# Robust Permanent Income and Pricing with Filtering*

Lars Peter Hansen†

University of Chicago

Thomas J. Sargent‡

Stanford University and Hoover Institution

and

Neng E. Wang§

Stanford University

January 24, 2002

*We thank Harald Uhlig, Darrell Duffie, Kenneth Singleton, and Ming Huang for criticisms of earlier drafts and seminar presentations. Hansen and Sargent thank the National Science Foundation for research grants that supported their research for this paper.

†l-hansen@uchicago.edu.

‡The corresponding author. sargent@stanford.edu.

§newang@stanford.edu.

*running head:* robust permanent income

*Address correspondence*:

Prof. Thomas J. Sargent
Department of Economics
Stanford University
Stanford, CA 94305

email: sargent@stanford.edu

**Abstract**

A planner and agent in a permanent income economy cannot observe part of the state, regard their model as an approximation, and value decision rules that are robust across a set of models. They use robust decision theory to choose allocations. Equilibrium prices reflect the preference for robustness and so embody a 'market price of Knightian uncertainty'. We compute market prices of risk and compare them with a model that assumes that the state is fully observed. We use detection error probabilities to constrain a single parameter that governs the taste for robustness.

Key Words: Kalman filter, approximating model, Knightian uncertainty, robustness, equity premium, market price of uncertainty, permanent income.

# 1 Introduction

This paper studies decision making and asset pricing in the presence of model uncertainty and an imperfectly measured state vector. Agents treat their model as a good approximation to an unknown 'true model'. Doubts about the model make agents want decision rules that work well for a set of models close to their approximating model. We formalize model uncertainty using a robust decision theory cast in terms of an explicit set of models. We augment previous work by formulating how a robust decision maker should proceed when parts of the state that are useful for forecasting are not observed.

We formulate a discounted linear-quadratic control problem with an unobserved state, then apply it to compute equilibrium asset prices within a stochastic growth model calibrated to U.S. data.[1] We use the stochastic growth model as a laboratory to study how agents' preference for decisions robust to model misspecification affects equilibrium allocation and asset prices.

Our laboratory is a model that Hansen, Sargent, and Tallarini (1999) estimated from time series on consumption and investment for the post 1970's U.S. In HST's model, the representative consumer faces an exogenous endowment process that is a sum of two serially correlated stochastic components. HST assumed that the representative consumer sees the state vector, including current and lagged values of both components of the endowment process. At their maximum likelihood parameter estimates, HST could actually infer the two stochastic components of the endowment process from the data on consumption and investment used to estimate the model.

In this paper, we recast HST's model by concealing elements of the state from the consumer. We allow the consumer to see current and lagged values of only the aggregate endowment and not its components. We follow HST in imputing model uncertainty to the representative agent, inspiring a preference for robust estimators and decision rules. The representative agent uses robust filtering and control, both to choose a consumption, savings plan and to price risky claims.

This setting requires that we reconstruct HST's decision and pricing theory to incorporate effects of model uncertainty that influence filtering. We accomplish this by building on results of Hansen and Sargent (2000), who have modified and extended the linear quadratic robust decision and filtering theory of Basar and Bernhard (1995) and Whittle (1990) to discounted problems of a type that are especially relevant to economics and finance. We show how to adapt HST's pricing formulas when the state is unobserved. We follow HST in defining a multiplicative adjustment to a stochastic discount factor that reflects the representative agent's preference for robustness. We use this adjustment to compute a 'market price of model uncertainty' and study how it affects the market price of risk.

We want quantitative estimates of how filtering affects the market price of model uncertainty. Our hunch was originally that confounding the representative agent's problem by adding filtering can raise the market price of model uncertainty, thereby helping to explain the equity premium.[2] Quantifying the effects of a preference for robustness on the market price of model uncertainty requires that we find a way to discipline the one parameter that in our framework describes that preference. We use Bayesian statistical detection theory to discipline that parameter, along the lines described by Anderson, Hansen, and Sargent (2000). When we keep detection error probabilities constant across the no-filtering-needed model of HST and the filtering-needed model of this paper (to be dubbed the HSW model), we find little additional effect on the market price of uncertainty from making agents filter. We suspect that this reflects that the detection error probabilities do not properly penalize the added complexity of the approximating model that is used by the agent who must filter. We are not yet prepared to concede that the above hunch is misguided.

The following issue arises in asset pricing models in which the hidden Markov structure of an endowment or dividend process impels an agent to filter. Without a preference for robustness, such a model is observationally equivalent to another with a fully observed state following a more complicated stochastic process. The agents forecast future returns using that state and its stochastic process. Indeed, the solution of the filtering problem in the original hidden Markov model produces this more complicated state and stochastic process for the endowment or dividend. Thus, rather than positing the filtering problem, one could simply begin with that richer state and law of motion. Positing the hidden Markov model can only be defended as a parsimonious way of specifying a richer stochastic process for the observable data.

We show that a preference for robustness causes the filtering and decision problems to interact in a way that destroys the preceding observational equivalence. We highlight this result by also constructing what we call a 'comparison model' that shuts down the interaction between the filtering and decision problems. This model allows us to identify an additional dimension of model misspecification (or 'deception') that concerns the robust decision maker when he takes into account that the richer representation of the dividend or endowment process is itself the result of solving a filtering problem. We also display numerical calculations that show the quantitative effects of this additional source of misspecification.

We are interested in the HST model partly for studying the market price of Knightian uncertainty, and partly as a laboratory for applying robust decision methods more generally. The combined robust filtering and control methods described in this paper have applications in various macroeconomic models.[3]

The remainder of this paper is organized as follows. Section 2 describes key asset pricing formulas and gives a representation of the market price of risk in terms of HST's market price of Knightian uncertainty. Section 3 describes the basic robust decision theory, the set of models used to represent Knightian uncertainty, and three salient models from within this set. Section 4 recasts HST's model in a notation compatible with Hansen and Sargent's (2000) machinery for joint filtering and control. Section 5 describes detection error probabilities and how they can be used to discipline $\theta$, the single parameter that measures preferences for robustness. Section 6 describes HST's observational equivalence result, the foundation of their empirical strategy and ours. Section 7 reformulates HST's model by causing the planner and the agent to estimate the state. A preference for robustness makes the filtering problem interact with the control problem in a way that it does not when the model is treated as known. Section 9 reports the computed multi-period market prices for our model. Section 10 concludes and suggests fruitful next steps. Three appendices describe technical details about constructing detection error probabilities, robust decision rules, and multi-period asset prices.

# 2 Asset Pricing Theory in Brief

Let $p_{t+1}$ be a payoff at $t+1$ and $q_t$ be its price at $t$. Asset pricing theories[4] start from the Euler equation

$$q_t = \mathbb{E}[m_{t+1} p_{t+1} | \mathcal{J}_t] \equiv \mathbb{E}_t \left[ m_{t+1} p_{t+1} \right], \tag{1}$$

where $\mathbb{E}$ is the mathematical expectation with $\mathcal{J}_t$ a time $t$ $\sigma$-algebra, and $m_{t+1}$ a stochastic discount factor. To give content to (1), we must specify a model (i.e., a probability distribution) with respect to which $\mathbb{E}$ is evaluated. For most of this paper, we let $\mathbb{E}$ be evaluated with respect to the planner's approximating model. We show how a preference for robustness modifies the ordinary formula for the stochastic discount factor in consumption-based asset pricing models.[5]

Using the definition of a conditional covariance and the Cauchy-Schwarz inequality, we obtain the inequality

$$\frac{q_t}{\mathbb{E}_t m_{t+1}} \geq \mathbb{E}_t p_{t+1} - \frac{\sigma_t(m_{t+1})}{\mathbb{E}_t m_{t+1}} \sigma_t(p_{t+1}), \tag{2}$$

where $\frac{\sigma_t(m_{t+1})}{\mathbb{E}_t m_{t+1}}$ is called the *market price of risk*. Notice that the left side is the ratio of the price of a claim to payoff $p_{t+1}$ to the price of a risk-less claim on one unit of consumption next period. The right side then relates this price ratio to the mean and standard deviation of the payoff. Inequality (2) becomes an equality for payoffs on the conditional mean-standard deviation frontier. Hansen and Jagannathan's statement of the equity premium puzzle is that data on asset market returns and prices give values of the market price of risk that are too high to be reconciled with many particular models of the stochastic discount factor $m_{t+1}$. This is because those theories make the conditional standard deviation of the stochastic discount factor $\sigma_t(m_{t+1})$ too small.[6] Two classic theories of the discount factor $m_{t+1}$ are

- Theory 1: $m_{t+1} = \beta$, used in Shiller (1981), where $\beta \in (0, 1)$ is a constant.
- Theory 2: $m_{t+1} = m^f_{t+1} \equiv \beta \frac{u'(c_{t+1})}{u'(c_t)}$ used in LeRoy (1973), Lucas Jr. (1978), and Breeden (1979), where $u(c_t)$ is a constant relative risk aversion one-period utility function, and $c_t$ is consumption of a representative consumer.

Both of these theories have small $\sigma_t(m_{t+1})$: the former theory makes it zero by definition, the latter makes it small under a constant relative risk aversion utility function evaluated at aggregate U.S. consumption growth rates.[7]

This paper uses HST's:

- Theory 3: $m_{t+1} = m^f_{t+1} m^u_{t+1}$, where $m^u_{t+1}$ is a multiplicative adjustment to the stochastic discount factor that reflects agents' aversion to model uncertainty.

HST call $m^u_{t+1}$ the *market price of Knightian uncertainty*. They deduce measures of it using the robust decision theory to be described below. Those measures reflect agents' doubt about the approximating model that they use to evaluate the conditional expectation in the asset pricing formula (1). HST showed that for empirically plausible parameterizations of model uncertainty, $m^u_{t+1}$ possesses substantial variability, raising the theoretical value of the equity premium, thereby helping to explain the equity premium puzzle. Below, we will define what is empirically plausible in terms of the probability of erroneously distinguishing among the alternative models to be described in the next section.

## 3 Three Salient Models

This section presents a brief overview of the robust decision theory that underlies the rest of the paper. Fear of model misspecification makes a decision maker want a decision rule to work well for a set of models. We consider a class of models indexed by a vector process $v_t$, with state $x_t$, control $u_t$, and i.i.d. Gaussian shock process $w_t$ with mean zero and identity covariance matrix:[8]

$$x_{t+1} = Ax_t + Bu_t + C\left[w_{t+1} + v_t\right].$$

We use the vector $v_t$ to represent model misspecifications around an *approximating* model; $v_t \equiv 0$ in the approximating model. We impose the following bound on the specification error:

$$\frac{1}{1-\beta}\mathbb{E}_{x_0}\left[\sum_{t=0}^{\infty}\beta^t v_t \cdot v_t\right] \leq \eta_0.$$

The parameter $\eta_0$ sets the average size of the potential model misspecifications where the average on the left side is taken across states and over time. Otherwise $v_t$ can feed back arbitrarily on the history of $x_t$. In this way, $v_t$ represents misspecified dynamics. The robustness parameter $\theta$ below can be interpreted as a Lagrange multiplier on the above constraint.[9]

Within this class of models, three are especially important:

- An unknown *true* model has $v_t = \bar{v}_t \neq 0$.
- An approximating model has $v_t = 0$.

- A constrained worst case model has $v_t = \hat{v}_t \neq 0$, where $\hat{v}_t$ is a process that depends on $\eta_0$.

The true model actually generates the data. The approximating model is the decision maker's model.[10] Figure 1 depicts these three models graphically. The worst case model $\hat{v}$ is created as a by product of the process of designing a rule to work well over the entire set of models in the circle.

We consider a decision maker who, when he fears no specification error (i.e., believes $v \equiv 0$), has preferences ordered by

$$V_0 = \mathbb{E} \sum_{t=0}^{\infty} \{-\beta^t R(x_t, u_t)\} \tag{3}$$

where $R(x, u)$ is a quadratic function. In (3), $\mathbb{E}$ is the mathematical expectation taken with respect to the approximating model. We want to evaluate (3) under a time-invariant decision rule $u = -Fx$. For fixed $F$, write the one-period return function $R_F(x) = R(x, -Fx)$.

For fixed $F$, we want to evaluate

$$V_F(x_0) = \mathbb{E}_{x_0} \sum_{t=0}^{\infty} \left[-\beta^t R_F(x_t)\right], \tag{4}$$

under the approximating model. Under the approximating model ($v_t = 0$), equation (3) can be evaluated as the fixed point of the recursion

$$V_F(x) = -R_F(x) + \beta \mathbb{E}_x V_F(x^*),$$

where the superscript $^*$ denotes a next period value and $\mathbb{E}_x$ is the conditional expectation evaluated with respect to the approximating model. This is an ordinary Bellman equation.

Now suppose we admit specification error, so that multiple models are in play, multiple probability distributions, with respect to each of which a mathematical expectation in (3) might be taken. We want a way to evaluate continuation utility that is conservative with respect to model misspecification, meaning that it admits the presence of multiple models. Anderson, Hansen, and Sargent (2000) construct a distorted expectations operator $\mathcal{R}$ that delivers a conservative evaluation of a next period continuation value and that serves as a constant in a robustness bound. It is conservative in the following sense. Let

$$\mathcal{R}(V) = \inf_{v} J(v) \equiv J(\hat{v}), \tag{5}$$

where

$$J(v) = \theta v'v + \mathbb{E}_x V(x^*), \tag{6}$$

$$x^* = A_o x + C(w + v), \tag{7}$$

$$\hat{v} = \theta^{-1}(I - \theta^{-1}C'\Omega C)^{-1}C'\Omega A_o x, \tag{8}$$

and where $A_o = A - BF$ and $x'\Omega x$ is part of the value function for the zero sum game defined by (25) and (20) below[11] Note that the dependence of $J(v)$ on $v$ comes through the distorted transition law (7) induced by $v$. The definition of inf in (5) implies that for any distortion $v$,

$$\mathbb{E}V[A_o x + C(w + v)] \geq J(\hat{v})(x) - \theta v'v.$$

The left side of this equation is the expectation of the one-period continuation value evaluated under a particular model indexed by the distortion $v$. The inequality thus bounds the rate at which performance deteriorates with respect to model misspecification as measured by $v'v$. Furthermore, under the approximating model ($v = 0$), $J(\hat{v}) = \mathcal{R}(V)$ gives a conservative estimate, i.e., a lower bound, of the one-period continuation value.[12]

# 4 Reformulating HST's Model

Our ultimate goal is to modify HST's model by concealing the state of the economy, thereby impelling the planner to estimate it. To accomplish this, it is convenient to rearrange HST's model to avail ourselves of the results of Hansen and Sargent (2000). We recount and recast HST's model.

