $t+1$ grows, the computational advantages of this randomness will eventually disappear for large t. Under the new Markov law, the state vector evolves as follows:

$$\tilde{\theta}_{t+1} \sim \psi(\cdot|\tilde{S}_t)$$

$$\tilde{Y}_{t+1} \sim \nu\left(\cdot|\tilde{X}_t, \tilde{\theta}_{t+1}\right) \tag{2}$$

$$\tilde{Z}_{t+1} \sim \Gamma(\cdot|\tilde{Y}_{t+1}, \tilde{X}_t, \tilde{\theta}_{t+1}) \tag{3}$$

$$\tilde{S}_{t+1} = \Phi\left(\tilde{\theta}_{t+1}, \tilde{Y}_{t+1}, \tilde{Z}_{t+1}, \tilde{X}_t, \tilde{S}_t\right). \tag{4}$$

By using (2), we have altered the joint process for the state and signal relative to that given in section 2; but we have left unaltered the conditional distributions that are the targets of our simulator. We can hit these targets using the signals recorded in the data in conjunction with simulations based on (2).

Proposition 4.5. *Under Assumption 4.1, the $\tilde{\cdot}$ Markov process shares κ_t with the original process as the conditional distribution for $(\tilde{Z}_t, \tilde{S}_t)$ conditioned on $\tilde{\mathcal{Y}}_t$, constructed from the history of observations on $\tilde{Y}_t$ from the $\tilde{\cdot}$ Markov process, provided that the same prior over θ and the same initialization of he sufficient statistic vector are used.*

Proof. We prove this by induction. Supposing that the result holds for κ_t, we establish the result for κ_{t+1}. Substituting the restriction given in Assumption 4.1 into (1) gives

$$\varsigma_{t+1}(dz^*, dz, ds, d\theta) = \frac{\Gamma(dz^*|Y_{t+1}, D_t, z, \theta)\nu(Y_{t+1}|D_t, z, \theta)\psi(d\theta|s)\kappa_t(dz, ds)}{\int \nu(Y_{t+1}|D_t, z, \theta)\psi(d\theta|s)\kappa_t(dz, ds)}. \tag{5}$$

The numerator of this formula motivates our construction of the $\tilde{\cdot}$ process. At date t first generate $\tilde{\theta}_{t+1}$ conditioned on $\tilde{S}_t$ by drawing from ψ, next generate $\tilde{Y}_{t+1}$ conditioned on $\tilde{X}_t$ and $\tilde{\theta}_{t+1}$ by drawing from ν, and finally generate $\tilde{Z}_{t+1}$ conditioned on $\tilde{Y}_{t+1}$, $\tilde{X}_t$ and $\tilde{\theta}_{t+1}$ by drawing from Γ. Formula (5) then conditions on $\tilde{Y}_{t+1}$ in addition to $\tilde{\mathcal{Y}}_t$. As before, we impute κ_{t+1} from

$$E\left[f(\tilde{Z}_{t+1}, \tilde{S}_{t+1})|\tilde{\mathcal{Y}}_{t+1}\right] = \int f[z^*, \Phi(\tilde{\theta}_{t+1}, \tilde{Y}_{t+1}, z^*, D_t, z, s)]\varsigma_{t+1}(dz^*, dz, ds, d\theta),$$

which holds for any Borel measurable function f. Therefore, κ_{t+1} is the same for our $\tilde{\cdot}$ construction. □

We now show how to use a simulation method to construct the particle filtering solution for this alternative Markov process.

Algorithm 4.6. *At date t there are N particles. A particle is specified as $(\tilde{Z}_t^{[i]}, \tilde{S}_t^{[i]})$ where $^{[i]}$ indexes a particle.*

1. *Draw $\tilde{\theta}_{t+1}^{[i]}$ from $\psi\left(\theta|\tilde{S}_t^{[i]}\right)$.*
2. *Construct weights*

$$w_{t+1}^{[i]} = \frac{\nu\left(Y_{t+1}|D_t, \tilde{Z}_t^{[i]}, \tilde{\theta}_{t+1}^{[i]}\right)}{\sum_i^N \nu\left(Y_{t+1}|D_t, \tilde{Z}_t^{[i]}, \tilde{\theta}_{t+1}^{[i]}\right)}.$$

   Sample from the weighted empirical distribution. Draw $(\tilde{Z}_t^{[i]}, \tilde{S}_t^{[i]}, \tilde{\theta}_{t+1}^{[i]})$ from a multinomial distribution with probability $w_{t+1}^{[i]}$.

3. *Draw $\tilde{Z}_{t+1}^{[i]}$ from $\Gamma\left(dz^*|Y_{t+1}, D_t, Z_t^{[i]}, \tilde{\theta}_{t+1}^{[i]}\right)$.*
4. *Construct $\tilde{S}_{t+1}^{[i]} = \Phi\left(\tilde{S}_{t+1}^{[i]}, Y_{t+1}, Z_{t+1}^{[i]}, D_t, \tilde{X}_t^{[i]}, \tilde{\theta}_{t+1}^{[i]}\right)$.*
5. *Replace particle $(\tilde{Z}_t^{[i]}, \tilde{S}_t^{[i]})$ with $(\tilde{Z}_{t+1}^{[i]}, \tilde{S}_{t+1}^{[i]})$.*

By including the additional randomness to regenerate parameters θ, this algorithm avoids problems with standard particle filtering methods.

Under the particle approximation, the predictive density for Y_{t+1} is:

$$\ell_t(y^*|\mathcal{Y}_t) \approx \frac{1}{N}\sum_{i=1}^{N} \nu(y^*|D_t, \tilde{Z}_t^{[i]}, \tilde{\theta}_{t+1}^{[i]})$$

where the θ's come from the first step of the algorithm. The resulting estimate of the state density is:

$$\int q_{t+1}(z^*,\theta)\pi(d\theta) \approx \sum_{i=1}^{N} w_{t+1}^{[i]}\gamma\left(z^*|Y_{t+1}, D_t, \tilde{Z}_t^{[i]}, \tilde{\theta}_{t+1}^{[i]}\right).$$

The density for θ conditional on z^* has density

$$\frac{q_{t+1}(z^*,\theta)}{\int q_{t+1}(z^*,\theta)\pi(d\theta)} \approx \frac{1}{N}\sum_{i=1}^{N}\psi\left(\theta|\tilde{S}_{t+1}^{[i]}\right).$$

5 Utility-based Simulation

We consider utility-based adjustment for simulation. As argued by Thurn et al. (2002), the objective of a decision problem can be useful guide on where to focus the accuracy of the approximation. In what follows we show to use a dynamic objective to alter the evolution in way that supports the calculation of an implied price of uncertainty.