## 4.1 HST's Model

This section describes HST's model, a linear quadratic stochastic growth model with a habit. A planner values a scalar process $s$ of consumption services according to

$$V_0 = \mathbb{E} \sum_{t=0}^{\infty} \beta^t \{-(s_t - \mu_b)^2\}. \tag{9}$$

The service $s$ is produced via the household technology

$$\begin{aligned} s_t &= (1+\lambda)c_t - \lambda h_{t-1}, \\ h_t &= \delta_h h_{t-1} + (1-\delta_h)c_t, \end{aligned} \tag{10}$$

where $\lambda \geq 0$ and $\delta_h \in (0,1)$, $c$ is a scalar consumption process, $\mu_b$ is a preference parameter governing curvature of the utility function, and $h$ is a scalar stock of household habits.[13] A linear technology converts a scalar endowment $d$ into consumption or capital:

$$\begin{aligned} k_t &= \delta_k k_{t-1} + i_t, \\ c_t + i_t &= \gamma k_{t-1} + d_t. \end{aligned} \tag{11}$$

Here $k_t, i_t, d_t$ are the capital stock, gross investment, and the exogenous stochastic endowment at time $t$, respectively. The parameter $\gamma$ is the constant marginal product of capital and $\delta_k$ is the depreciation factor for capital. Combining (11) leads to

$$c_t + k_t = Rk_{t-1} + d_t, \tag{12}$$

where $R = \gamma + \delta_k$. Relation (12) makes the gross return on a one-period risk-free asset be $R$.

HST assumed the following two-component model for the endowment:[14]

$$\begin{aligned} d_{t+1} &= \mu_d + d^1_{t+1} + d^2_{t+1}, \\ d^1_{t+1} &= g_1 d^1_t + g_2 d^1_{t-1} + c_1 w^1_{t+1}, \\ &\equiv (\phi_1 + \phi_2) d^1_t - \phi_1\phi_2 d^1_{t-1} + c_1 w^1_{t+1}, \\ d^2_{t+1} &= a_1 d^2_t + a_2 d^2_{t-1} + c_2 w^2_{t+1}, \\ &\equiv (\alpha_1 + \alpha_2) d^2_t - \alpha_1\alpha_2 d^2_{t-1} + c_2 w^2_{t+1} \end{aligned} \tag{13}$$

where $w_{t+1} = \begin{bmatrix} w^1_{t+1} \\ w^2_{t+1} \end{bmatrix}$ is an i.i.d. Gaussian disturbance vector with mean zero and identity covariance matrix. The two-component specification (13) allows separate permanent and transitory components of $d_t$, and is a specification often found in the micro literature on permanent income models.[15] HST also assumed that the planner observes current and lagged values of *both* components $d^i_t, i = 1, 2$, at all $t$. Later in this paper, we shall withdraw from the planner knowledge of the history of the individual components of the endowment process, and let only the history of their sum be observed.

## 4.2 Features of the HST Model

HST show that optimal consumption can be expressed

$$c_t = \frac{1}{1+\lambda}(\mu_b - \mu_{st}) + \frac{\lambda}{1+\lambda} h_{t-1}, \tag{14}$$

where $\mu_{st}$ is the shadow price of services in the planning problem. It obeys

$$\mu_{st} = \mu_b + \psi_0 \sum_{j=0}^{\infty} R^{-j} \mathbb{E}_t d_{t+j} + \psi_1 h_{t-1} + \psi_2 k_{t-1}. \tag{15}$$

HST show that (15) implies that

$$\mu_{st} = \mu_{st-1} + \nu' w_t, \tag{16}$$

where $\nu$ is a vector, so that $\mu_s$ is a martingale.

Equations (15) and (14) imply that $\mu_b$ has no effect on the allocation, because $\mu_b - \mu_{st}$ does not depend on $\mu_b$. However, $\mu_b$ does affect prices, including the market price of risk. HST show that the shadow price of consumption, $\mathcal{M}_t^c$, the marginal utility of consumption in the solution of the planning problem, satisfies

$$\mathcal{M}_t^c = (1+\lambda) + (1-\delta_h)\mathbb{E}_t \left[ \sum_{\tau=1}^{\infty} \beta^{\tau} \delta_h^{\tau} (-\lambda)(\mu_b - s_{t+\tau}) \right], \tag{17}$$

where $\mu_b - s_t = \mu_{st}$. The stochastic discount factor (without a preference for robustness) is

$$m_{t+1,t}^f = \beta \frac{\mathcal{M}_{t+1}^c}{\mathcal{M}_t^c}. \tag{18}$$

Finally, note that the coefficient of relative risk aversion for the one period utility function $-(s_t - \mu_b)^2$ is $\frac{s_t}{\mu_b - s_t}$.

## 4.3 Recasting the State Vector

The main purpose of this paper is to alter HST's model by changing assumptions about what the planner observes. To accomplish this, we first recast the model so that it conforms to a framework of Hansen and Sargent (2000) for getting robust solutions of joint filtering and control problems. To set HST's model into the Hansen and Sargent (2000) form, we redefine the state vector. Thus, we let the state vector be

$$x_t = \begin{bmatrix} h_{t-1} \\ k_{t-1} \\ d_{t-1} \\ 1 \\ d_t \\ d_t^1 \\ d_{t-1}^1 \end{bmatrix} \equiv \begin{bmatrix} f_t \\ y_t \\ z_t \end{bmatrix}, \tag{19}$$

with the partitioning of the state

$$f_t \equiv \begin{bmatrix} h_{t-1} \\ k_{t-1} \\ d_{t-1} \\ 1 \end{bmatrix}, \quad y_t \equiv d_t, \text{ and } z_t \equiv \begin{bmatrix} d_t^1 \\ d_{t-1}^1 \end{bmatrix}.$$

Please note that although $d_{2t}, d_{2t-1}$ are not explicitly included in the state vector, they can be recovered from the $d_t, d_{1t}$ components.

### 4.3.1 Reason for State Partitioning

We partitioned the state because we anticipate formulating a robust decision problem in which part of the state, namely $z_t$, is unobserved. Even with incomplete information, we assume that the first two components $f_t$ and $y_t$ are known to the decision maker or can be correctly inferred from current and past information. However, later we shall assume that the third component, $z_t$, consists of states that are hidden from the decision maker. The decision maker uses current and past data to make inferences about this vector. In many problems, there is a redundancy in the available information. For our prediction algorithms, it is important to eliminate redundant information. We accomplish this by eliminating $f_t$ from the information set. Current and past values of $y_t$ are sufficient to generate the current information set. Knowledge of $f_t$ or its history conveys no additional information.

In terms of the permanent income model, the partitioned law of motion can be written in the recursive form

$$\begin{bmatrix} f^* \\ y^* \\ z^* \end{bmatrix} = \begin{bmatrix} A_{ff} & A_{fy} & 0 \\ A_{yf} & A_{yy} & A_{yz} \\ A_{zf} & A_{zy} & A_{zz} \end{bmatrix} \begin{bmatrix} f \\ y \\ z \end{bmatrix} + \begin{bmatrix} B_f \\ 0 \\ 0 \end{bmatrix} u + \begin{bmatrix} 0 \\ C_y \\ C_z \end{bmatrix} w, \tag{20}$$

where superscript $^*$ denotes a next period value, $C_y = [c_1 \;\, c_2]$, and $C_z = \begin{bmatrix} c_1 & 0 \\ 0 & 0 \end{bmatrix}$.

Notice that $f^*$ is an exact function of $f$, current $y$, and the control $u$. No information is conveyed by the $f$ vector. Notice also that $C_y C_y'$ is nonsingular, so the entire $y^*$ vector is required to capture the arrival of new information next period. In what follows we will sometimes use the shorthand notation

$$x^* = Ax + Bu + Cw, \tag{21}$$

to depict the state evolution where

$$A = \begin{bmatrix} A_{ff} & A_{fy} & 0 \\ A_{yf} & A_{yy} & A_{yz} \\ A_{zf} & A_{zy} & A_{zz} \end{bmatrix} \equiv \begin{bmatrix} A_f \\ A_y \\ A_z \end{bmatrix}, \quad B = \begin{bmatrix} B_f \\ 0 \\ 0 \end{bmatrix}, \quad \text{and} \quad C = \begin{bmatrix} 0 \\ C_y \\ C_z \end{bmatrix}. \tag{22}$$

We can express the objective function (9) as

$$\mathbb{E} \sum_{t=0}^{\infty} \{\beta^t r(f_t, y_t, u_t)\}, \tag{23}$$

where

$$r(f, y, u) = -(f' \; y') R \begin{pmatrix} f \\ y \end{pmatrix} - u'Qu - 2u'W \begin{pmatrix} f \\ y \end{pmatrix}. \tag{24}$$

The objective function (23) does not depend directly on $z_t$. Instead $z_t$ enters the problem only as an information vector that helps predict $y_t$, which does appear in the objective function.

The robust control problem with objective (9) and transition law (20) is just a rewriting of HST's problem. They solved this problem using a robust decision theory, which we now briefly recount.

## 4.4 Robustness via a Two-Player Game

HST compute a robust decision rule by solving the two person game defined by the fixed point of

$$-x'\Omega x - a = \max_u \min_v \{r(f, y, u) - \beta \mathbb{E} x^{*\prime} \Omega x^* - \beta a + \beta \theta v'v\}, \tag{25}$$

subject to

$$x^* = Ax + Bu + C(w + v). \tag{26}$$

The equilibrium of the game is a pair of decision rules

$$\begin{aligned} u &= -Fx, \\ \hat{v} &= \kappa x, \end{aligned} \tag{27}$$

where $F$ and $\kappa$ are given by (B.6) to (B.9) in Appendix B, with volatility matrix $C$. The decision rule for $\hat{v}$ induces a 'worst case' adjustment to the conditional mean of the innovation $w$. In effect, a robust rule for $u$ is constructed by planning against this worst case $\hat{v}$. Please note that this worst case model is not the decision maker's model: his model has $v = 0$. The decision maker admits multiple models surrounding his approximating $v = 0$ model and doesn't know enough to unify the multiple models by choosing a unique prior distribution over them. The worst case model is simply a byproduct of the planning process.

### 4.5 Approximating and Distorted Models

The min-max decision theory leads to two salient models: the approximating and the distorted or worst-case model, both evaluated under the robust decision rule $u_t = -Fx_t$. The former becomes the economist's (and also the planner's and the agent's) model of the time series on quantities $(c_t, i_t)$; the latter gives the measure to be used for pricing risky securities. The distortion of the worst-case model vis a vis the approximating model boosts rates of return for risky assets, giving rise to 'Knightian uncertainty premia.'

Then under the control law $u = -Fx$, the *approximating model* is

$$f^* = (A_f - B_f F)x,$$
$$y^* = A_y x + C_y w,$$
$$z^* = A_z x + C_z w. \tag{28}$$

The *distorted* or worst case law of motion is

$$f^* = (A_f - B_f F)x,$$
$$y^* = (A_y + C_y \kappa)x + C_y w,$$
$$z^* = (A_z + C_z \kappa)x + C_z w. \tag{29}$$

Evidently, the distorted model can be obtained from the approximating model by displacing the zero conditional mean of $w_{t+1}$ in the approximating model by $\hat{v}_t = \kappa x_t$. The Radon-Nikodym derivative, or likelihood ratio, of the distorted conditional probability of $x_{t+1}$ with respect to the approximating conditional probability is

$$m^u_{t+1,t} = \frac{\exp[-\frac{1}{2}(w_{t+1} - \hat{v}_t)'(w_{t+1} - \hat{v}_t)]}{\exp[-\frac{1}{2}w'_{t+1}w_{t+1}]}, \tag{30}$$
$$m^u_{t+1,t} = \exp(w'_{t+1}\hat{v}_t - \frac{1}{2}\hat{v}'_t\hat{v}_t). \tag{31}$$

HST show that this Radon-Nikodym derivative is the market price of Knightian uncertainty that appears in the multiplicative adjustment of the stochastic discount factor

$$m_{t+1,t} = m^f_{t+1,t} m^u_{t+1,t},$$

where $m^f_{t+1,t} = \beta \frac{\mathcal{M}^c_{t+1}}{\mathcal{M}^c_t}$ is the "ordinary" ($\theta = +\infty$) stochastic discount factor without a preference for robustness. Here $\mathcal{M}^c_{t+1}$ is the shadow price of time $t+1$ consumption in the planning problem without a preference for robustness and $m^f_{t+1,t}$ is an intertemporal marginal rate of substitution between consumption rates at $t+1$ and $t$.

Evidently, $\theta$ is a critical parameter influencing $m^u_{t+1,t}$ through its impact on $\hat{v}_t$. For $\theta = +\infty$, there is no preference for robustness, $\kappa = 0$, and $m^u_{t+1,t} = 1$. Lowering $\theta$ increases the taste for robustness and allows $m^u_{t+1,t}$ to depart from unity and become stochastic and variable. This increases the volatility of the stochastic discount factor $m_{t+1}$.

We require a way of thinking about reasonable values of $\theta$. As we see in the next section, different settings of $\theta$ lead to different probabilities of detecting differences of the approximating model from the worst case model from a time series on $x_t$ of given length. We shall use the detection statistics to guide our setting of $\theta$.

# 5 Detection Error Probabilities

Anderson, Hansen, and Sargent (2000) link the preference-for-robustness parameter $\theta$ and detection error probabilities, a link that we shall use below to discipline our choice of plausible $\theta$'s. Detection error probabilities can be calculated using likelihood ratio tests. Thus, consider two alternative models. Model A is the approximating model, and model B is the distorted model associated with the worst case shock implied by $\theta$. Consider a fixed sample of observations. Let $L_{ij}$ be the likelihood of that sample for model $j$ assuming that model $i$ generates the data. Define the log likelihood ratio

$$r_i \equiv \log \frac{L_{ii}}{L_{ij}},$$

where $j \neq i,$ and $i = A, B$. Now consider the probabilities of two kinds of mistakes. First, assume that model A generates the data and calculate

$$p_A = \text{Prob}(\text{mistake}|A) = \text{freq}(r_A \leq 0).$$

Thus, $p_A$ is the frequency of negative log likelihood ratios $r_A$ when model A is true. Similarly, $p_B = \text{Prob}(\text{mistake}|B) = \text{freq}(r_B < 0)$ is the frequency of negative log likelihood ratios $r_B$ when model B is true. Call the probability of a detection error

$$p(\theta) = \frac{1}{2}(p_A + p_B). \tag{32}$$

Here, $\theta$ is the robustness parameter used to generate a particular model B. Appendix A shows in detail how to estimate the detection error probability by using simulations. We propose to set $p(\theta)$ to a reasonable number, then invert $p(\theta)$ to find a plausible value of $\theta$.

# 6 Observational Equivalence

We follow HST and define $\sigma \equiv -\theta^{-1}$; $\sigma$ is the risk-sensitivity parameter of Whittle (1990) and Jacobson (1973). HST's two-part empirical strategy rested on the fact that the likelihood function for quantity data $(c_t, i_t)$ has a ridge that makes $(\beta, \sigma)$ not separately identifiable. However, $(\beta, \sigma)$ pairs that are observationally equivalent for quantities can have very different implications for asset prices, as summarized by the market price of risk. HST's strategy was, first, to estimate the model's free parameters from quantity observations; and, second, to select a $(\beta, \sigma)$ pair from the likelihood function ridge that matches market based measures of the market price of risk.

The free parameters of HST's model are $[\lambda, \delta_h, \delta_k, \gamma, g_1, g_2, a_1, a_2, c_1, c_2]$ and a locus of $(\sigma, \beta)$ pairs. Using data on quantities $(c_t, i_t)$ alone, HST computed maximum likelihood estimates of these parameters for geometrically detrended quarterly U.S. time series from $1970I$ to $1996III$. HST proved:

*Observational Equivalence Proposition:*
Fix all parameters except $\beta$ and $\sigma$. Suppose $\beta R = 1$. There exists a $\underline{\sigma} < 0$ such that the optimal $(c, i)$ plan with $\sigma = 0$ is also the optimal $(c, i)$ plan for any $\sigma$ satisfying $\underline{\sigma} < \sigma \leq 0$ and a smaller discount factor $\hat{\beta}(\sigma)$ satisfying[16]

$$\hat{\beta}(\sigma) = \frac{1}{R} + \frac{\sigma \eta^2}{R - 1}, \tag{33}$$

where $\eta^2 = \nu \cdot \nu$, and $\nu$ is the vector that appears in the martingale representation (16) for the shadow price of consumption services $\mu_{st}$. Representation (16) comes from the solution of the planning problem when $\sigma = 0$.

Recall the decomposition of the stochastic discount factor

$$m_{t+1,t} = m^f_{t+1,t} m^u_{t+1,t},$$

where $m^f_{t+1,t} = \beta \frac{\mathcal{M}^c_{t+1}}{\mathcal{M}^c_t}$ is the "ordinary" ($\sigma = 0$) stochastic discount factor without a preference for robustness and $m^u_{t+1,t}$ is the likelihood ratio defined above. The marginal utility of consumption $\mathcal{M}^c_{t+1}$ is tied down by the quantities $(c_t, i_t, k_{t-1}, h_{t-1})$ and so is identical across observationally equivalent $(\beta, \sigma)$ pairs satisfying (33). However, $m^u_{t+1,t}$ *does* depend on $\sigma \equiv -\theta^{-1}$, through formula (30). Increasing the absolute value of $\sigma$ generally increases the norm of $\hat{v}_t$ and affects the stochastic discount factor.

It will ameliorate the equity premium puzzle[17] – the low theoretical volatility of the stochastic discount factor – if we can somehow increase the volatility of $m^u_{t+1,t}$. HST note that

$$\mathbb{E}[(m^u_{t+1,t})^2 | \mathcal{J}_t] = \exp(\hat{v}_t' \hat{v}_t).$$

Because $\mathbb{E}[m^u_{t+1,t} | \mathcal{J}_t] = 1$ by construction, it follows that the conditional standard deviation of $m^u_{t+1,t}$,

$$\sigma(m^u_{t+1,t} | \mathcal{J}_t) = \sqrt{\exp(\hat{v}_t' \hat{v}_t) - 1}. \tag{34}$$

HST call $\sigma(m^u_{t+1,t} | \mathcal{J}_t)$ the market price of Knightian uncertainty. The robustness parameter $\theta$ affects $\text{std}(m^u_{t+1,t} | \mathcal{J}_t)$ through $\hat{v}_t$.

In summary, in HST's model,

- Variations in the robustness parameter $\sigma$ have no effect on quantities, in the sense that there is an offsetting change in $\beta$ that leaves $F$ and all quantities unaltered.
- $(\beta, \sigma)$ pairs that are observationally equivalent for quantities affect the market price of risk through the market price of uncertainty (34).

# 7 Two Models with Filtering

We now turn to the main purpose of this paper. We modify one assumption in HST's model. We assume that the planner does not observe the entire state. In particular, we assume that the planner observes the history of $d_t$ but not its individual components. This assumption can be expressed by saying that the planner observes current and past values of $f, y$ but never sees $z$ in (20). Because $z$ contains information about future values of $y$, the planner is impelled to estimate $z$, and to base decisions on that estimate. The planner is induced jointly to solve robust control and filtering problems.

## 7.1 An Elementary Problem with Filtering

Hansen and Sargent (2000) show how to modify the two-player game (25) to incorporate unobserved elements of the state vector. They begin with an elementary formulation of a game that is designed to induce robust filtering and control, and show how that elementary game can be transformed to a simpler game, taking the form of (45), (46) via a two-step procedure involving a first step that solves a filtering problem.

Now the decision maker enters a period knowing the components of the state $f, y$ but having only an estimator $\check{z}$ of $z$, whose covariance matrix about $z$, $\Sigma$, is known. To express a preference for robust filtering and control, Hansen and Sargent consider the following dynamic game:

$$-\check{x}'\Omega\check{x} - a = \max_{u}\min_{v,v_z}\{r(f,y,u) - \beta\mathbb{E}\check{x}^{*\prime}\Omega^*\check{x}^* - \beta a^* + \beta\theta(v'v + v_z'v_z)\}, \tag{35}$$

subject to

$$x^* = Ax + Bu + C(w+v), \tag{36}$$

and

$$z = \check{z} + G_z(w_z + v_z). \tag{37}$$

Here $w_z$ is another i.i.d. Gaussian process, independent of $w$; $w_z$ has mean zero and identity covariance matrix; $w_z$ is the error in reconstructing the hidden part of the state. The matrix $G_z$ is a Cholesky factor of a covariance matrix $\Sigma \equiv \mathbb{E}(z - \check{z})(z - \check{z})'$, namely, $G_z G_z' = \Sigma$, and $\check{z}$ is an estimate of $z$ constructed from current and past observed values of $y$. This game assumes that the maximizing agent arrives at the current period with an estimate $\check{z}$ of the subcomponent $z$ of the state $x$. To promote robustness, the game also lets the minimizing agent distort the conditional mean $v_z$ of the state-reconstruction error $w_z$, allowing it to depend on the history of the state. One step of minimizing and maximizing in (35) will 'backdate' the value function as parameterized by $\Omega^*, a^*$ and 'update' the factored covariance matrix $G_z$.

Thus, this game produces

A. A *backward* (in time) recursion mapping $\Omega^*$ into $\Omega$ and $a^*$ into $a$.

B. An estimator $\check{z}^*$ of next period's hidden state $z^*$.

C. A *forward* (in time) recursion mapping $\Sigma$ into $\Sigma^*$, which generates a covariance matrix to be used for next period's version of the problem.

D. A robust adjustment to the estimate of the current state $z$.

Building on work of Basar and Bernhard (1995), Hansen and Sargent (2000) show that item C is the same recursion associated with an ordinary Kalman filter, and that $\check{z}^*$ from item B is the ordinary Kalman filter estimate of the state. Thus, the ordinary Kalman filter solves a filtering problem that embeds a preference for robustness. Although the Kalman filter is used to construct $\check{z}$ given current and past data on $y$, item D makes a conservative adjustment in the estimated $z$ aimed at making the control law more robust.

## 7.2 Interactions of Filtering and Decisions

Hansen and Sargent (2000) show that (35), (36), (37) can be reformulated in terms of an ordinary Kalman filtering problem and a particular ordinary robust control problem without filtering. In particular, they show that the solution of (35), (36), (37) can also be obtained via the following three-step procedure:

- Step 1: For the purpose of solving the filtering part of the problem, form the small state space system:

$$
\begin{aligned}
z_{t+1} &= A_{zz} z_t + C_z w_{t+1}, \\
y_{t+1} &= A_{yz} z_t + C_y w_{t+1}.
\end{aligned} \tag{38}
$$

Form the ordinary Kalman filter for the system, i.e., the Kalman filter for the system matrices[18]

$$
[A_{zz}, C_z, A_{yz}, C_y, C_z C_y'].
$$

In particular, solve the Ricatti equation for $\Sigma \equiv E(z - \check{z})(z - \check{z})'$,

$$
\Sigma = [A_{zz}\Sigma A_{zz}' + C_z C_z'] - [A_{zz}\Sigma A_{yz}' + C_z C_y'] \times [A_{yz}\Sigma A_{yz}' + C_y C_y']^{-1} [A_{zz}\Sigma A_{yz}' + C_z C_y']'. \tag{39}
$$

Form the Kalman gain

$$
K = [A_{zz}\Sigma A_{yz}' + C_z C_y'] \times [A_{yz}\Sigma A_{yz}' + C_y C_y']^{-1}.
$$

Define the covariance matrix of errors in forecasting $\begin{bmatrix} y_{t+1} \\ z_{t+1} \end{bmatrix}$ from $\{y_s, s \leq t\}$,

$$
\Lambda = \begin{bmatrix} A_{yz}\Sigma A_{yz}' + C_y C_y' & A_{yz}\Sigma A_{zz}' + C_y C_z' \\ A_{zz}\Sigma A_{yz}' + C_z C_y' & A_{zz}\Sigma A_{zz}' + C_z C_z'; \end{bmatrix}. \tag{40}
$$

Factor $\Lambda$ according to

$$\Lambda = \begin{bmatrix} \check{C}_y & 0 \\ \check{C}_z & \tilde{C}_z \end{bmatrix} \begin{bmatrix} \check{C}_y & 0 \\ \check{C}_z & \tilde{C}_z \end{bmatrix}' \equiv \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix}, \tag{41}$$

where $\check{C}_y$ is the Cholesky factor of $\Lambda_{11}$, $\check{C}_z = K\check{C}_y$, and $\tilde{C}_z$ is the Cholesky factor of $[\Lambda_{22} - \Lambda_{21}\Lambda_{11}^{-1}\Lambda_{12}]$. Note that $\Sigma = \Lambda_{22} - \Lambda_{21}\Lambda_{11}^{-1}\Lambda_{12}$ and that the Kalman gain is $K = \Lambda_{21}{\Lambda_{11}}^{-1}$. By construction, $\check{C}_y$ and $\tilde{C}_z$ are nonsingular.

- Step 2: Write the state evolution equation as:[19]

$$\begin{aligned} x^* &= A\check{x} + Bu + Cw + A(x - \check{x}), \\ &= A\check{x} + Bu + C^* w^*, \end{aligned}$$

where

$$C^* = \begin{bmatrix} 0 & 0 \\ \check{C}_y & 0 \\ \check{C}_z & \tilde{C}_z \end{bmatrix},$$

and $w^*$ is a normally distributed vector with mean zero and covariance matrix $I$, which we partition as:

$$w^* = \begin{bmatrix} \check{w} \\ \tilde{w} \end{bmatrix}. \tag{42}$$

The vector $w^*$ is the shock in an innovations representation for the $(y, z)$ process.[20] Note that the dimension of the composite shock $w^*$ is $1 + 2 = 3$, where 1 is the dimension of $y$ and 2 is the dimension of $z$. Recall that the dimension of $w$ in the original transition law (26) with full state observation was 2.

We can use the process $w^*$ to form a law of motion for the predicted state. By construction, the shock $\breve{w}$ is in the information set of the decision-maker next period: it is revealed by $y^*$ (remember that $\check{C}_y$ is by construction nonsingular). Also, by construction $\breve{w}$ is independent of $\tilde{w}$. Therefore, the law of motion for the predicted state is obtained by replacing $\tilde{w}$ with zero in the following representation:

$$\breve{x}^* = A\breve{x} + Bu + \check{C}\breve{w}, \tag{43}$$

where

$$\check{C} = \begin{bmatrix} 0 \\ \check{C}_y \\ \check{C}_z \end{bmatrix}.$$

The $f^*$ and $y^*$ components of $x^*$ match those for $x^*$ because both components are in the decision maker's information set tomorrow. However, $z^*$ and $\breve{z}^*$ will differ. For future reference, we also define

$$\tilde{C} = \begin{bmatrix} 0 \\ 0 \\ \tilde{C}_z \end{bmatrix}, \tag{44}$$

and note that $x^* = \breve{x} + \tilde{C}\tilde{w}$.

- Step 3: Compute the decision rule for $u$ that solves

$$-\check{x}'\Omega\check{x} - a = \max_{u} \min_{\check{v},\, \tilde{v}} \{r(f, y, u) - \beta \mathbb{E}\check{x}^{*'}\Omega\check{x}^* - \beta a + \beta\theta(\check{v}'\check{v} + \tilde{v}'\tilde{v})\}, \tag{45}$$

subject to

$$\check{x}^* = A\check{x} + Bu + \check{C}(\check{w} + \check{v}) + \tilde{C}\tilde{v}. \tag{46}$$

In this game, the composite vector $w^*$ disguises model misspecification. The two-dimensional misspecification term $\tilde{v}$ appears in the evolution for the *predicted* state $\check{x}^*$, but is hidden in the evolution for the *actual* state vector $x^*$. The predicted state $\check{z}^*$ is created by the agent and not directly observed. The $\tilde{v}$ misspecification appears in the agent's perception of how $z^*$ will evolve and is thereby transmitted into how $\check{z}^*$ is constructed.