5.1 Recursive utility

Following Kreps and Porteus (1978) and Epstein and Zin (1989) form a process of continuation values:

$$\begin{aligned} V_t &= \left[(\zeta C_t)^{1-\vartheta} + \exp(-\delta)\left[\mathcal{R}_t(V_{t+1})\right]^{1-\vartheta}\right]^{\frac{1}{1-\vartheta}} \\ \mathcal{R}_t(V_{t+1}) &= \left(E\left[(V_{t+1})^{1-\omega}|\mathcal{X}_t\right]\right)^{\frac{1}{1-\omega}} \end{aligned} \tag{6}$$

where $\vartheta > 0$, and $\omega > 0$. For now we do calculation under the more complete information set to motivate a change of probability measure. The parameter ζ is a scale factor for the continuation value, and we will have cause to adjust this parameter in studying limiting cases. In the special case in which $\vartheta = 1$, the recursion as given is not well defined. It turns out by an appropriate scaling, we can represent preferences in this special case by using the Cobb-Douglas form:[1]

$$V_t = (\zeta C_t)^{1-\exp(-\delta)}\left[\mathcal{R}_t(V_{t+1})\right]^{\exp(-\delta)}. \tag{7}$$

Notice that recursion (6) maps next-periods continuation V_{t+1} and the current-period C_t into the current period continuation value V_t. Next we exploit homogeneity and write

$$\left(\frac{V_t}{C_t}\right)^{1-\vartheta} = (\zeta)^{1-\vartheta} + \exp(-\delta)\left(\mathcal{R}_t\left[\left(\frac{V_{t+1}}{C_{t+1}}\right)\left(\frac{C_{t+1}}{C_t}\right)\right]\right)^{1-\vartheta}$$

where we first divide by C_t and then we raise both sides to the power $1-\vartheta$. Finally we write

$$\exp\left[\frac{\delta(1-\omega)}{1-\vartheta}\right]\left[\left(\frac{V_t}{C_t}\right)^{1-\vartheta} - (\zeta)^{1-\vartheta}\right]^{\frac{1-\omega}{1-\vartheta}} = E\left[\left(\frac{V_{t+1}}{C_{t+1}}\right)^{1-\omega}\left(\frac{C_{t+1}}{C_t}\right)^{1-\omega}|\mathcal{X}_t\right]. \tag{8}$$

[1] Represent $[1-\exp(\delta)](\zeta^\star)^{1-\vartheta} = \zeta^{1-\vartheta}$ and take limits as ϑ tends to unity. This results in the Cobb-Douglas recursion with $\zeta^\star$ replacing ζ.

There are some noteworthy special cases. First suppose that $\omega = \vartheta$. In this case we obtain the expected utility recursion. Next suppose that $(\zeta)^{1-\vartheta}$ converges to zero and that $\exp\left[\frac{\delta(1-\omega)}{1-\vartheta}\right]$ converges to $\exp(\epsilon)$ defined to be the solution to the equation:

$$\exp(\epsilon)\left(\frac{V_t}{C_t}\right)^{1-\omega} = E\left[\left(\frac{V_{t+1}}{C_{t+1}}\right)^{1-\omega}\left(\frac{C_{t+1}}{C_t}\right)^{1-\omega}|\mathcal{X}_t\right]. \tag{9}$$

This choice of ϵ by design makes the future as "important as possible" and thus gives an interesting limiting case.

In what follows we construct solutions for $\frac{V_t}{C_t}$ for a given specification of the consumption dynamics. We suppose that $\log C_{t+1} - \log C_t$ is among the components of Y_{t+1} and we suppose initially that parameters and states are observed by the decision maker. This leads us to search for a solution of the form:

$$\log V_t - \log C_t = \upsilon(X_t, \theta).$$

The equations (8) and its limiting counterpart (9) now become "fixed point" equations for υ. In the case of (9), let

$$e(x,\theta) = \exp\left[(1-\omega)\upsilon(x,\theta)\right],$$

and rewrite equation (9) as

$$\exp(\epsilon)e(x,\theta) = E\left[e(X_{t+1},\theta)\left(\frac{C_{t+1}}{C_t}\right)^{1-\omega}|X_t = x, \theta\right]$$

which is an eigenvalue equation with a positive eigenfunction e and an eigenvalue $\exp(\epsilon)$ where ϵ depends on the state and consumption dynamics including the parameter vector θ.

5.2 A convenient change of measure

We follow Hansen and Scheinkman (2009) by using this positive eigenfunction to build a change in probability measure that preserves the Markov structure.[2] Notice in particular that (9) implies

$$\exp(-\epsilon)E\left[\frac{e(X_{t+1},\theta)}{e(X_t,\theta)}\exp\left[(1-\omega)(\log C_{t+1} - \log C_t)\right]|X_t = x, \theta\right] = 1.$$

The positive random variable

$$\frac{\hat{M}_{t+1}}{\hat{M}_t} = \exp(-\epsilon)\frac{e(X_{t+1},\theta)}{e(X_t,\theta)}\exp\left[(1-\omega)(\log C_{t+1} - \log C_t)\right]$$

is the Radon-Nikodym derivative for change in the transition density of the Markov process. Moreover, once we initialize $\hat{M}_0$ at say unity, we may build a so-called multiplicative martingale conditioned on the parameter vector θ.

[2] An analogous change of measure is used in the Markov theory for large deviations.

Associated with a multiplicative martingale is a change of probability measure, as is well known from the applied probability literature. Moreover, this change of measure preserves the Markov structure. Under the change of measure, the expectation conditioned on date zero information including the unknown parameter is:

$$\hat{E}(R_t|X_0,\theta) = E\left(\frac{\hat{M}_t}{\hat{M}_0} R_t | X_0, \theta\right)$$

for a random variable R_t that is $\mathcal{X}_t$ measurable.

The implied distortion for the filtering distribution is what is important for us. Let $\hat{q}_{t+1}$ denote the density for Z_{t+1} and θ conditioned on $\mathcal{Y}_{t+1}$ under this change of measure. Of course this density will depend on our initial distributions.

Proposition 5.1. *If the joint distorted prior $\hat{q}_0$ for Z_0, θ satisfies:*

$$\hat{q}_0(z,\theta)\lambda(dz)\pi(d\theta) \propto \exp(\epsilon\tau)e(D_0,z,\theta)q_0(z,\theta)\lambda(dz)\pi(d\theta),$$

for some τ, then

$$\exp(\epsilon t)\hat{q}_t(z,\theta) \propto \exp(\epsilon t)e(D_t,z,\theta)q_t(z,\theta)$$

for all $t \geq 0$ where ϵ depends on θ and the constant of proportionality depends only on $\mathcal{Y}_t$.

Proof. Recall that

$$q_{t+1}(z^*,\theta) \propto \int \gamma\left(z^*|Y_{t+1}, D_t, z, \theta\right) \eta\left(Y_{t+1}|D_t, z, \theta\right) q_t(z,\theta)\lambda(dz)$$

where the proportionality constant can depend on the conditioning vector $\mathcal{Y}_{t+1}$. Thus

$$\begin{aligned}\exp[-(t+1+\tau)\epsilon]e(D_{t+1},z^*,\theta)q_{t+1}(z^*,\theta) \propto \exp(-\epsilon)\int &\frac{e(D_{t+1},z^*,\theta)}{e(D_t,z,\theta)}\exp\left[(1-\gamma)(\log C_{t+1} - \log C_t)\right] \\ &\times \gamma\left(z^*|Y_{t+1}, D_t, z\right)\eta\left(Y_{t+1}|D_t, z\right) \\ &\times e(D_t,z,\theta)\exp[-\epsilon(t+\tau)]q_t(z,\theta)\lambda(dz)\end{aligned}$$

given that

$$\exp\left[(1-\gamma)(\log C_{t+1} - \log C_t)\right]$$

is in the conditioning information set $\mathcal{Y}_{t+1}$. Thus

$$\hat{q}_{t+1}(z^*,\theta) \propto \exp[-(t+1)\epsilon]e(D_{t+1},z^*,\theta)q_{t+1}(z^*,\theta),$$

provided that

$$\hat{q}_t(z,\theta) \propto \exp(-t\epsilon)e(D_t,z,\theta)q_t(z^*,\theta).$$

The conclusion follows by induction since the date zero proportionality holds by assumption.

□

This result shows how the change of measure that alters the dynamic evolution affects the implied stationary distribution. This link will prove valuable in designing filtering algorithms that adapt to the concerns that investors might have for model specification.