As mentioned above, Hansen and Sargent (2000) derive the three-step procedure (45)-(46) from the more elementary recursive specification of a game, (35), (36), (37), that involves both the unknown state and the control. Several things about this procedure are remarkable. First, filtering is done using an ordinary (*i.e.*, non-robust) Kalman filter.[21] Second, the two-player game (45)–(46) is associated with an ordinary robust decision problem that treats the state as observed and given by

$$\begin{bmatrix} y' & f' & \check{z}' \end{bmatrix}.$$

Third, there is an interaction between the filtering problem and the control problem due to robustness. The interaction comes from the presence of the term $\tilde{C}\tilde{v}$, which captures the ability of the minimizing agent to deceive the maximizing agent by altering the gap between the estimated and actual value of the unobserved part of the state $z$. Below, we shall expand upon this third point by describing a comparison model that, inappropriately according to the elementary recursive game that induces (45)–(46), ignores this avenue of deception.

A solution of (45) and (46) is a decision rule

$$u_2 = -F_2 \check{x},$$

and laws of motion for the worst case means

$$\check{v}_2 = \check{\kappa}_2 \check{x},$$

$$\tilde{v}_2 = \tilde{\kappa}_2 \check{x}, \tag{47}$$

where coefficients $F_2$ and $\kappa_2 \equiv \begin{bmatrix} \check{\kappa}_2 \\ \tilde{\kappa}_2 \end{bmatrix}$ are determined by equations (B.6) to (B.9) in Appendix B with volatility matrix

$$C_2 = \begin{bmatrix} 0 & 0 \\ \check{C}_y & 0 \\ \check{C}_z & \tilde{C}_z \end{bmatrix}.$$

We can use these worst case means to form the distorted law of motion to be used for asset pricing and detection error probabilities. Thus, the approximating model under the robust rule is

$$\check{x}^* = (A - BF_2)\check{x} + \check{C}\check{w}, \tag{48}$$

and

$$x^* = \check{x}^* + \tilde{C}\tilde{w}. \tag{49}$$

The distorted model under the robust rule is

$$\check{x}^* = (A - BF_2 + \check{C}\check{\kappa}_2 + \tilde{C}\tilde{\kappa}_2)\check{x} + \check{C}\check{w}, \tag{50}$$

or

$$\begin{aligned} \check{x}^* &= (A - BF_2)\check{x} + \check{C}\check{w}, \\ x^* &= \check{x}^* + \tilde{C}\tilde{w} + \check{C}\check{v} + \tilde{C}\tilde{v}. \end{aligned} \tag{51}$$

These are the representations that we need to calculate detection error probabilities and the market price of uncertainty.

## 7.3 Comparison Model

To highlight an interaction between filtering and control, we display another game that emerges from ignoring that interaction. This game is formed by the following two step procedure.

- Step 1. Perform steps 1 and 2 above.
- Step 2. Solve a recursive game (45) where the extremization is now subject to the transition equation

$$\check{x}^* = A\check{x} + Bu + \check{C}(\check{v} + \check{w}), \tag{52}$$

The difference between (46) and (52) is the absence of $\tilde{v}$ from the latter. The elementary recursive game referred to above directs Hansen and Sargent (2000) to include this term in (46). This term embodies an interaction between filtering and control for inducing robustness.

Notice that game (45)–(52) comes from replacing the original transition equation for $y$ in (20) with the ordinary Kalman filter 'innovations representation' for $y$, then treating the innovations representation as though it were the original model of the $y$

process in HST. This pushes the original representation of the $y$ process in (25) into the background and replaces it with another that ignores its hidden state structure, then proceeding as in HST. The robust decision rule and the worst case means are solved by (B.6) to (B.9) in Appendix B by setting the volatility matrix equal to

$$\check{C} = \begin{bmatrix} 0 \\ \check{C}_y \\ \check{C}_z \end{bmatrix}.$$

This two-step procedure without the interaction term was appropriate in analyses like those of Detemple (1986), Dothan and Feldman (1986), Gennotte (1986) and Veronesi (1999), that study asset pricing in the face of filtering without a preference for robustness. With a preference for robustness, the procedure is not correct.

We call (45)–(52) the 'comparison model'. Although Hansen and Sargent (2000) show that it does not give the robust solution to the joint filtering and control problem, we compute market prices of risk and detection error probabilities for the comparison model as well as for (45)–(46).

# 8 Market Price of Uncertainty under Filtering

This section and Appendix C describe how to compute market prices of uncertainty. We extend HST's calculations to pricing multi-period returns.

## 8.1 One-Period Market Price of Uncertainty

We can compute the market price of uncertainty by again using a Radon-Nikodym derivative of the distorted model of $x^*$ respect to the approximating model. Write:

$$x^* = \check{x}^* + \tilde{C}\tilde{w}.$$

Form

$$m^{u*} = \exp(w^* \cdot v^* - \frac{1}{2} v^* \cdot v^*),$$

where $v^* = \begin{bmatrix} \check{v} \\ \tilde{v} \end{bmatrix}$. While this will generate the correct pricing formulas, we can also use the conditional expectation

$$\mathbb{E}[m^{u*} | \check{w}, v^*] = \exp(\check{w} \cdot \check{v} - \frac{1}{2} \check{v} \cdot \check{v}),$$

since $\check{w}$ is the innovation to the information set of economic agents. Below, inspired by (2), we compute the conditional standard deviation of $m^u$ to measure the boost in the market price of risk contributed by uncertainty aversion.

We extend the calculations to multi-period returns because the effects of filtering on prices of risk operate through $\tilde{v}$ and appear only in prices of multi-period returns.

## 8.2 Multi Period Market Prices of Uncertainty

To derive formulas for multi-period market prices of uncertainty with filtering, we impose the permanent income control law and let the resulting state evolution under the approximating model with filtering be

$$\check{x}_{t+1} = A^* \check{x}_t + \check{C} \check{w}_{t+1},$$

and under the (constrained) worst-case model

$$\check{x}_{t+1} = \check{A} \check{x}_t + \check{C} \check{w}_{t+1}.$$

Here $A^* = A - BF$ and $\check{A} = A - BF + \check{C}\check{\kappa}_2 + \tilde{C}\check{\kappa}_2$, so that $\check{A}$ includes captures the feedback of both $\check{v}$ and $\tilde{v}$ on the state.

We want to form the ratio of conditional densities for the observed state vector

$$y^j_{t+j} = \begin{bmatrix} y_{t+1} \\ y_{t+2} \\ \vdots \\ y_{t+j} \end{bmatrix},$$

under the two models for each $j$. To represent this ratio, we construct the conditional means and shock weighting matrices for $y^j_{t+j}$ in terms of the composite shock vector

$$\check{w}^j_{t+j} = \begin{bmatrix} \check{w}_{t+1} \\ \check{w}_{t+2} \\ \vdots \\ \check{w}_{t+j} \end{bmatrix}.$$

Then we can write

$$y^j_{t+j} = H^*_j \check{x}_t + G^*_j \check{w}^j_{t+j},$$

under the approximating model and

$$y^j_{t+j} = \check{H}_j \check{x}_t + \check{G}_j \check{w}^j_{t+j},$$

under the worst case model for some matrices $H^*_j$, $\check{H}_j$, $G^*_j$ and $\check{G}_j$. Form the likelihood ratio:

$$m^u_{t+j,t} = \frac{|\det(G^*_j)|}{|\det(\check{G}_j)|} \frac{\exp\left\{-\frac{1}{2}\left[(\check{G}_j)^{-1}(y^j_{t+j} - \check{H}_j\check{x}_t)\right] \cdot \left[(\check{G}_j)^{-1}(y^j_{t+j} - \check{H}_j\check{x}_t)\right]\right\}}{\exp\left\{-\frac{1}{2}\left[(G^*_j)^{-1}(y^j_{t+j} - H^*_j\check{x}_t)\right] \cdot \left[(G^*_j)^{-1}(y^j_{t+j} - H^*_j\check{x}_t)\right]\right\}}.$$

Since we evaluate this under the approximating model, we can write

$$(G_j^*)^{-1}(y_{t+j}^j - H_j^* \check{x}_t) = \check{w}_{t+j}^j,$$

and

$$\begin{aligned}(\check{G}_j)^{-1}(y_{t+j}^j - \check{H}_j \check{x}_t) &= (\check{G}_j)^{-1} G_j^* (G_j^*)^{-1}(y_{t+j}^j - H_j^* \check{x}_t + H_j^* \check{x}_t - \check{H}_j \check{x}_t), \\ &= (\check{G}_j)^{-1} G_j^* [\check{w}_{t+j}^j - (G_j^*)^{-1}(\check{H}_j - H_j^*)\check{x}_t],\end{aligned}$$

and substitute this into the likelihood ratio. In Appendix C, we obtain a formula for $m_{t+j,t}^u$ from which we can readily compute $\sigma_t(m_{t+j,t}^u)$, which is the $j$-period market price of Knightian uncertainty. We proceed to construct recursions for $\check{H}_j, H_j^*, \check{G}_j, G_j^*$.

## 8.3 Conditional Means

Consider first the recursive construction of the conditional mean matrices. Let $U = \begin{bmatrix} 0_{1\times 4} & 1 & 0_{1\times 2} \end{bmatrix}$ denote a selection matrix designed so that $y_{t+j} = U x_{t+j}$. Let $\check{H}_1 = U\check{A}$, and use the recursion

$$\check{H}_{k+1} = \begin{bmatrix} \check{H}_k \\ U(\check{A})^{k+1} \end{bmatrix}$$

to construct $\check{H}_j$. Then the conditional mean for $y_{t+j}^j$ is $\check{H}_j \check{x}_t$, which captures the contributions of both $\check{v}$ and $\tilde{v}$. Form $H_j^*$ analogously with $A^*$ used in place of $\check{A}$ for the approximating model.

## 8.4 Shock Dependence

Consider next the recursive construction of the matrices encoding shock dependence. Let $\check{C}_1 = \check{C}$, $C_1^* = \check{C}$ and define $\check{C}$ and $G^*$ recursively as follows,

$$\check{C}_{k+1} = [\check{A}\check{C}_k \vdots \check{C}],$$

$$C_{k+1}^* = [A^* C_k^* \vdots \check{C}]. \tag{53}$$

Using these matrices and the facts that $\check{G}_1 = U\check{C}$ and $G_1^* = U\check{C}$ as inputs into the recursion, we have

$$\check{G}_{k+1} = \begin{bmatrix} \check{G}_k & 0 \\ U\check{A}\check{C}_k & \check{G}_1 \end{bmatrix},$$

$$G_{k+1}^* = \begin{bmatrix} G_k^* & 0 \\ UA^*C_k^* & G_1^* \end{bmatrix}. \tag{54}$$

# 9 Results

This section presents estimates of market prices of Knightian uncertainty for three models: HST's, ours, which we dub the HSW model, and the comparison model. The first assumes that both components of the endowment process are observed, while the second and third assume that only the sum is observed, impelling agents to filter. The HSW model takes into account the interaction between filtering and decision making under a preference for robustness, while the comparison model suppresses that interaction.

## 9.1 HST Empirical Procedure

HST estimated the identifiable parameters by maximizing a Gaussian likelihood function. They estimated the model from geometrically detrended time series on $c_t, i_t$.

They found that given the parameters, the model reveals time series of the two components $d_t^1, d_t^2$ of the endowment shock. They recovered those two components and used them to construct the state $x_t$ for computing the mean distortion $\hat{v}_t$ and $m_{t+1,t}^u$.

## 9.2 Filtering the Endowment Process

In this paper we use HST's parameter estimates, but want to assume that the planner and agents do not see the components $d_t^i$, only their sum $d_t$, up to a constant. We form $d_t$ as the sum of the components $d_t^1 + d_t^2$ recovered by HST, then use the Kalman filter to construct filtered estimates of the components based on the history of the sum $d_t$ up to time $t$. In that way, we form $\check{z}_t$ as a component of $\check{x}_t$. We then use $\check{x}_t$ to form $\check{v}_t, \tilde{v}_t$ and $m_{t+1,t}^u$. In more detail, we form $\check{z}_{t+1}$ recursively from

$$\check{w}_{t+1} = \check{C}_y^{-1}(y_{t+1} - A_y \check{x}_t), \tag{55}$$

$$\check{z}_{t+1} = A_z \check{x}_t + \check{C}_z \check{w}_{t+1}. \tag{56}$$

Here $\check{C}_y^{-1} w_{t+1}$ is the innovation in $y_{t+1}$. This is a standard recursive application of the Kalman filter to construct state estimates.

Figure 2 shows the two components of the endowment process recovered by HST. Figure 3 shows the filtered estimates of these two components. Not surprisingly, the filtered components are smoother than their true counterparts. Below, we calculate $m_{t+1,t}^u$ based on these filtered components.

Table 1 reports HST's estimates of the free parameters of their model with habit persistence. For those parameters, Figure 4 shows the locus of $(\beta, \sigma)$ pairs that are observationally equivalent for HST's model, the HSW model, and the comparison model. These were computed by evaluating the exact formula (33). The locus for the comparison model is virtually identical with that for HST's model, while the locus for the HSW model is steeper, reflecting the larger innovation 'volatility' coming from $\tilde{C}_z$. By construction, all three loci go through the same point at $\sigma = 0$.

Table 2 computes the median market prices of risk from one to four periods for HST's model for some combinations[22] of parameter values $(\mu_b, \sigma)$. The preference specification makes $\mu_b$ a curvature parameter. Table 8 reports coefficients of relative risk aversion associated with various values of $\mu_b$, which we formed by injecting the derivatives of the utility function $u_{cc}$ and $u_c$ injected into the standard formula for the coefficient of relative risk aversion:

$$r_c \equiv \frac{-cu_{cc}}{u_c}. \tag{57}$$

We apply the chain rule to calculate risk aversion coefficient defined in terms of consumption, as in (57). The marginal utility of consumption is related to that for services by $u_c = (1+\lambda)u_s = -2(1+\lambda)(s-\mu_b)$. and The second derivative of the utility function with respect to consumption is $u_{cc} = -2(1+\lambda)^2$. Therefore, we take the coefficient of relative risk aversion for consumption gambles to be

$$r_c = c\frac{1+\lambda}{\mu_b - s}.$$

We evaluated (57) along the $c, s$ realization of HST. Table 8 records various quantiles of the resulting coefficients of relative risk aversion for different bliss points $\mu_b$.[23]

In tables 3, 4, 5 and 6, as $\sigma$ varies, we alter $\beta$ according to the observational equivalence formula (33). That $(\beta, \sigma)$ respect the observational equivalence formula (33) implies that $F$ stays fixed for all three models and all values of $\sigma$ (this is what the observational equivalence proposition means). However, the worst case shock $v$ varies with $\sigma$, and across models, because the volatility matrices (the $C$'s) vary across models and also time series of the state vector.[24] Tables 3 to 6 report the market prices of uncertainty for 1 to 4 periods for all three models.[25]

Consider a comparison between the HSW and benchmark models. Recall that the difference between the games underlying these models is that the HSW game has an additional perturbation not present in the benchmark model. It requires more than one time period for this perturbation to be reflected in the market price of uncertainty. For a given choice of robustness penalty parameter $\theta = -1/\sigma$, there is virtually no difference in the one period median market prices of uncertainty between the HSW game and the benchmark game. The additional perturbation, however, enhances the uncertainty prices for longer time horizons in the HSW model. For instance, Table 4 shows that that the HSW model leads to a fifty percent increase in the market price of Knightian uncertainty for horizon four. However, the meaning of $\theta$ or $\sigma$ is different across models. For a given $\sigma$, the worst-case model associated with the HSW model is, from a statistical vantage point, further away from the approximating model than is the worst-case model that is associated with the benchmark model. An agent who learns statistically may more readily detect such model departures from historical data. We now study detection probabilities in more detail.

For each of our three models, Table 7 records the detection error probabilities for distinguishing the approximating model from the worst case model affiliated with a given $\sigma$. Each of these was calculated by counting frequencies from 20,000 simulations of the detection error statistics described in Appendix A. Each simulation started from HST's estimate of the initial condition for the state, and contained the same number of periods as the data set that HST used to estimate their model. For a given $\sigma$, the detection error probability is lower for the HSW model than for the HST model, meaning that it is easier to distinguish the worst case model from the approximating model in the HSW case. In figures 5, 6, 7 and 8, we plot the relationship between the market price of Knightian uncertainty and the detection error probability, using statistics from tables 3, 4, 5 and 6 and 7. For each model for each pricing horizon, there is a tight inverse relation between the detection error probability and the market price of uncertainty. For the shorter pricing horizons, the market price of uncertainty is actually lower for a given detection error probability for the HSW model than for the other models. At horizon 4, however, the loci of detection-error probabilities and market prices of uncertainty of all three models coincide. The graph at horizon 4 demonstrates that the link between the detection error probabilities and the market prices of uncertainty that was discussed and documented in Anderson, Hansen, and Sargent (2000) extends to at least some models with hidden Markov states, provided that we look beyond the initial response.

## 9.3 Detection Error Probabilities and Model Complexity

The last subsection indicated that for a given detection error probability, all three models give rise to nearly the same market prices of uncertainty for the four period pricing horizon. We suspect that it is not really appropriate to compare detection-error probabilities as we have across different approximating models. Those models are associated with games that assume different types of perturbations. The worst-case model for the HSW game was derived by looking at more complicated perturbations than those allowed for the benchmark game. In studying detection, we explored only pairwise comparisons between worst-case models and the approximating model. In statistically exploring a richer class of perturbations, it may instead be reasonable to imagine detection problems with a more complicated family of alternative models. Enlarging the family of alternative models might make statistical detection more challenging. Our current detection comparison misses the additional complexity that emerges from adding candidate models into the choice set of a hypothetical statistician. When some of the state variables are hidden from decision makers, taking account of this complexity might well boost the market price of uncertainty.

## 10 Concluding Remarks

This paper has shown how to adapt the asset pricing theory of HST to a setting where part of the state is not observed, putting the planner and the agents into a situation where they have both to filter and to control. By using results of Hansen and Sargent (2000), the joint filtering and control problem can be broken in two, the first being an ordinary Kalman filtering problem, and the second being an ordinary robust control problem with observed state. HST's formulation of asset pricing then applies directly, including their formula for the market price of Knightian uncertainty in terms of the Radon-Nikodym derivative of the distorted with respect to the approximating model. This two step procedure still embodies an interaction between filtering and control that is captured by an extra innovation volatility term in the control problem relative to what is found in the non-robust formulations of related problems by Detemple (1986), Dothan and Feldman (1986), Gennotte (1986) and Veronesi (1999) and others.

We used detection error probabilities to discipline our choice of the critical robustness parameter $\sigma = -\theta^{-1}$ across models. For fixed detection error probabilities, we find that the market price of risk measured using the approximating model does not increase in moving from HST's specification to ours. The explanation is this. For fixed $\sigma$, the added confusion caused by the filtering problem increases the gap between the distorted and the approximating model by enlarging the mean distortion $v_t$, making deviations between the approximating and distorting models easier to detect statistically. Adjusting $\sigma$ toward zero to compensate for this effect erases much of the boost in the market price of risk coming from the increased volatility from filtering.

However, we are doubt this apparent irrelevance result because in comparing detection errors across models it may be important to adjust the likelihood ratios for the differing complexities of the models. We suspect that adjusting for model complexity would alter our interpretation of the above findings.

We intend this paper partly as a prolegomenon to a paper in which we alter the specification of the trend in HST's model. Instead of positing a known geometric trend, we would like to work with a stochastic trend model, say by letting the endowment process have repeated unit roots. That specification is capable of matching 'trend breaks' in productivity growth. The filtering machinery in this paper then applies directly to the problem of estimating an unobserved trend component of GDP growth, allowing for breaks. HST's model could be re-estimated under such a modification.

The joint robust filtering and control problem has many potential applications in macroeconomics and monetary economics. A class of examples that especially interests us has stochastic unobserved trends in productivity or 'potential GDP', estimates of which enter monetary policy rules. See Cagetti, Hansen, Sargent, and Williams (2000) for a formulation in continuous time.

Table 1: Parameter Estimates from HST

| | |
|---|---|
| $\beta$ | 0.9971 |
| $\delta_h$ | 0.6817 |
| $\lambda$ | 2.4433 |
| $\alpha_1$ | 0.8131 |
| $\alpha_2$ | 0.1888 |
| $\phi_1$ | 0.9978 |
| $\phi_2$ | 0.7044 |
| $\mu_d$ | 13.7099 |
| $c_1$ | 0.1084 |
| $c_2$ | 0.1551 |

Table 2: Multi-Period Market Price of Model Uncertainty (with habit persistence)

Panel A: 1 period

| $\mu_b \diagdown \sigma$ | 0 | -0.000025 | -0.00005 | -0.000075 | -0.00010 | -0.00015 |
|---|---|---|---|---|---|---|
| 24 | 0 | 0.0174 | 0.0348 | 0.0523 | 0.0697 | 0.1048 |
| 30 | 0 | 0.0284 | 0.0568 | 0.0853 | 0.1140 | 0.1718 |
| 36 | 0 | 0.0394 | 0.0789 | 0.1186 | 0.1586 | 0.2399 |

Panel B: 2 period

| $\mu_b \diagdown \sigma$ | 0 | -0.000025 | -0.00005 | -0.000075 | -0.00010 | -0.00015 |
|---|---|---|---|---|---|---|
| 24 | 0 | 0.0246 | 0.0493 | 0.0740 | 0.0989 | 0.1491 |
| 30 | 0 | 0.0402 | 0.0805 | 0.1211 | 0.1620 | 0.2454 |
| 36 | 0 | 0.0557 | 0.1118 | 0.1685 | 0.2260 | 0.3450 |

Panel C: 3 period

| $\mu_b \diagdown \sigma$ | 0 | -0.000025 | -0.00005 | -0.000075 | -0.00010 | -0.00015 |
|---|---|---|---|---|---|---|
| 24 | 0 | 0.0302 | 0.0604 | 0.0909 | 0.1215 | 0.1835 |
| 30 | 0 | 0.0492 | 0.0987 | 0.1487 | 0.1994 | 0.3035 |
| 36 | 0 | 0.0683 | 0.1372 | 0.2073 | 0.2790 | 0.4298 |

Panel D: 4 period

| $\mu_b \diagdown \sigma$ | 0 | -0.000025 | -0.00005 | -0.000075 | -0.00010 | -0.00015 |
|---|---|---|---|---|---|---|
| 24 | 0 | 0.0348 | 0.0699 | 0.1051 | 0.1407 | 0.2131 |
| 30 | 0 | 0.0569 | 0.1142 | 0.1722 | 0.2314 | 0.3540 |
| 36 | 0 | 0.0789 | 0.1588 | 0.2405 | 0.3248 | 0.5049 |

Table 3: One-Period Median Market Price of Model Uncertainty (with habit persistence)

Panel A: The HST model

| $\mu_b \diagdown \sigma$ | 0 | -0.000025 | -0.00005 | -0.000075 |
|---|---|---|---|---|
| 24 | 0 | 0.0174 | 0.0348 | 0.0523 |
| 30 | 0 | 0.0284 | 0.0568 | 0.0853 |
| 36 | 0 | 0.0394 | 0.0789 | 0.1186 |

Panel B: The comparison model

| $\mu_b \diagdown \sigma$ | 0 | -0.000025 | -0.00005 | -0.000075 |
|---|---|---|---|---|
| 24 | 0 | 0.0175 | 0.0350 | 0.0525 |
| 30 | 0 | 0.0285 | 0.0570 | 0.0857 |
| 36 | 0 | 0.0395 | 0.0792 | 0.1191 |

Panel C: The HSW model

| $\mu_b \diagdown \sigma$ | 0 | -0.000025 | -0.00005 | -0.000075 |
|---|---|---|---|---|
| 24 | 0 | 0.0175 | 0.0350 | 0.0526 |
| 30 | 0 | 0.0285 | 0.0571 | 0.0858 |
| 36 | 0 | 0.0396 | 0.0793 | 0.1193 |

Table 4: Two-Period Median Market Price of Model Uncertainty (with habit persistence)

Panel A: The HST model

| $\mu_b \diagdown \sigma$ | 0 | -0.000025 | -0.00005 | -0.000075 |
|---|---|---|---|---|
| 24 | 0 | 0.0246 | 0.0493 | 0.0740 |
| 30 | 0 | 0.0402 | 0.0805 | 0.1211 |
| 36 | 0 | 0.0557 | 0.1118 | 0.1685 |

Panel B: The comparison model

| $\mu_b \diagdown \sigma$ | 0 | -0.000025 | -0.00005 | -0.000075 |
|---|---|---|---|---|
| 24 | 0 | 0.0247 | 0.0495 | 0.0744 |
| 30 | 0 | 0.0403 | 0.0808 | 0.1216 |
| 36 | 0 | 0.0559 | 0.1122 | 0.1692 |

Panel C: The HSW model

| $\mu_b \diagdown \sigma$ | 0 | -0.000025 | -0.00005 | -0.000075 |
|---|---|---|---|---|
| 24 | 0 | 0.0310 | 0.0622 | 0.0935 |
| 30 | 0 | 0.0506 | 0.1016 | 0.1531 |
| 36 | 0 | 0.0702 | 0.1412 | 0.2135 |

Table 5: Three-Period Median Market Price of Model Uncertainty (with habit persistence)

Panel A: The HST model

| $\mu_b \setminus \sigma$ | 0 | -0.000025 | -0.00005 | -0.000075 |
|---|---|---|---|---|
| 24 | 0 | 0.0302 | 0.0604 | 0.0909 |
| 30 | 0 | 0.0492 | 0.0987 | 0.1487 |
| 36 | 0 | 0.0683 | 0.1372 | 0.2073 |

Panel B: The comparison model

| $\mu_b \setminus \sigma$ | 0 | -0.000025 | -0.00005 | -0.000075 |
|---|---|---|---|---|
| 24 | 0 | 0.0303 | 0.0607 | 0.0912 |
| 30 | 0 | 0.0494 | 0.0991 | 0.1493 |
| 36 | 0 | 0.0686 | 0.1378 | 0.2082 |

Panel C: The HSW model

| $\mu_b \setminus \sigma$ | 0 | -0.000025 | -0.00005 | -0.000075 |
|---|---|---|---|---|
| 24 | 0 | 0.0436 | 0.0874 | 0.1318 |
| 30 | 0 | 0.0711 | 0.1431 | 0.2164 |
| 36 | 0 | 0.0987 | 0.1993 | 0.3034 |

Table 6: Four-Period Median Market Price of Model Uncertainty (with habit persistence)

Panel A: The HST model

| $\mu_b \setminus \sigma$ | 0 | -0.000025 | -0.00005 | -0.000075 |
|---|---|---|---|---|
| 24 | 0 | 0.0348 | 0.0699 | 0.1051 |
| 30 | 0 | 0.0569 | 0.1142 | 0.1722 |
| 36 | 0 | 0.0789 | 0.1588 | 0.2405 |

Panel B: The comparison model

| $\mu_b \setminus \sigma$ | 0 | -0.000025 | -0.00005 | -0.000075 |
|---|---|---|---|---|
| 24 | 0 | 0.0350 | 0.0702 | 0.1056 |
| 30 | 0 | 0.0571 | 0.1147 | 0.1730 |
| 36 | 0 | 0.0793 | 0.1595 | 0.2416 |

Panel C: The HSW model

| $\mu_b \setminus \sigma$ | 0 | -0.000025 | -0.00005 | -0.000075 |
|---|---|---|---|---|
| 24 | 0 | 0.0551 | 0.1108 | 0.1673 |
| 30 | 0 | 0.0900 | 0.1816 | 0.2760 |
| 36 | 0 | 0.1250 | 0.2537 | 0.3894 |

Table 7: Detection Error Probability

Panel A: The HST model

| $\mu_b \diagdown \sigma$ | 0 | -0.000025 | -0.00005 | -0.000075 |
|---|---|---|---|---|
| 24 | 0.5000 | 0.4605 | 0.4254 | 0.3844 |
| 30 | 0.5000 | 0.4390 | 0.3739 | 0.3216 |
| 36 | 0.5000 | 0.4165 | 0.3371 | 0.2576 |

Panel B: The comparison model

| $\mu_b \diagdown \sigma$ | 0 | -0.000025 | -0.00005 | -0.000075 |
|---|---|---|---|---|
| 24 | 0.5000 | 0.4637 | 0.4237 | 0.3857 |
| 30 | 0.5000 | 0.4370 | 0.3796 | 0.3231 |
| 36 | 0.5000 | 0.4169 | 0.3380 | 0.2647 |

Panel C: The HSW model

| $\mu_b \diagdown \sigma$ | 0 | -0.000025 | -0.00005 | -0.000075 |
|---|---|---|---|---|
| 24 | 0.5000 | 0.4410 | 0.3781 | 0.3191 |
| 30 | 0.5000 | 0.4063 | 0.3092 | 0.2238 |
| 36 | 0.5000 | 0.3731 | 0.2466 | 0.1481 |

Table 8: Implied Coefficients of Relative Risk Aversion

| quantile$\diagdown \mu_b$ | 18 | 24 | 30 | 36 |
|---|---|---|---|---|
| .25 | 13.1 | 5.1 | 3.1 | 2.3 |
| .5 | 14.2 | 5.2 | 3.2 | 2.3 |
| .75 | 15.4 | 5.4 | 3.3 | 2.4 |

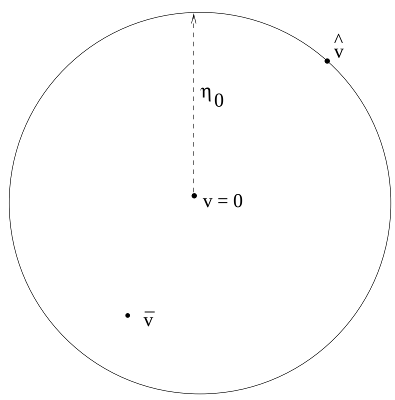


Figure 1: Three models: the approximating model $v = 0$, the true model $v = \bar{v}$, and the worst-case model $\hat{v}$.