5.3 Robustness and risk sensitivity

Since the work of Jacobson (1973) and Whittle (1990), there is a well known connection between risk sensitivity, robustness and large-deviation theory. Hansen and Sargent (1995), Maenhout (2004) and Hansen et al. (2006) show how to adapt the Jacobson and Whittle formulation to recursive utility. The risk aversion parameter $1 - \omega$ is related to an entropy penalization parameter ξ used to discipline a concern about model misspecification via the formula:

$$\xi = -\frac{1}{1-\omega},$$

which is positive provided that $\omega > 1$. When states or parameters are unknown, robust valuation and pricing lead to the use of the robust-adjusted density:

$$\tilde{q}_t(z,\theta) \propto \exp(\epsilon\tau)e(D_t, z, \theta)q_t(z,\theta),$$

In the case of known parameters, the stationary density that emerges as a result of the the change of measure of produces the correct scaling. In other words

$$\tilde{q}_t(z,\theta) \propto \hat{q}_t(z,\theta),$$

however when parameters are unknown we must include an eigenfunction adjustment:

$$\tilde{q}_t(z,\theta) \propto \exp(\epsilon t)\hat{q}_t(z,\theta)$$

because $\exp(\epsilon t)$ depends on unknown parameters. In what follows we will use a recursive approach for this adjustment. Given a date t a numerical approximation for $\tilde{q}_t$ that includes the $\exp(\epsilon t)$ adjustment, we will simulate using the $\hat{}$ evolution and then make an $\exp(\epsilon)$ adjustment to construct $\tilde{q}_{t+1}$. This repeated adjustment implements the implicit change in the reference model under robust learning implicit in Basar and Bernhard (1989) and delineated in Hansen and Sargent (2005).

6 Examples

We consider robust utility-based adjustments to filtering methods applied to four example economies.

6.1 Unknown mean growth rate

This example gives an illustration of robust Kalman filtering.

$$\begin{aligned} Y_{t+1} &= \mu + Z_t + GW_{t+1} \\ Z_{t+1} &= AZ_t + BW_{t+1} \end{aligned}$$

where X is a scalar hidden state and μ an unknown parameter but G, A and B are known. This can be viewed as a state space system when written as:

$$\begin{aligned} Y_{t+1} &= \mu + Z_t + GW_{t+1} \\ \begin{bmatrix} Z_{t+1} \\ \mu \end{bmatrix} &= \begin{bmatrix} A & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} Z_t \\ \mu \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} W_{t+1}. \end{aligned} \quad (10)$$

In this system Y_{t+1} is the growth rate of consumption. Kalman filtering gives recursive estimators of Z_t and μ. Call these $\bar{Z}_t$ and $\bar{\mu}_t$ and their joint covariance matrix:

$$\begin{bmatrix} \Sigma_t & \Lambda_t \\ \Lambda_t & \Pi_t \end{bmatrix}.$$

To construct a change of measure, we use the eigenfunction:

$$e(z) = \exp\left[-\frac{z}{\xi(1-A)}\right].$$

which is independent of μ. The eigenvalue associated with this eigenfunction is

$$\exp(\epsilon) \propto \exp\left(-\frac{1}{\xi}\mu\right)$$

where the proportionality constant does not depend on μ. Then the shock W_{t+1} has distorted dynamics with precision: I and mean conditioned on X_t and θ:

$$-\frac{1}{\xi}\left(G' + B'\chi\right)$$

where

$$\chi = \frac{1}{1-A}.$$

This leads us to the distorted dynamics:

$$\begin{aligned} Y_{t+1} &= \mu + Z_t - \frac{1}{\xi}G\left(G' + B'\chi\right) + G\hat{W}_{t+1} \\ Z_{t+1} &= AZ_t - \frac{1}{\xi}B\left(G' + B'\chi\right) + B\hat{W}_{t+1} \end{aligned}$$

where $\{\hat{W}_{t+1}\}$ is multivariate sequence of independent, standard normally distributed random vectors.

Applying the Kalman filter gives estimates for $\hat{\mu}_t$ and $\hat{Z}_t$ with the same covariance matrix as from the original Kalman filter. The recursive updating equations are:

$$\begin{aligned} \hat{Z}_{t+1} =& A\hat{Z}_t - \frac{1}{\xi}B\left(G' + B'\chi\right) \\ &+ \left[\frac{\Lambda_t + \Sigma_t + BG'}{\Sigma_t + 2\Lambda_t + \Pi_t + GG'}\right]\left[Y_{t+1} - \hat{Z}_t - \hat{\mu}_t + \frac{1}{\xi}G\left(G' + B'\chi\right)\right] \\ \hat{\mu}_{t+1} =& \hat{\mu}_t \\ &+ \left[\frac{\Lambda_t + \Pi_t}{\Sigma_t + 2\Lambda_t + \Pi_t + GG'}\right]\left[Y_{t+1} - \hat{Z}_t - \hat{\mu}_t + \frac{1}{\xi}G\left(G' + B'\chi\right)\right] \end{aligned} \tag{11}$$

This distorted system will produce $\hat{Z}_t$ sequences that are systematically negative in contrast to the original Kalman filter. On the other hand, $\hat{\mu}_t$ will be distorted in a positive direction. An eigenvalue adjustment will offset this latter distortion.

We now consider a recursive specification of the eigenfunction adjustment. Let $\tilde{Z}_t$ and $\tilde{\mu}_t$ be the adjusted conditional estimates of Z_t and μ_t With joint covariance matrix:

$$\begin{bmatrix} \Sigma_t & \Lambda_t \\ \Lambda_t & \Pi_t \end{bmatrix}.$$

These estimates are computed recursively from the distorted dynamics except that at each date we make an eigenvalue adjustment:

$$\begin{aligned}
\tilde{Z}_{t+1} =& -\frac{1}{\xi}\Lambda_{t+1} + A\tilde{Z}_t - \frac{1}{\xi}B\left(G' + B'\chi\right) \\
&+ \left[\frac{\Lambda_t + \Sigma_t + BG'}{\Sigma_t + 2\Lambda_t + \Pi_t + GG'}\right]\left[Y_{t+1} - \tilde{Z}_t - \tilde{\mu}_t + \frac{1}{\xi}G\left(G' + B'\chi\right)\right] \\
\tilde{\mu}_{t+1} =& -\frac{1}{\xi}\Pi_{t+1} + \tilde{\mu}_t \\
&+ \left[\frac{\Lambda_t + \Pi_t}{\Sigma_t + 2\Lambda_t + \Pi_t + GG'}\right]\left[Y_{t+1} - \tilde{Z}_t - \tilde{\mu}_t + \frac{1}{\xi}G\left(G' + B'\chi\right)\right]
\end{aligned}$$

The inclusion of the terms

$$\begin{matrix} -\frac{1}{\xi}\Lambda_{t+1} \\ -\frac{1}{\xi}\Pi_{t+1} \end{matrix}$$

relative to recursion (11) are outcome of the eigenvalue adjustment. This is recognized as a special case of the robust counterpart to the Kalman filter.

To initialize the Kalman filter, suppose that the prior for μ and Z_0 are independent implying that $\Lambda_0 = 0$. We restrict Z_0 to be in the stationary distribution implying that

$$\begin{aligned}
\bar{Z}_0 &= 0 \\
\Sigma_0 &= \frac{|B|^2}{1 - A^2}
\end{aligned}$$

Under the $\tilde{\cdot}$ change of measure,

$$\begin{aligned}
\tilde{Z}_0 &= -\frac{|B|^2}{\xi(1 - A^2)(1 - A)} \\
\tilde{\mu}_0 &= \bar{\mu}_0 - \frac{\tau}{\xi}\Pi_0
\end{aligned}$$

We illustrate the outcome of the Kalman filter, in large part to motivate what follows. In this illustration we assume that $B = \begin{bmatrix} 0 & b \end{bmatrix}$ where $b = .0003$, $G = \begin{bmatrix} g & 0 \end{bmatrix}$ where $g = .005$, $A = .978$. We initialize the prior for μ by setting:

$$\begin{aligned}
\bar{\mu}_0 &= .005 \\
\Pi_0 &= .002^2.
\end{aligned}$$

The parameter $\tau .23$ and $\xi = .1$. The Kalman filter iterations for the covariance matrix converge. We report the original conditional distribution for the hidden state Z_t and its distorted counterpart in Figure 1. The concern for robustness shifts the state distribution to the left by reducing its mean conditioned on the history of consumption signals.