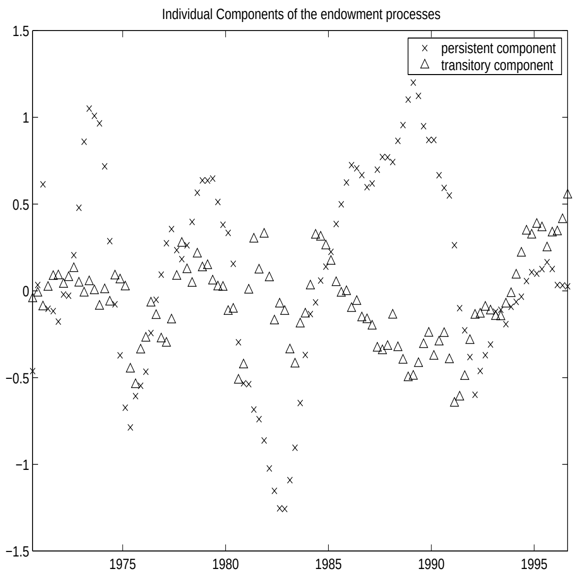


Figure 2: Actual permanent and transitory components of endowment process from Hansen, Sargent, Tallarini (1999) model.

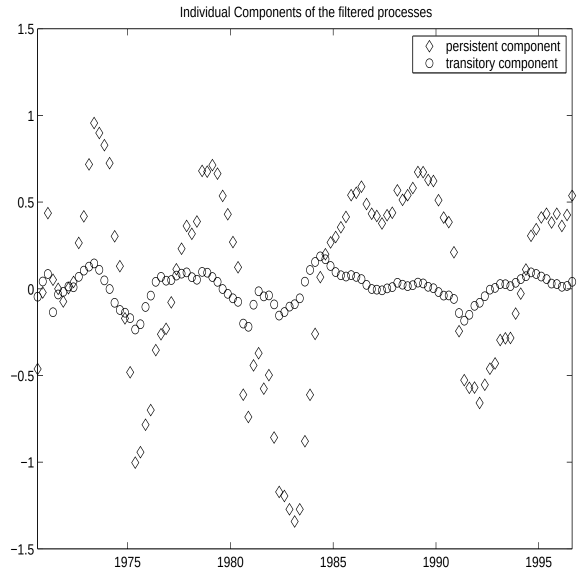


Figure 3: Filtered estimates of permanent and transitory components of endowment process from Hansen, Sargent, Tallarini (1999) model.

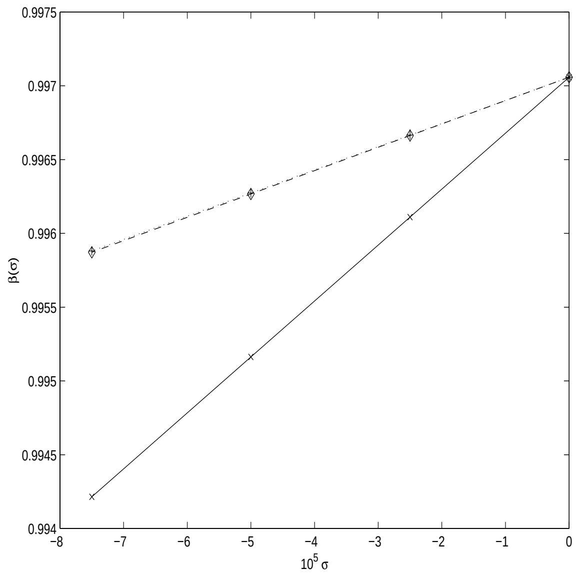


Figure 4: Observationally equivalent $(\beta, \sigma)$ pairs, $\beta$ on the vertical axis. The steeper line is for the HSW model, the overlapping less steep line is for the HST model and the comparison model.

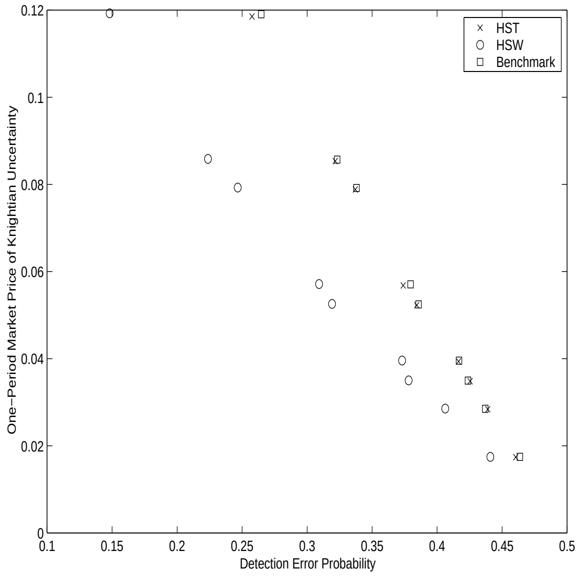


Figure 5: One-Period Market Price of Knightian Uncertainty versus Detection Error Probability

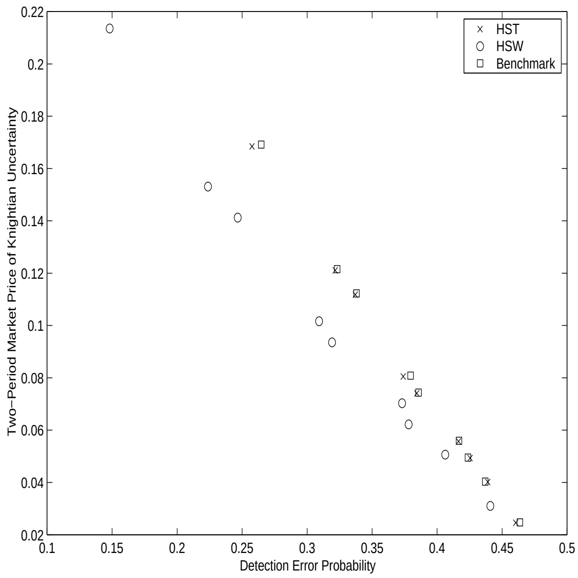


Figure 6: Two-Period Market Price of Knightian Uncertainty versus Detection Error Probability

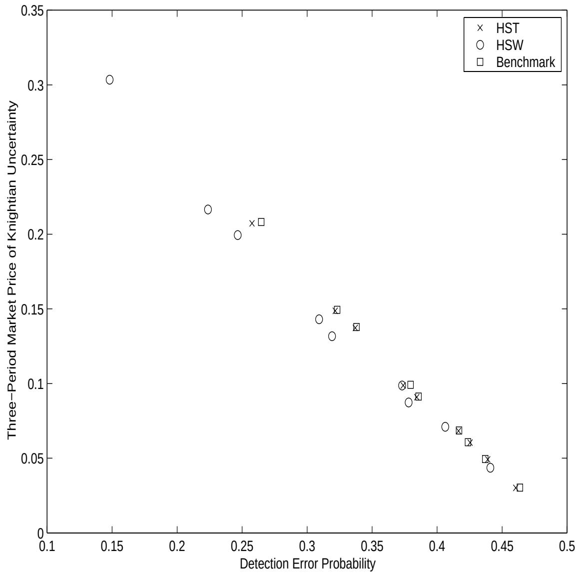


Figure 7: Three-Period Market Price of Knightian Uncertainty versus Detection Error Probability

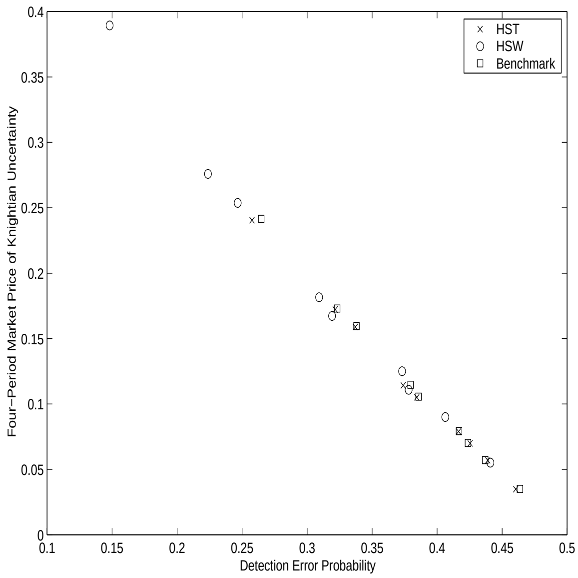


Figure 8: Four-Period Market Price of Knightian Uncertainty versus Detection Error Probability

# Appendices

# A Detection Error Probabilities

This appendix describes how we compute the detection error probabilities. First, we describe detection error probabilities for the basic HST model, and then for the HSW and comparison models.

## A.1 Likelihood Ratio under the Approximating Model

Represent the *approximating* model as

$$x_{t+1} = A_o x_t + C w_{t+1}, \tag{A.1}$$

where $w_{t+1}$ is a sequence of i.i.d. Gaussian vectors with mean zero and covariance matrix $I$. In this part, we assume that the true data generating process is this *approximating* model.

Represent the *distorted* model as

$$\begin{aligned} x_{t+1} &= A_o x_t + C(\check{w}_{t+1} + v_t), \\ &= \hat{A} x_t + C \check{w}_{t+1}. \end{aligned} \tag{A.2}$$

Define $v^A$ as the worst case shock assuming that the underlying data generating process is the approximating model, i.e., $v^A = \kappa x^A$ and $\hat{A} = A_o + C\kappa$, where $x^A$ is generated under (A.1). Hence, we can express the innovation under the worst case model as:

$$\check{w}_{t+1} = w_{t+1} - v_t^A. \tag{A.3}$$

The log likelihood function under the approximating model is

$$\log L_{AA} = -\frac{1}{T}\sum_{t=0}^{T-1}\{\log\sqrt{2\pi} + \frac{1}{2}(w_{t+1}\cdot w_{t+1})\}. \tag{A.4}$$

The likelihood function for the distorted model, given that (A.1) is the data generating process, is

$$\begin{aligned}\log L_{AB} &= -\frac{1}{T}\sum_{t=0}^{T-1}\{\log\sqrt{2\pi} + \frac{1}{2}(\check{w}_{t+1}\cdot \check{w}_{t+1})\},\\ &= -\frac{1}{T}\sum_{t=0}^{T-1}\{\log\sqrt{2\pi} + \frac{1}{2}(w_{t+1} - v_t^A)'(w_{t+1} - v_t^A)\}.\end{aligned} \tag{A.5}$$

Hence, the likelihood ratio assuming that the approximating model is the data generating process, $r_A$ is :

$$\begin{aligned}r_A &\equiv \log L_{AA} - \log L_{AB},\\ &= \frac{1}{2T}\sum_{t=0}^{T-1}\{\check{w}_{t+1}\cdot\check{w}_{t+1} - w_{t+1}\cdot w_{t+1}\},\\ &= \frac{1}{T}\sum_{t=0}^{T-1}\{\frac{1}{2}v_t^{A\prime}v_t^A - v_t^{A\prime}w_{t+1}\}.\end{aligned} \tag{A.6}$$

## A.2 Likelihood Ratio under the Distorted Model

Now suppose that the data generating process is the *distorted* model, described as follow

$$\begin{aligned}x_{t+1} &= (A_o + C\kappa)x_t + C\epsilon_{t+1},\\ &\equiv \hat{A}x_t + C\epsilon_{t+1},\end{aligned} \tag{A.7}$$

where $\hat{A} = A_o + C\kappa$. Under the approximating model, we have,

$$x_{t+1} = A_o x_t + C\check{\epsilon}_{t+1}. \tag{A.8}$$

Hence, $\check{\epsilon}_{t+1} = \epsilon_{t+1} + v_t^B$, where $v^B = \kappa x_t^B$ and $x_t^B$ is the time series generated under (A.7).