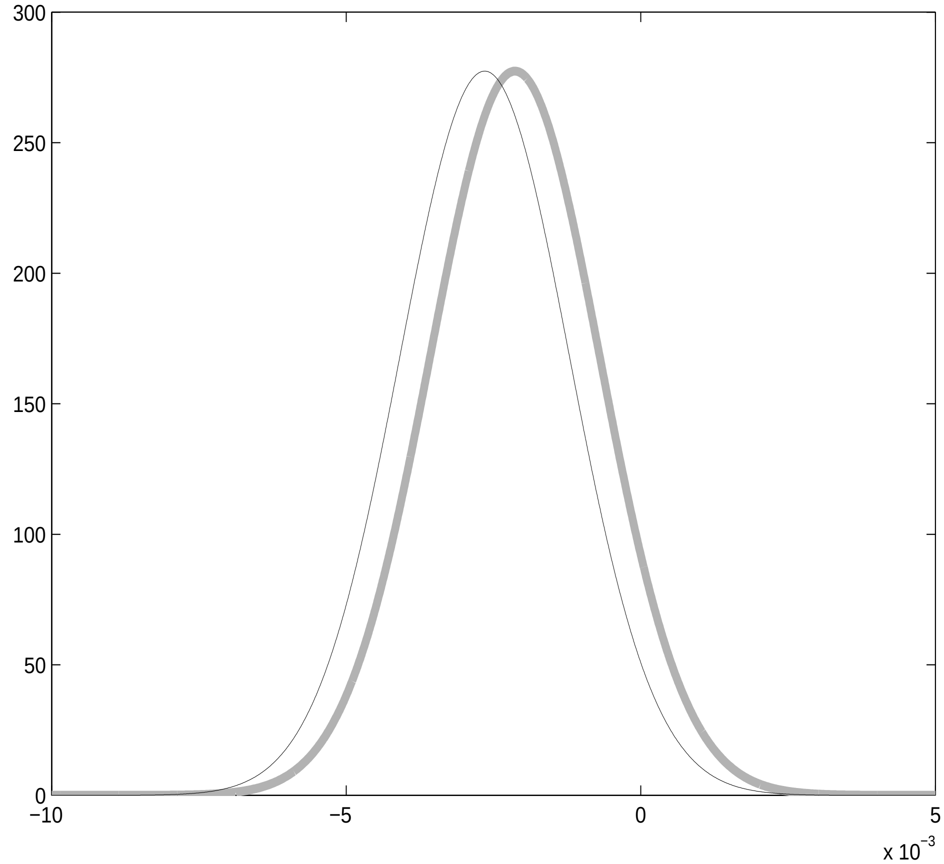

Figure 1: Conditional density for the hidden state. Thick line is the density without the robustness adjustment and the thin line adjusts for robustness.

We report the distortion in the conditional mean for consumption scaled by the conditional standard deviation g in Figure 2. When there is a unitary elasticity of intertemporal substitution, this can be interpreted as a price of uncertainty. See Hansen and Sargent (2010).

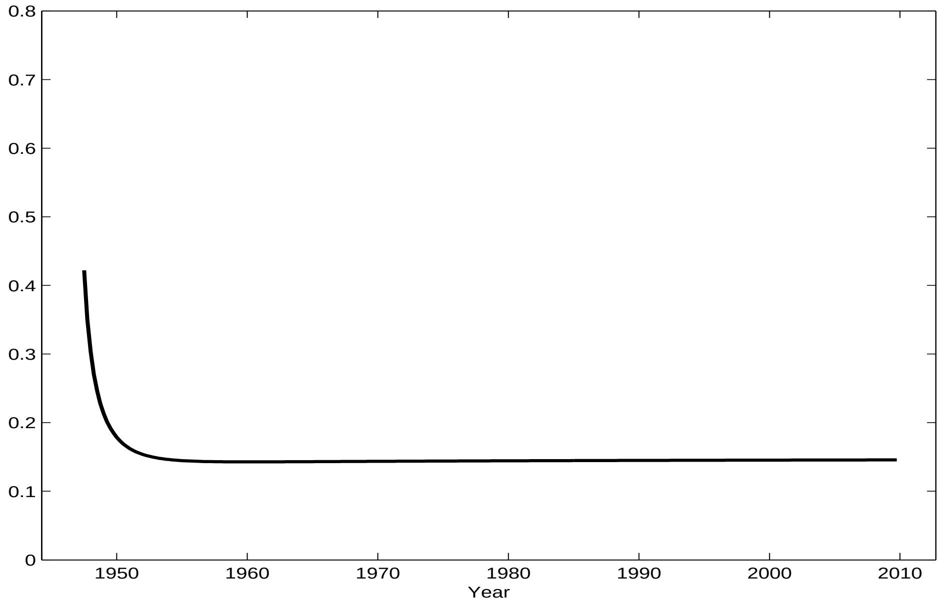

Figure 2: The time series for the one-period price of uncertainty with $\tau = 23$.

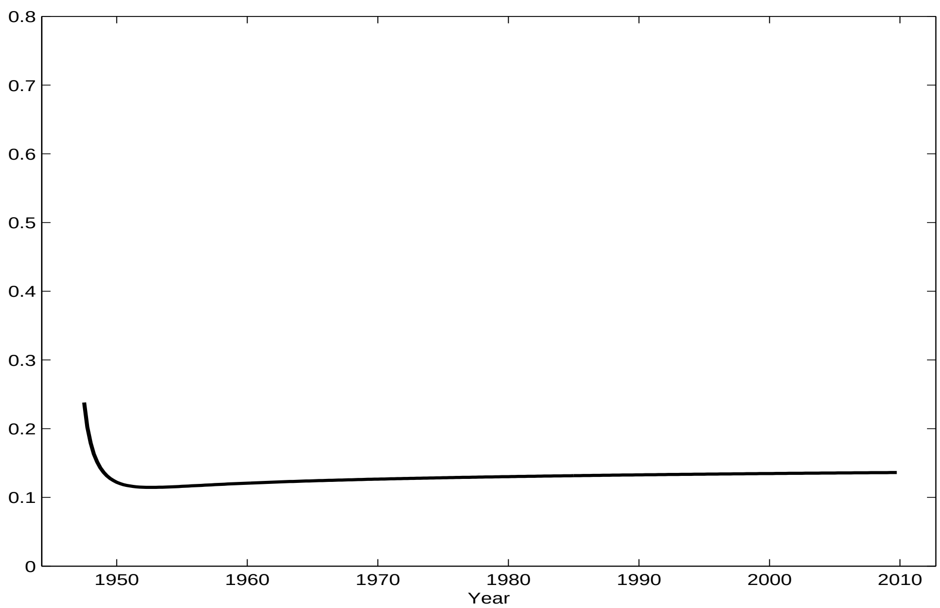

Figure 3: The time series for the one-period price of uncertainty with $\tau = 0$.

For the Kalman filtering model this distortion varies smoothly with respect to calendar time. There is an initial phase of active learning after when distortion becomes flat over time. Lowering τ diminishes the initial uncertainty. This is reflected in Figure 3 which presumes that $\tau = 0$:

6.2 Conditioning on A and G

We now make the estimation a bit more ambitious by letting the shock coefficient $B = \begin{bmatrix} b & 0 \end{bmatrix}$ be estimated by the decision maker. As a consequence, we can no longer apply the Kalman filter, but instead we rely on sufficient statistics in conjunction with the particle filter to compute uncertainty prices numerically. We suppose that A and $G = \begin{bmatrix} 0 & g \end{bmatrix}$ are given.

The eigenvalue formula of interest is:

$$\exp(\epsilon) \propto \exp\left(-\frac{1}{\xi}\mu + \frac{1}{2}\left[\frac{1}{\xi(1-A)}\right]^2 b^2\right).$$

We impose the priors

$$Z_0 \sim \mathcal{N}\left(0, \frac{b^2}{1-A^2}\right)$$
$$\mu \sim \mathcal{N}\left(\bar{\mu}_0, \Pi_0\right)$$

which are assumed to be independent conditioned on b. We assume the marginal prior on b^2 to be a generalized inverse Gaussian distribution with density of the form:

$$\propto \exp\left[-\alpha_1 \frac{1}{b^2} - \alpha_2 b^2\right] (b^2)^{-\alpha_3}.$$

Following Geweke (1993), we include a term in the exponential of both $\frac{1}{b^2}$ and b^2. As we will see the latter contribution has further motivation as an adjustment for robustness.[3]

Next we consider the distorted priors conditioned on b:

$$Z_0 \sim \mathcal{N}\left(-\frac{b^2}{\xi(1-A^2)(1-A)}, \frac{b^2}{1-A^2}\right)$$
$$\mu \sim \mathcal{N}\left(\bar{\mu}_0 - \frac{\tau}{\xi}\Pi_0, \Pi_0\right)$$

and the marginal prior on b^2

$$\propto \exp\left[-\alpha_1\left(\frac{1}{b^2}\right) - \alpha_2 b^2 - \frac{\tau}{2}\left(\frac{1}{\xi(1-A)}\right)^2 b^2\right] (b^2)^{-\alpha_3}.$$

For the purposes of building sufficient statistics we treat the hidden state as data, and we use two contributions to the likelihood

$$\exp\left[\frac{1}{g^2}\sum_{u=0}^{t-1}(Y_{u+1} - Z_u)\mu - \frac{t}{2g^2}\mu^2\right] \exp\left[-\frac{1}{2b^2}\sum_{u=0}^{t-1}(Z_{u+1} - AZ_u)^2\right] (b^2)^{-t/2}$$

[3] Geweke (1993) does not consider models with hidden states, and the estimation problem he explores is thus distinct from ours.

This leads us to construct the sufficient statistics:

$$
\begin{aligned}
S_{t+1}^{[1]} &= \sum_{u=0}^{t} \frac{1}{g^2}(Y_{u+1} - Z_u) + S_0^{[1]} = \frac{1}{g^2}(Y_{t+1} - Z_t) + S_t^{[1]} \\
S_{t+1}^{[2]} &= (t+1)\frac{1}{g^2} + S_0^{[2]} = \frac{1}{g^2} + S_t^{[2]} \\
S_{t+1}^{[3]} &= \sum_{u=0}^{t} \frac{1}{2}(Z_{u+1} - AZ_u)^2 + S_0^{[3]} = \frac{1}{2}(Z_{t+1} - AZ_t)^2 + S_t^{[3]} \\
S_{t+1}^{[4]} &= t + 1 + S_0^{[4]} = 1 + S_t^{[4]}.
\end{aligned}
$$

where[4]

$$
S_0^{[1]} = \frac{\bar{\mu}_0}{\Pi_0}, S_0^{[2]} = \frac{1}{\Pi_0}, S_0^{[3]} = \alpha_1, S_0^{[4]} = 2\alpha_3.
$$

Then the date t posteriors are:

$$
\begin{aligned}
\mu &\sim \mathcal{N}\left(\frac{S_t^{[1]}}{S_t^{[2]}}, \frac{1}{S_t^{[2]}}\right) \\
b^2 &\propto \exp\left[-S_t^{[3]}\left(\frac{1}{b^2}\right) - \alpha_2 (b)^2\right](b^2)^{-S_t^{[4]}/2}.
\end{aligned}
$$

When we use the distorted law of motion we use the likelihood contributions:

$$
\begin{aligned}
&\exp\left[\frac{1}{2g^2}\sum_{u=0}^{t-1}\left(Y_{u+1} - Z_u + \frac{g^2}{\xi}\right)\mu - \frac{t}{g^2}\mu^2\right] \\
&\times \exp\left[-\frac{1}{2b^2}\sum_{u=0}^{t-1}\left(Z_{u+1} - AZ_u + \frac{b^2}{\xi(1-A)}\right)^2\right](b^2)^{-t/2}
\end{aligned}
$$

Notice that

$$
\begin{aligned}
&\exp\left[-\frac{1}{2b^2}\sum_{u=0}^{t-1}\left(Z_{u+1} - AZ_u + \frac{b^2}{\xi(1-A)}\right)^2\right](b^2)^{-t/2} \\
&\propto \exp\left(-\frac{1}{2b^2}\sum_{u=0}^{t-1}(Z_{u+1} - AZ_u)^2 - b^2\frac{t}{2}\left[\frac{1}{\xi(1-A)}\right]^2\right)(b^2)^{-t/2}.
\end{aligned}
$$

We now use the sufficient statistics

$$
\begin{aligned}
\hat{S}_{t+1}^{[1]} &= \frac{1}{g^2}\left(Y_{t+1} - Z_t + \frac{g^2}{\xi}\right) + \hat{S}_t^{[1]} \\
\hat{S}_{t+1}^{[2]} &= \frac{1}{g^2} + \hat{S}_t^{[2]} \\
\hat{S}_{t+1}^{[3]} &= \frac{1}{2}(Z_{t+1} - AZ_t)^2 + \hat{S}_t^{[3]} \\
\hat{S}_{t+1}^{[4]} &= 1 + \hat{S}_t^{[4]}.
\end{aligned}
$$

[4] It is sometimes more convenient to use $S_{t+1}^{[1]}/S_{t+1}^{[2]}$ in place of $S_{t+1}^{[1]}$ and $S_{t+1}^{[3]}/S_{t+1}^{[4]}$ in place of $S_{t+1}^{[3]}$ as sufficient statistics.

where

$$\hat{S}_0^{[1]} = \frac{\bar{\mu}_0}{\Pi_0} - \frac{\tau}{\xi}, \hat{S}_0^{[2]} = S_0^{[2]} = \frac{1}{\Pi_0}, \hat{S}_0^{[3]} = S_0^{[3]} = \alpha_1, \hat{S}_0^{[4]} = S_0^{[4]} = 2\alpha_3.$$

Notice that $\hat{S}_t^{[j]} = S_t^{[j]}$ for $j = 2, 3, 4$.

Then the date t ˆ posteriors are:

$$\mu \sim \mathcal{N}\left(\frac{\hat{S}_t^{[1]}}{\hat{S}_t^{[2]}}, \frac{1}{\hat{S}_t^{[2]}}\right)$$
$$b^2 \propto \exp\left[-\hat{S}_t^{[3]}\left(\frac{1}{b^2}\right) - \alpha_2 b^2 - \left(\frac{t+\tau}{2}\right)\left[\frac{1}{\xi(1-A)}\right]^2 b^2\right](b^2)^{-\hat{S}_t^{[4]}/2}.$$

The distorted law of motion, however, is not what we use in our analysis. Instead we now must make an "eigenvalue" adjustment. This leads us to the ˜ evolution. Suppose the date t ˜ density $\tilde{q}_t$ has been computed and that there is an associated vector of sufficient statistics $\tilde{S}_t^{[j]}$ for $j = 1, 2, 3, 4$. We use $\exp(\epsilon)$ to alter the evolution the one-period evolution and hence the construction of the sufficient statistics.