The loglikelihood function $\log L_{BB}$ for the *distorted* model, assuming that the *distorted* model generates the data is

$$\log L_{BB} = -\frac{1}{T}\sum_{t=0}^{T-1}\{\log\sqrt{2\pi} + \frac{1}{2}(\epsilon_{t+1}\cdot\epsilon_{t+1})\}. \tag{A.9}$$

The log likelihood function $\log L_{BA}$ for the approximating model, assuming that the distorted model (A.7) generates the data is,

$$\begin{aligned}\log L_{BA} &= -\frac{1}{T}\sum_{t=0}^{T-1}\{\log\sqrt{2\pi} + \frac{1}{2}(\check{\epsilon}_{t+1}\cdot\check{\epsilon}_{t+1})\},\\ &= -\frac{1}{T}\sum_{t=0}^{T-1}\{\log\sqrt{2\pi} + \frac{1}{2}(\epsilon_{t+1} + v_t^B)'(\epsilon_{t+1} + v_t^B)\}.\end{aligned} \tag{A.10}$$

Hence, the likelihood ratio $r_B$, assuming that the *distorted* model is the data generating process is

$$\begin{aligned}r_B &\equiv \log L_{BB} - \log L_{BA},\\ &= \frac{1}{2T}\sum_{t=0}^{T-1}\{\check{\epsilon}_{t+1}\cdot\check{\epsilon}_{t+1} - \epsilon_{t+1}\cdot\epsilon_{t+1}\},\\ &= \frac{1}{T}\sum_{t=0}^{T-1}\{\frac{1}{2}v_t^{B'}v_t^B + v_t^{B'}\epsilon_{t+1}\}.\end{aligned} \tag{A.11}$$

## A.3 The Detection Error Probability

The detection error probability is defined as,

$$p(\theta) = \frac{1}{2}(p_A + p_B), \tag{A.12}$$

where $p_i = \text{freq}(r_i \leq 0), i = A, B$. We attach equal prior weights to model $A$ and $B$. To compute $p(\theta)$, we simulate a large number of trajectories and calculate the empirical detection error probability.

## A.4 The HSW and the Comparison Models

For the HSW model, this appendix describes in detail how we simulated the approximating and worst case models and evaluated their likelihood functions to calculate the detection error probabilities.

### A.4.1 Simulating data under the worst case model

First, simulate under the worst case model, described by the following law of motion:

$$\begin{aligned} y^* &= A_y\check{x} + \check{C}_y(\check{w} + \check{v}), \\ \check{z}^* &= A_z\check{x} + \check{C}_z\check{C}_y^{-1}(y^* - A_y\check{x}) + \tilde{C}_z\tilde{v}, \\ f^* &= A_f\check{x} + B_f u, \\ &= (A_f - B_f F)\check{x}, \end{aligned} \tag{A.13}$$

given the initial condition $\check{x}_0$ from HST (after appropriate transformation to the newly defined state vector notation in order to make $C_yC_y'$ nonsingular.) Note that

$$v = \begin{bmatrix} \check{v} \\ \tilde{v} \end{bmatrix} = \begin{bmatrix} \check{\kappa} \\ \tilde{\kappa} \end{bmatrix} \check{x} \equiv \kappa\check{x}, \tag{A.14}$$

where $\check{x}$ is generated under (A.13).

First, given initial $\check{x}$ value from HST, calculate $\check{v} = \check{\kappa}\check{x}$, draw a $\check{w}$ from $\mathcal{N}(0,1)$, calculate $y^*$. Second, compute next period's $\check{z}^*$ using $y^*$ and $\tilde{v} = \tilde{\kappa}\check{x}$. Third, calculate next period endogenous $f^*$ using the third equation in (A.13). Finally, construct $\check{x}^* = \begin{bmatrix} f^* & y^* & \check{z}^* \end{bmatrix}^T$ for the next period and repeat this procedure.

### A.4.2 Simulating data under the approximating model

Perform the same procedure under the approximating model, except that now simulation is done under the following law of motion:

$$
\begin{aligned}
y^* &= A_y \check{x} + \check{C}_y \check{w}, \\
\check{z}^* &= A_z \check{x} + \check{C}_z \check{C}_y^{-1}(y^* - A_y \check{x}), \\
f^* &= A_f \check{x} + B_f u, \\
&= (A_f - B_f F)\check{x}.
\end{aligned}
\tag{A.15}
$$

Note that there is no $\check{v}$ or $\tilde{v}$ appearing in the simulation.

## A.5 Simulation under the comparison model

### A.5.1 Simulating data under the worst case model

In the spirit of Section A.4, from the initial condition on $\check{x}$, we simulate using

$$
\begin{aligned}
y^* &= A_y \check{x} + \check{C}_y (\check{w} + \check{v}), \\
\check{z}^* &= A_z \check{x} + \check{C}_z \check{C}_y^{-1}(y^* - A_y \check{x}), \\
f^* &= A_f \check{x} + B_f u, \\
&= (A_f - B_f F)\check{x}.
\end{aligned}
\tag{A.16}
$$

Given the initial condition $\check{x}_0$, we iterate out the simulated data series for $\{y\}_{t=1}^T$.

### A.5.2 Simulating data under the approximating model

Perform the following simulation

$$\begin{aligned} y^* &= A_y \check{x} + \check{C}_y \check{w}, \\ \check{z}^* &= A_z \check{x} + \check{C}_z \check{C}_y^{-1}(y^* - A_y \check{x}), \\ f^* &= A_f \check{x} + B_f u, \\ &= (A_f - B_f F)\check{x}. \end{aligned} \tag{A.17}$$

Note that these equations for simulation under the approximating model for the comparion model are the same as those for simulation under the approximating model for the HSW model (A.15).

## A.6 Likelihood ratio for the HSW model

Given one realization of simulated data $\{y_t\}_{t=1}^T$, (whether (A.15) or (A.13) generates the data,) we can compute the likelihood under the worst case and approximating models as follows.

### A.6.1 Likelihood under the worst case model

The likelihood under the worst case model is

$$\sum_{t=1}^{T} \left[ -\frac{1}{2}(y_{t+1} - A_y \check{x}_t - \check{C}_y \check{v}_t)'(\check{C}_y \check{C}_y')^{-1}(y_{t+1} - A_y \check{x}_t - \check{C}_y \check{v}_t) \right], \tag{A.18}$$

where $\check{x}_t$ is filtered using the Kalman filter under worst case model:

$$\begin{aligned} \check{z}_{t+1} &= A_z \check{x}_t + \check{C}_z \check{C}_y^{-1}(y_{t+1} - A_y \check{x}_t) + \tilde{C}_z \check{v}_t, \\ f_{t+1} &= (A_f - B_f F)\check{x}_t. \end{aligned} \tag{A.19}$$

Again note that $\tilde{v}_t = \tilde{\kappa}\check{x}_t$ and

$$\check{x}_{t+1} = \begin{bmatrix} f_{t+1} \\ y_{t+1} \\ \check{z}_{t+1} \end{bmatrix}. \tag{A.20}$$

Equation (A.19) generates the filtered state. Then we can compute $\check{v}$ and hence construct the log likelihood defined in (A.18).

### A.6.2 Likelihood under the approximating model

The likelihood under the approximating model is

$$\sum_{t=1}^{T} \left[ -\frac{1}{2}(y_{t+1} - A_y\check{x}_t)'(\check{C}_y\check{C}_y')^{-1}(y_{t+1} - A_y\check{x}_t) \right], \tag{A.21}$$

where $\check{x}_t$ is filtered using the following Kalman filter under worst case model:

$$\begin{aligned} \check{z}_{t+1} &= A_z\check{x}_t + \check{C}_z\check{C}_y'^{-1}(y_{t+1} - A_y\check{x}_t), \\ f_{t+1} &= (A_f - B_fF)\check{x}_t. \end{aligned} \tag{A.22}$$

With input $\{y_t\}_{t=1}^T$ and initial condition $\check{x}_0$, we construct the filtered state for the comparison model assuming that the approximating model generates the data based on (A.27).

## A.7 Likelihood ratio for the comparison model

Given one draw from, say simulated data $\{y_t\}_{t=1}^T$, whether (A.17) or (A.16) generates the data, we can compute the likelihood under the worst case and approximating models.

### A.7.1 Likelihood under worst case model

First compute under worst case model:

$$\sum_{t=1}^{T} \left[ -\frac{1}{2}(y_{t+1} - A_y \check{x}_t - \check{C}_y \check{v}_t)' (\check{C}_y \check{C}_y')^{-1} (y_{t+1} - A_y \check{x}_t - \check{C}_y \check{v}_t) \right], \tag{A.23}$$

where $\check{x}_t$ is filtered using the Kalman filter under worst case model:

$$\begin{aligned} \check{z}_{t+1} &= A_z \check{x}_t + \check{C}_z \check{C}_y^{-1} (y_{t+1} - A_y \check{x}_t), \\ f_{t+1} &= (A_f - B_f F) \check{x}_t. \end{aligned} \tag{A.24}$$

Again note that $\tilde{v}_t = \tilde{\kappa} \check{x}_t$ and

$$\check{x}_{t+1} = \begin{bmatrix} f_{t+1} \\ y_{t+1} \\ \check{z}_{t+1} \end{bmatrix}. \tag{A.25}$$

Equation A.24 generates the filtered state. Then we may compute $\check{v}$ and hence, construct the loglikelihood defined in (A.23).

### A.7.2 Likelihood under approximating model

The likelihood under the approximating model is

$$\sum_{t=1}^{T} \left[ -\frac{1}{2}(y_{t+1} - A_y \check{x}_t)' (\check{C}_y \check{C}_y')^{-1} (y_{t+1} - A_y \check{x}_t) \right], \tag{A.26}$$

where $\check{x}_t$ is filtered using the Kalman filter under worst case model:

$$\check{z}_{t+1} = A_z \check{x}_t + \check{C}_z \check{C}_y^{-1}(y_{t+1} - A_y \check{x}_t),$$
$$f_{t+1} = (A_f - B_f F)\check{x}_t. \tag{A.27}$$

With input $\{y_t\}_{t=1}^T$ and initial condition $\check{x}_0$, we construct the filtered state for the comparison model assuming that the approximating model generates the data based on (A.27).

# B Computing Robust Decision Rules

Consider a general optimization problem in a discounted linear quadratic environment when the agent is concerned about model misspecification. Let $x_t$ be an $(n \times 1)$ state vector, $\bar{u}_t$ be an $(k \times 1)$ control variable, and $w_t$ be an $(m \times 1)$ Gaussian noise hitting the system at time $t$. The state vector is assumed to follow,

$$x_{t+1} = \bar{A}x_t + B\bar{u}_t + Cw_{t+1} \tag{B.1}$$

where $\bar{A}$ is an $(n \times n)$ , $B$ is an $(n \times k)$ and $C$ is an $(n \times m)$ matrix, respectively. We define the time homogeneous instantaneous return function, $r(x, \bar{u})$ to have the quadratic form:

$$r(x, \bar{u}) = -(x' \ \bar{u}') \begin{bmatrix} \bar{R} & W \\ W' & Q \end{bmatrix} \begin{pmatrix} x \\ \bar{u} \end{pmatrix}, \tag{B.2}$$

where $\bar{R}$ is an $(n \times n)$, $Q$ is a $(k \times k)$ and $W$ is an $(n \times k)$ matrix, respectively. Her concern about the model uncertainty is summarized by the parameter $\theta$. She solves the following minmax optimization problem:

$$\begin{aligned}\tilde{V}(x) &= \sup_{\bar{u}} \inf_{v} r(x,\bar{u}) + \beta \left[\theta v'v + E\tilde{V}(\bar{A}x + B\bar{u} + C(w+v))\right], \\ &= -x'\Omega x - a. \end{aligned} \tag{B.3}$$

To eliminate the cross product between the state vector and the control variable, we define

$$\begin{aligned} R &= \bar{R} - WQ^{-1}W', \\ A &= \bar{A} - BQ^{-1}W', \\ u &= \bar{u} + Q^{-1}W'x. \end{aligned} \tag{B.4}$$

The above transformation converts the law of motion (B.1) to the following equivalent representation:

$$x_{t+1} = Ax_t + Bu_t + Cw_{t+1}. \tag{B.5}$$

The agent's optimal decision rule and the worst case shock take the form of

$$\begin{aligned} u &= -\mathcal{F} \circ \mathcal{D}(\Omega)x, \\ \hat{v} &= \theta^{-1}(I - \theta^{-1}C'\Omega C)^{-1}C'\Omega \left[A - B\mathcal{F} \circ \mathcal{D}(\Omega)\right] x \equiv \kappa x, \end{aligned} \tag{B.6}$$

where

$$\begin{aligned} \mathcal{F}(\Omega) &= \beta \left[Q + \beta B'\Omega B\right]^{-1} B'\Omega A, \\ \mathcal{D}(\Omega) &= \Omega + \theta^{-1}\Omega C(I - \theta^{-1}C'\Omega C)^{-1}C'\Omega, \\ \kappa &= \theta^{-1}(I - \theta^{-1}C'\Omega C)^{-1}C'\Omega \left[A - B\mathcal{F} \circ \mathcal{D}(\Omega)\right]. \end{aligned} \tag{B.7}$$

$\mathcal{D}$ captures the notion of robustness through its second term and $\mathcal{F}$ is the standard decision rule for discounted linear quadratic regular problem. To compute the solution of the optimizers from (B.6) and (B.7), we first need to compute the value function $\Omega$.

This can be achieved by solving the following fixed point problem

$$\Omega = \mathcal{T} \circ \mathcal{D}(\Omega), \tag{B.8}$$

where

$$\begin{aligned} \mathcal{T}(P) &= R + \mathcal{F}(P)'Q\mathcal{F}(P) + \beta[A - B\mathcal{F}(P)]'P[A - B\mathcal{F}(P)], \\ &= R + \beta A'(P - \beta PB(Q + \beta B'PB)^{-1}B'P)A. \end{aligned} \tag{B.9}$$

# C Multi-period market prices of Knightian uncertainty

This appendix describes the detailed computations of multi-period market prices of Knightian uncertainty with perfectly observable state vector, as in HST. The one-period market prices of Knightian uncertainty in HST's calculation is subsumed here. The notation in this appendix is self-contained. Some notation from the text has been recycled.[26]

The law of motion under the approximating model is

$$x_{t+1} = A^* x_t + C\check{w}_{t+1},$$

and under the worst-case model is

$$x_{t+1} = \check{A} x_t + C\check{w}_{t+1}.$$

There is perfect observability of the state vector $x$, as assumed under HST.