We construct

$$\tilde{S}_{t+1}^{[1]} = \tilde{S}_t^{[1]} - \frac{1}{\xi} + \frac{1}{g^2}\left(Y_{t+1} - Z_t + \frac{g^2}{\xi}\right) = \tilde{S}_t^{[1]} + \frac{1}{g^2}(Y_{t+1} - Z_t)$$
$$\tilde{S}_{t+1}^{[2]} = \hat{S}_{t+1}^{[2]} = S_{t+1}^{[2]}$$
$$\tilde{S}_{t+1}^{[3]} = \hat{S}_{t+1}^{[3]} = S_{t+1}^{[3]}$$
$$\tilde{S}_{t+1}^{[4]} = \hat{S}_{t+1}^{[4]} = S_{t+1}^{[4]}.$$

The sufficient statistic $\tilde{S}_{t+1}^{[1]}$ gets updated just as $S_{t+1}^{[1]}$ but starts from a different initial condition. Then the date $t+1$ distributions for the parameters and hidden states are:

$$Z_{t+1} \sim \mathcal{N}\left(AZ_t - \frac{\chi}{\xi}b^2, b^2|b^2\right)$$
$$\mu \sim \mathcal{N}\left(\frac{\tilde{S}_{t+1}^{[1]}}{\tilde{S}_{t+1}^{[2]}}, \frac{1}{\tilde{S}_{t+1}^{[2]}}\right)$$
$$b^2 \propto \exp\left(-\tilde{S}_{t+1}^{[3]}\left(\frac{1}{b^2}\right) - \alpha_2 b^2 - \left(\frac{\tau}{2}\right)\left[\frac{1}{\xi(1-A)}\right]^2 b^2\right)(b^2)^{-\tilde{S}_{t+1}^{[4]}/2}$$

where the first distribution is conditioned on b^2.

When we use partial smoothing we must condition on more Y observation than Z observation. For this example, we just use a later date for the sufficient statistics based on Y than Z. Thus we use the likelihood:

$$\exp\left[\frac{1}{g^2}\sum_{u=0}^{t-1}(Y_{u+1} - Z_u)\mu - \frac{t}{2g^2}\mu^2\right]\exp\left[-\frac{1}{2b^2}\sum_{u=0}^{t-2}(Z_{u+1} - AZ_u)^2\right](b^2)^{(1-t)/2},$$

and we combine $S_{t+1}^{[1]}$ and $S_{t+1}^{[2]}$ in conjunction with $S_t^{[3]}$ and $S_t^{[4]}$, and similarly for the ˜ counterparts.

We report the one-period uncertainty prices in Figure 4. While there continues to be an initial phase in which parameter learning is prominent, the time series of uncertainty prices after this phase has some prominent peaks associated with recessions. Again a reduction in τ diminishes the impact of the learning phase as is evident in Figure 5. In this latter case we see a notable increase in the uncertainty prices in recent time periods.

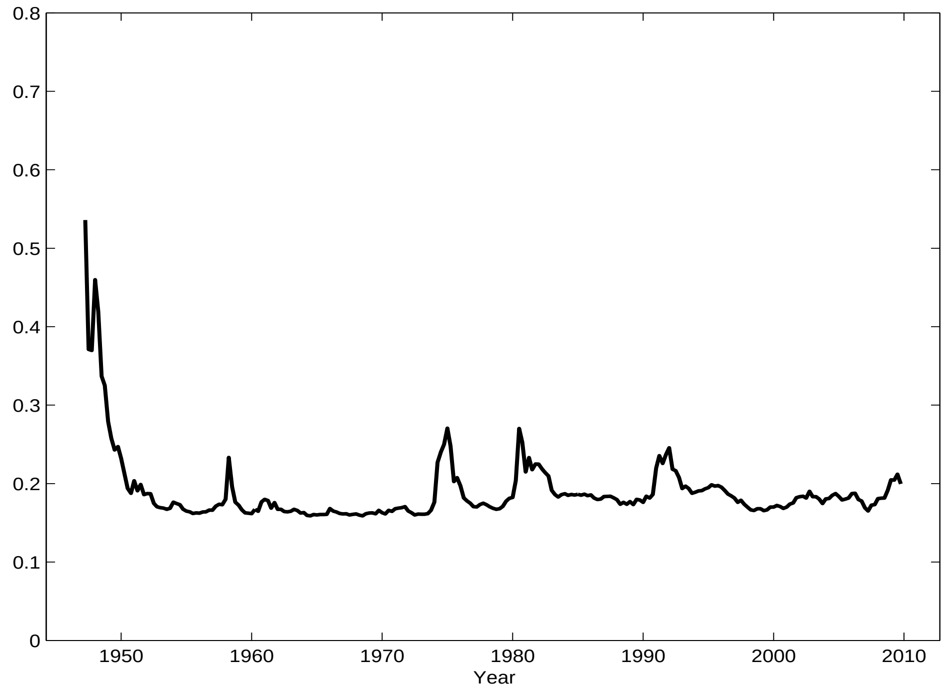

Figure 4: The time series for the one-period price of uncertainty when investors are uncertain about μ, B and the hidden state process Z. For this specification $\tau = 23$.

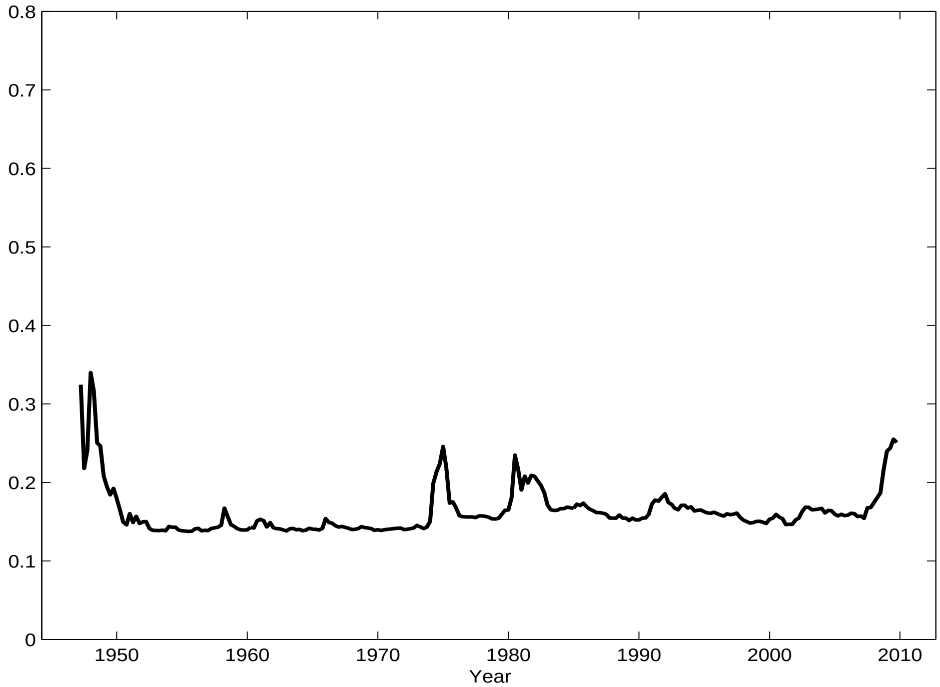

Figure 5: The time series for the one-period price of uncertainty when investors are uncertain about μ, B and the hidden state process Z. For this specification $\tau = 0$.

6.3 Conditioning only on G

Next we include the estimation of the parameter A in the analysis. For conceptual simplicity suppose that we have a discrete distribution over alternative values of A. To infer the distortion for this distribution, notice that scale factor for the generalized inverse Gaussian prior for b^2 depends on the parameter A, where we now view this density as a conditional density. The scale factor for the generalized inverse Gaussian is well known.[5] This scale dependence is offset when constructing the distorted prior distribution for A. We consider a uniform distribution for A defined on in interval $[.95, .99]$. Instead of using a discrete grid of points we constructed a truncated normal density for A that was essentially flat.

Robust learning induces the economic agents to distort priors towards large values of A and this is reflected in the upper two plots of figure 6. By the end of the sample the distorted distribution is largely concentrated at the endpoint.

[5]It depends on a modified Bessel function of the second kind.

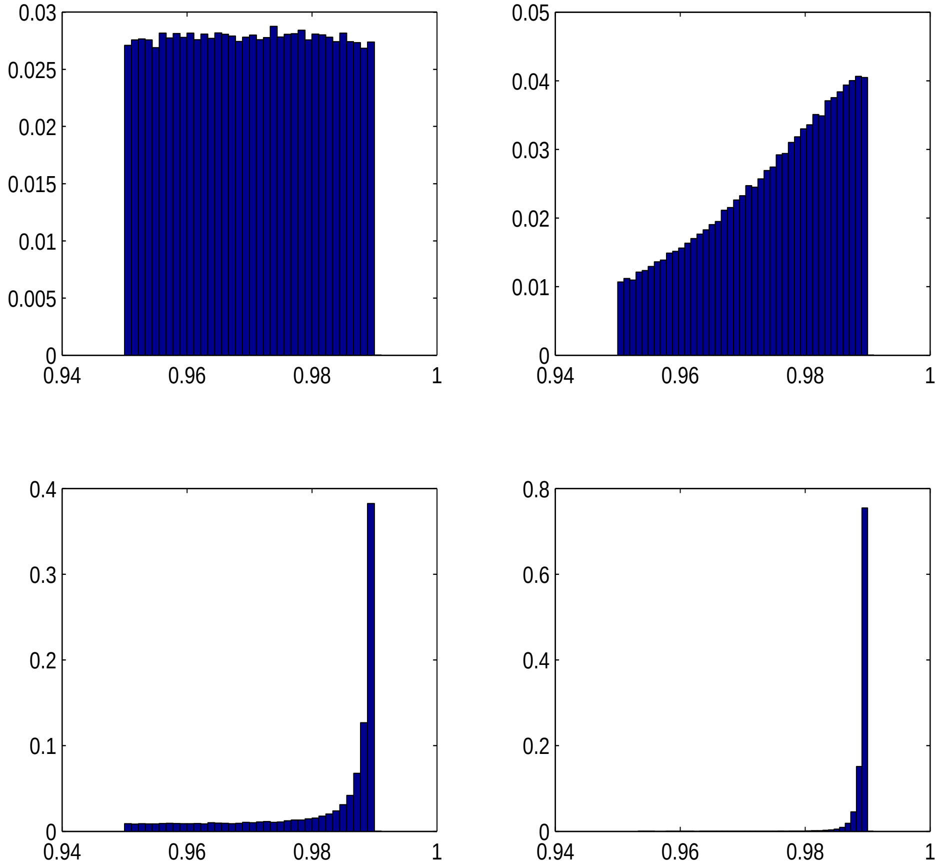

Figure 6: Histograms for the distribution of A. The first column gives the prior and slanted prior at the initial date, and the second column gives the end of sample posterior and slanted posterior.

The time series for the price of uncertainty is reported in figure 9.

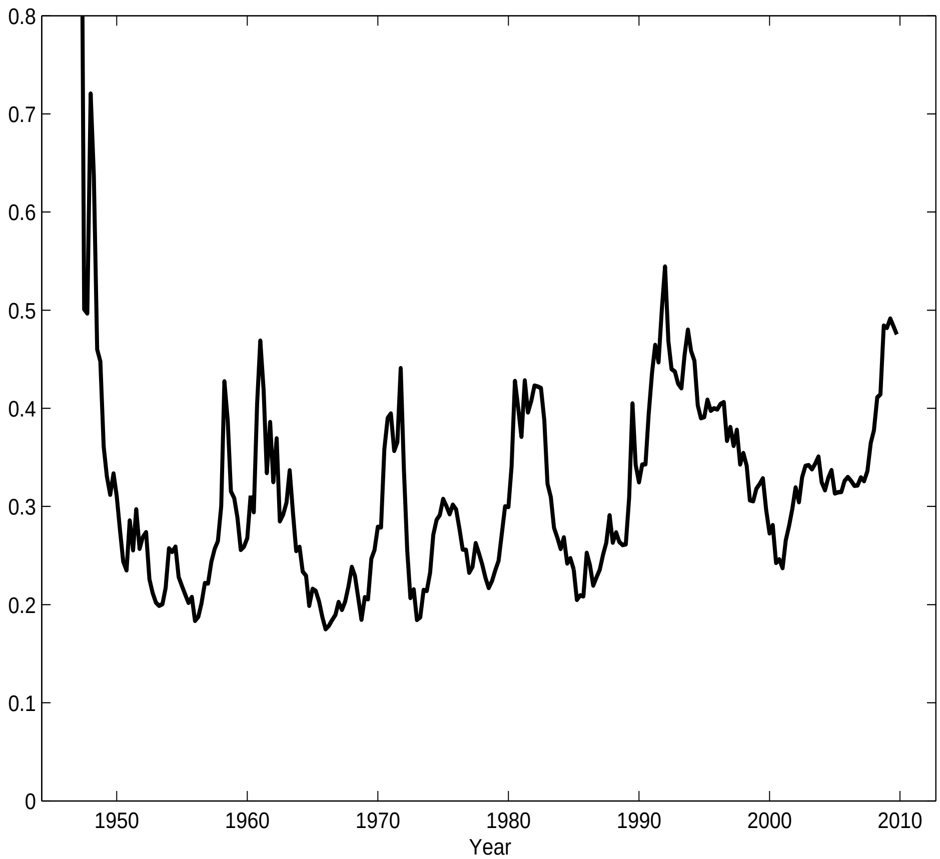

Figure 7: The time series for the one-period price of uncertainty when investors are uncertain about μ, A, B and the hidden state process Z. For this specification $\tau = 23$.

We next consider sensitivity to the choice of the prior value of τ. As an alternative we set $\tau = 0$ Prior and posterior histograms for the autoregressive are given in figure 8. The distorted distribution collapses more slowly to the upper bound of .99.