Define

$$x^j_{t+j} = \begin{bmatrix} x_{t+1} \\ x_{t+2} \\ \vdots \\ x_{t+j} \end{bmatrix}, \quad \text{and} \quad \check{w}^j_{t+j} = \begin{bmatrix} \check{w}_{t+1} \\ \check{w}_{t+2} \\ \vdots \\ \check{w}_{t+j} \end{bmatrix}. \tag{C.1}$$

Note that the dimension of $x^j_{t+j}$ is $(nj) \times 1$, where $n$ is the dimension of state vector $x$, for our analysis $n = 7$. Under the approximating model, $x^j_{t+j}$ follows by induction,

$$x^j_{t+j} = M^*_j x_t + N^*_j \check{w}^j_{t+j}, \tag{C.2}$$

where

$$M^*_j = \begin{bmatrix} A^* \\ (A^*)^2 \\ \vdots \\ (A^*)^j \end{bmatrix} = \begin{bmatrix} M^*_{j-1} \\ (A^*)^j \end{bmatrix}, \text{ and } N^*_j = \begin{bmatrix} C & 0 & \ldots & 0 \\ A^*C & C & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ (A^*)^{j-1}C & \ldots & A^*C & C \end{bmatrix} = \begin{bmatrix} N^*_{j-1} & 0 \\ S^*_{j-1} & N^*_1 \end{bmatrix}, \tag{C.3}$$

and where $S^*_j$ is define recursively as $S^*_j = \begin{bmatrix} (A^*)^j C & S^*_{j-1} \end{bmatrix}$, with $S^*_1 = A^*C$.

Under the worst-case model,

$$x^j_{t+j} = \check{M}_j x_t + \check{N}_j \check{w}^j_{t+j}, \tag{C.4}$$

where

$$\check{M}_j = \begin{bmatrix} \check{A} \\ \check{A}^2 \\ \vdots \\ \check{A}^j \end{bmatrix} = \begin{bmatrix} \check{M}_{j-1} \\ \check{A}^j \end{bmatrix}, \text{ and } \check{N}_j = \begin{bmatrix} C & 0 & \dots & 0 \\ \check{A}C & C & \dots & 0 \\ \dots\dots & \dots & \dots & \dots \\ (\check{A})^{j-1}C & \dots & \check{A}C & C \end{bmatrix} = \begin{bmatrix} \check{N}_{j-1} & 0 \\ \check{S}_{j-1} & \check{N}_1 \end{bmatrix}, \tag{C.5}$$

and where $\check{S}_j$ is defined recursively as $\check{S}_j = \begin{bmatrix} (\check{A})^j C & \check{S}_{j-1} \end{bmatrix}$, with $\check{S}_1 = \check{A}C$.

The initial conditions for these matrices are

$$N_1^* = \check{N}_1 = C,\ M_1^* = A^*,\ \text{and } \check{M}_1 = \check{A}.$$

Recall that under HST $\check{A} = A^* + C\kappa$, where $\kappa$ is the worst case shock coefficient.

Under the approximating model,

$$\check{w}_{t+j}^j = \left[(N_j^*)' N_j^*\right]^{-1} (N_j^*)'(x_{t+j}^j - M_j^* x_t). \tag{C.6}$$

Under the worst case model,

$$\begin{aligned} \left[(\check{N}_j)'\check{N}_j\right]^{-1} (\check{N}_j)'(x_{t+j}^j - \check{M}_j x_t) &= \left[(\check{N}_j)'\check{N}_j\right]^{-1} (\check{N}_j)'(x_{t+j}^j - M_j^* x_t - (\check{M}_j - M_j^*) x_t), \\ &= L_j \check{w}_{t+j}^j - O_j x_t, \end{aligned} \tag{C.7}$$

where

$$\begin{aligned} L_j &= \left[(\check{N}_j)'\check{N}_j\right]^{-1} (\check{N}_j)' N_j^*, \\ O_j &= \left[(\check{N}_j)'\check{N}_j\right]^{-1} (\check{N}_j)' \left[\check{M}_j - M_j^*\right]. \end{aligned} \tag{C.8}$$

Hence, the likelihood ratio

$$m^u_{t+j,t} = \frac{\exp\left[-\frac{1}{2}(L_j\breve{w}^j_{t+j} - O_j x_t)\cdot(L_j\breve{w}^j_{t+j} - O_j x_t)\right]}{\exp\left[-\frac{1}{2}\breve{w}^j_{t+j}\cdot\breve{w}^j_{t+j}\right]}. \tag{C.9}$$

We assume that $2L'_jL_j - I$ is positive definite and let $P_j$ be its Cholesky decomposition factor:

$$\begin{aligned} P'_jP_j &= 2L'_jL_j - I, \\ Q_j &= 2(P'_j)^{-1}L'_jO_j. \end{aligned} \tag{C.10}$$

The second moment of the market price of Knightian uncertainty hence can be expressed as follows:

$$\mathbb{E}_t\left[m^u_{t+j,t}\right]^2 = \left[\frac{1}{\sqrt{2\pi}}\right]^j \int_{-\infty}^{\infty} \Omega(\breve{w}^j_{t+j})d\breve{w}^j_{t+j}, \tag{C.11}$$

where

$$\Omega(\breve{w}^j_{t+j}) = \exp\left[-\frac{1}{2}(\breve{w}^j_{t+j})'P'_jP_j\breve{w}^j_{t+j} + x'_t(2O'_jL_jP_j^{-1})P_j\breve{w}^j_{t+j} - (O_jx_t)\cdot(O_jx_t)\right].$$

Hence, the conditional second moment of the market price of Knightian uncertainty is

$$\mathbb{E}_t\left[m^u_{t+j,t}\right]^2 = (\det P_j)^{-1}\exp\left[x'_tR_jx_t\right], \tag{C.12}$$

where

$$\begin{aligned} R_j &= \frac{1}{2}Q'_jQ_j - O'_jO_j, \\ &= 2O'_jL_jP_j^{-1}(P'_j)^{-1}L'_jO_j - O'_jO_j, \\ &= O'_j\left[2L_jL'_j - I\right]^{-1}O_j. \end{aligned} \tag{C.13}$$

Note that by construction, the conditional expectation of the market price of Knightian uncertainty is 1, namely, $\mathbb{E}_t m^u_{t+j,t} = 1$. Finally, the market price of Knightian uncertainty

$$\frac{\sigma_t(m^u_{t+j,t})}{\mathbb{E}_t(m^u_{t+j,t})} = \sqrt{(\det P_j)^{-1}\exp\left[x_t' R_j x_t\right] - 1}. \tag{C.14}$$

It seems we need to show that

$$\det(P_j) \leq 1$$

and $R_j$ is positive semidefinite.

HST's calculation is our special case with $j = 1$. Obviously, $L_1 = I$, $O_1 = \kappa$, $P_1 = I$, and $R_1 = \kappa'\kappa$. Hence,

$$\begin{aligned} \frac{\sigma_t(m^u_{t+1,t})}{\mathbb{E}_t(m^u_{t+1,t})} &= \sqrt{(\det P_1)^{-1}\exp\left[x_t' R_1 x_t\right] - 1}, \\ &= \sqrt{\exp\left[v_t \cdot v_t\right] - 1}, \end{aligned} \tag{C.15}$$

where $v_t = \kappa x_t$.

# Notes

[1]The combined estimation and control calculations extend Hansen and Sargent's (1995) formulation of a discounted risk-sensitive problem.

[2]See Cochrane and Hansen (1992), Constantinides and Duffie (1996), Mehra and Prescott (1985), Weil (1989) and Hansen and Jagannathan (1991).

[3]See Hansen and Sargent (2000) and Cagetti, Hansen, Sargent, and Williams (2000).

[4]See Harrison and Kreps (1979).

[5]As in HST, another way to interpret our calculations is as perturbing the measure with respect to which the expectation is evaluated, while retaining the ordinary formula for the stochastic discount factor.

[6]While Hansen and Jagannathan (1991) looked at the unconditional counterpart to this pricing inequality for multiple assets, Gallant and Tauchen (1990) studied the conditional version (2).

[7]See Hansen and Singleton (1983), Mehra and Prescott (1985), or Hansen and Jagannathan (1991) for alternative statements of this phenomenon.

[8]See Anderson, Hansen, and Sargent (2000) for an alternative specification of a class of models. Their approximating model is a controlled Markov process. They form a set of alternative models by multiplying the one-step transition density of the approximating model by a strictly positive function. It can be shown that the formulation

for the linear stochastic difference equation in this paper is consistent with Anderson, Hansen, and Sargent's.

[9]Thus, $\theta$ is $+\infty$ for $\eta_0 = 0$, and falls as $\eta_0$ rises above zero.

[10]This is called the reference model in much of the control theory literature.

[11]See Appendix B for a formula for $\Omega$.

[12]This $\mathcal{R}$ operator also appears in literature on recursive utility. See Kreps and Porteus (1978), Epstein and Zin (1989) and Duffie and Epstein (1992).

[13]For studies of preferences with habit formation, see Ryder and Heal (1973), Becker and Murphy (1988), Sudaresan (1989), Constantinides (1990) and Heaton (1993).

[14]The two parameterizations each for $d^1$ and $d^2$ are equivalent, the first being used in this paper and the second in HST.

[15]For HST, the two-component structure served also the purpose of assuring 'stochastic nonsingularity', meaning a spectral density of full rank for the observables $c_t, i_t$ for which they constructed a likelihood function for estimating free parameters.

[16]Formula (33) solves and simplifies an implicit function in HST.

[17]See Hansen and Jagannathan (1991) for this characterization of the equity premium puzzle.

[18]Here $C_z C_y'$ measures the covariance between the state and measurement errors.

[19]The spectral factorization achieved by (41) assures the equality $Cw + A(x - \check{x}) = C^{*}w^{*}$.

[20]See Anderson and Moore (1979) for a discussion of innovations representations, also called Wold representations.

[21] This differs from the procedure recommended by Basar and Bernhard (1995) and Whittle (1990). The difference stems from their using a different criterion, according to which the decision maker cares equally about past and future returns.

[22]These combinations include ones originally reported in HST and some additional ones besides.

[23]Given the time separabilities in preferences, there are important distinctions between consumption and wealth lotteries. See Constantinides (1990) for a discussion of this point and suggestions for other measures of risk aversion.

[24]The market prices of uncertainty are computed using the exact formula (34) while HST used an approximation.

[25]We chose a smaller range of $\sigma$'s because some of the $\sigma$'s in the tables are beyond the 'breakdown point' for the HSW model. See HST and Whittle (1990) for explanations of the breakdown point.

[26]We use $x_t$ for what was $\check{x}_t$ in the text.

# References


Anderson, E., L. P. Hansen, and T. J. Sargent (2000). Robustness, detection and the price of risk. manuscript, University of North Carolina, University of Chicago and Stanford University.

Basar, T. and P. Bernhard (1995). *$H_\infty$ Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*. Boston-Basel-Berlin: Birkhauser.

Becker, G. S. and K. M. Murphy (1988). A theory of rational addiction. *Journal of Political Economy 96*, 675–700.

Breeden, D. (1979). An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics 7*, 265–296.

Cagetti, M., L. P. Hansen, T. J. Sargent, and N. Williams (2000). manuscript, University of Chicago and Stanford University.

Cochrane, J. H. and L. P. Hansen (1992). Asset pricing explorations for macroeconomics. *Economic Perspectives*, 115–65. *NBER Macroeconomics Annual*, Cambridge, MA: The MIT Press.

Constantinides, G. M. (1990). Habit formation: A resolution of the equity premium puzzle. *Journal of Political Economy 98*, 519–43.

Constantinides, G. M. and D. Duffie (1996). Asset pricing with heterogeneous consumers. *Journal of Political Economy 104*, 219–240.

Detemple, J. (1986). Asset pricing in a production economy with incomplete information. *Journal of Finance 41*, 383–391.

Dothan, M. and D. Feldman (1986). Equilibrium interest rates and multiperiod bonds in a partially observed economy. *Journal of Finance 41*, 369–382.

Duffie, D. and L. G. Epstein (1992). Stochastic differential utility. *Econometrica 60*, 353–394.

Epstein, L. G. and S. E. Zin (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica 57*, 937–969.

Gallant, R., L. P. H. and G. Tauchen (1990). Using conditional moments of asset payoffs to infer the volatility of intertemporal marginal rates of substitution. *Journal of Econometrics 45*, 145–179.

Gennotte, G. (1986). Optimal portfolio choice under incomplete information. *Journal of Finance 41*, 733–749.

Hansen, L. P. and R. Jagannathan (1991). Implications of security market data for models of dynamic economies. *Journal of Political Economy 99*, 225–262.

Hansen, L. P. and T. J. Sargent (2000). Elements of robust control and filtering for macroeconomics. manuscript, University of Chicago and Stanford University.

Hansen, L. P., T. J. Sargent, and T. Tallarini (1999). Robust permanent income and pricing. *Review of Economic Studies 66*, 873–907.

Hansen, L. P. and K. J. Singleton (1983). Stochastic consumption, risk aversion, and the temporal behavior of asset returns. *Journal of Political Economy 91*, 249–265.

Harrison, J. and D. M. Kreps (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory 20*, 381–408.

Heaton, J. (1993). The interaction between time-nonseparable preferences and time aggregation. *Econometrica 61*, 353–385.

Jacobson, D. H. (1973). *Optimal Stochastic Linear Systems with Exponential Performance Criteria and Their Relation to Deterministic Differential Games*, Volume AC-18.

Kreps, D. and E. L. Porteus (1978). Temporal resolution of uncertainty and dynamic choice theory. *Econometrica 46*, 185–200.

LeRoy, S. (1973). Risk aversion and the martingale property of asset prices. *International Economic Review 14*, 436–446.

Lucas Jr., R. E. (1978). Asset prices in an exchange economy. *Econometrica 46*, 1429–1445.

Mehra, R. and E. C. Prescott (1985). The equity premium: A puzzle. *Journal of Monetary Economics 22*, 133–136.

Ryder, H.E., J. and G. Heal (1973). Optimal growth with intertemporally dependent preferences. *Review of Economic Studies 40*, 1–31.

Sudaresan, S. (1989). Intertemporally dependent preferences and the volatility of consumption and wealth. *Review of Financial Studies 2*, 73–89.

Veronesi, P. (1999). Stock market overreaction to bad news in good times: A rational expectations equilibrium model. *RFS 12*, 975–1007.

Weil, P. (1989). The equity premium puzzle and the riskfree rate puzzle. *JME 24*, 401–421.

Whittle, P. (1990). *Risk-sensitive Optimal Control*. New York: John Wiley & Sons.