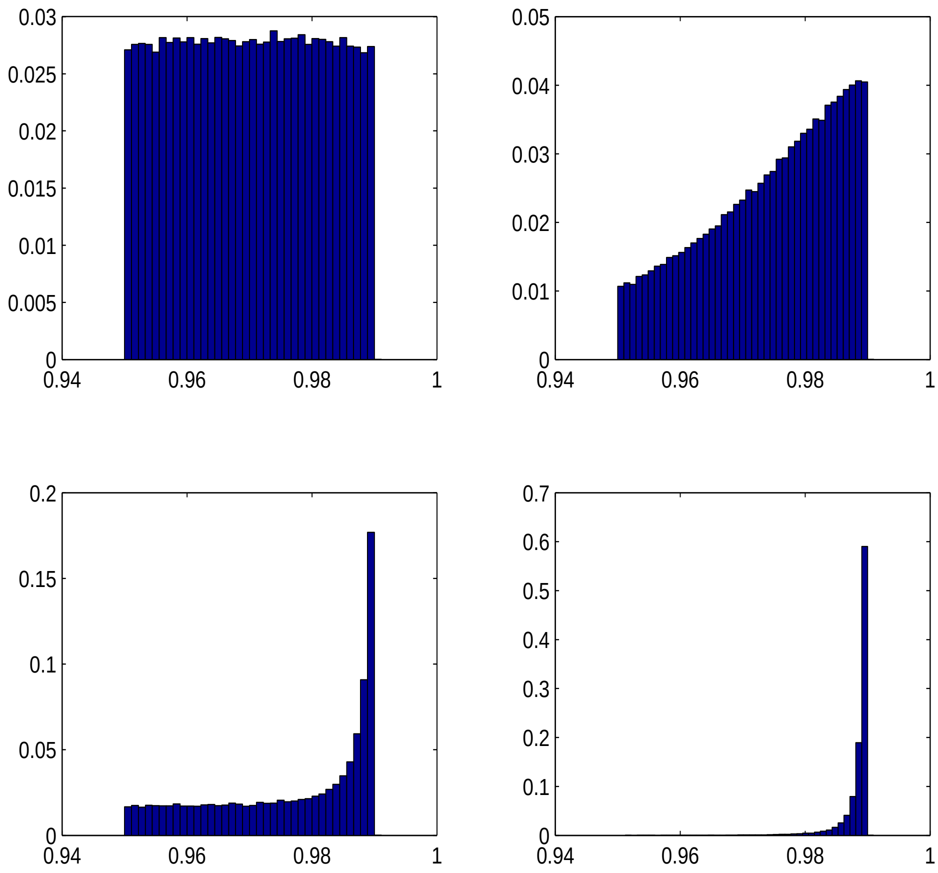

Figure 8: Histograms for the distribution of A. The first column gives the prior and slanted prior at the initial date, and the second column gives the end of sample posterior and slanted posterior.The time series for the one-period price of uncertainty when investors are uncertain about μ, A, B and the hidden state process Z. For this specification $\tau = 0$.

The important difference in the uncertainty prices is that the initial prices are now more modest than in our earlier calculations.

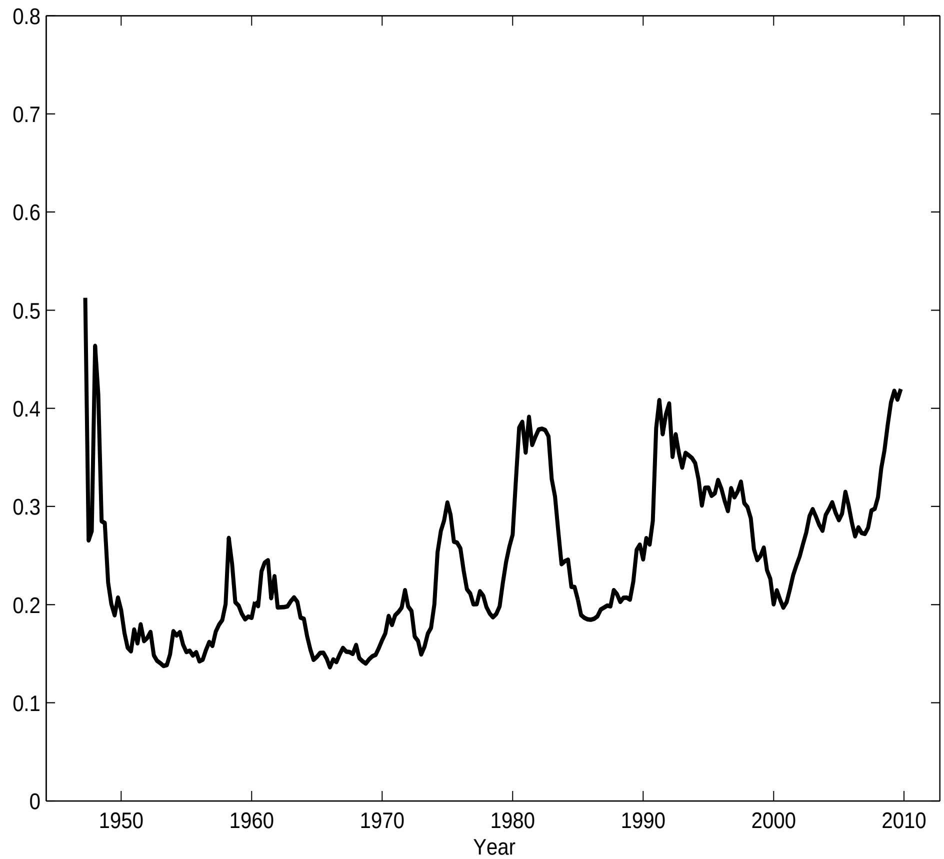

Figure 9: The time series for the one-period price of uncertainty when investors are uncertain about μ, A, B and the hidden state process Z. For this specification $\tau = 0$.

6.4 Stochastic volatility

To be added.

References

Basar, Tamar S. and Pierre Bernhard, eds. 1989. *Differential Games and Applications (Lecture Notes in Control and Information Sciences*. Springer.

Carvalho, C., M. Johannes, H. Lopes, and N. Polson. 2010. Particle Smoothing and Statistical Science. *Statistical Science* 25.

DeJong, David N., Hariharan Dharmarajan, Roman Liesenfeld, Guilherme V. Moura, and Jean-Froncois Richard. 2010. Efficient Filtering Likelihood Evaluation of State-Space Representations. University of Pittsburgh and Univeristat Kiel.

Epstein, L. and S. Zin. 1989. Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework. *Econometrica* 57 (4):937–969.

Fearnhead, P. 2002. MCMC, Sufficient Statistics and Particle Filter. *Journal of Computuational and Graphical Statistics* 11:848–862.

Geweke, John. 1993. A note on some limitations of CRRA utility. *Economic Letters* 71:341–345.

Hansen, Lars Peter. 2007. Beliefs, Doubts and Learning: Valuing Macroeconomic Risk. *American Economic Review* 97 (2):1–30.

Hansen, Lars Peter and Thomas J. Sargent. 1995. Discounted Linear Exponential Quadratic Gaussian Control. *IEEE Transactions on Automatic Control* 40.

———. 2005. Robust estimation and control under commitment. *Journal of Economic Theory* 124 (2):258–301.

———. 2007. Recursive robust estimation and control without commitment. *Journal of Economic Theory* 136 (1):1–27.

———. 2010. Fragile Beliefs and the Price of Uncertainty. *Quantitative Economics* 1 (1):129–162.

Hansen, Lars Peter and Jose Alexandre Scheinkman. 2009. Long-term Risk: An Operator Approach. *Econometrica* 77 (1):177–234.

Hansen, Lars Peter, Thomas J. Sargent, Guahar A. Turmuhambetova, and Noah Williams. 2006. Robust Control and Model Misspecification. *Journal of Economic Theory* 128:45–90.

Jacobson, David H. 1973. Optimal Stochastic Linear Systems with Exponential Performance Criteria and Their Relation to Deterministic Differential Games. *IEEE Transactions for Automatic Control* AC-18:1124–131.

Johannes, Michael S. and Nick Polson. 2007. Exact particle Filtering and Parameter Learning. University of Chicago and Columbia University.

Kreps, David M. and Evon L. Porteus. 1978. Temporal Resolution of Uncertainty and Dynamic Choice. *Econometrica* 46 (1):185–200.

Maenhout, Pascal J. 2004. Robust Portfolio Rules and Asset Pricing. *Review of Financial Studies* 17 (4):951–983.

Storvik, G. 2002. Particle Filters in State Space Models with the Presence of Unknown Static Parameters. *IEEE Trans. on Signal Processing* 50:281–289.

Thurn, Sebastian, John Langford, and Vandi Verma. 2002. Risk Sensitive Particle Filters. In *Advances in Neural Information Processing Systems 14*, vol. 2, edited by Thomas Diettrich, Suzanna Becker, and Zoubin Ghahramani. Cambridge, MA: MIT Press.

Whittle, P. 1969. A View of Stochastic Control Theory. *Journal of the Royal Statistical Society. Series A (General)* 132 (3):320–334.

Whittle, Peter. 1990. *Risk Sensitive and Optimal Control*. West Suffix, England: John Wiley and Sons.