Belief-dependent Utilities, Aversion to State-Uncertainty and Asset Prices*

Pietro Veronesi

University of Chicago, NBER and CEPR

First Version: August 2000

This Version: November 5, 2001

*I thank Federico Bandi, John Y. Campbell, Micaela Della Torre, Paolo Ghirardato, Lars P. Hansen, Fulvio Ortu, Nick Polson, Tano Santos, Costis Skiadas, Per Stromberg, Fernando Zapatero and seminar participants at Harvard University, University of Chicago, University of Southern California, University of Minnesota, the 2001 NBER Asset Pricing meeting in Los Angeles and the 2001 SAET conference in Ischia for comments and suggestions. My most sincere thanks to Marciano Siniscalchi for the lengthy conversations on many of the issues discussed in this paper. Of course, errors are my own. This work is supported by the James S. Kemper Faculty Research Fund at the Graduate School of Business, the University of Chicago.

Belief Dependent Utilities, Aversion to State-Uncertainty and Asset Prices

Pietro Veronesi

University of Chicago - GSB

1101 East 58th Street Chicago IL 60637

Abstract

This paper reinterprets standard axioms in choice theory to introduce the concepts of *"belief dependent" utility functions* and *aversion to "state-uncertainty,"* and it shows that this type of preferences helps to explain the various stylized facts of asset returns, including a high equity risk premium, a low risk-free rate, a high return volatility, stock return predictability and volatility clustering. In one particular specification consistent with habit formation preferences, I also argue that "aversion to state uncertainty" gives rise to "*aversion to long run risk*," that is, to the uncertainty surrounding the long run average of future consumption. In order to solve for asset prices and returns under general conditions about the hidden state variable, the paper also develops a discretization methodology to obtain approximate analytical solutions. In a parsimonious parametrization, I then show that the model calibrated to real consumption generates unconditional moments for asset returns that closely match the empirical ones. Finally, due to the estimated time-variation in the *dispersion* of the conditional distribution on the drift rate of consumption, the model also generates a time series of *conditional* return volatility in line with the ex-post integrated volatility of stock returns.

JEL Classification Number: D81, G12

Keywords: State-dependent Preferences, Uncertainty Aversion, Asset Pricing

1 Introduction

This paper introduces the concepts of *"belief dependent" utility functions* and *aversion to "state-uncertainty"* and shows in a standard pure-exchange economy that this type of preferences helps to explain the various stylized facts of stock returns, including a high equity risk premium, a low risk-free rate, a high return volatility, stock return predictability and volatility clustering.

In a nutshell, a "belief-dependent" utility function is a generalization of the more common "state-dependent" utility function, where the "state" is not fully observable. For example, consider the case of utility functions that depend on the agent's health status (see e.g. Zeckhauser (1970, 1973), Arrow (1974), Viscusi and Evans (1990)). For a whole range of diseases or risk factors, such as diabetes, high cholesterol etc., it is more reasonable to assume that agents may not be fully aware of their health status, because the diagnostic techniques are imprecise and check-ups infrequent (see e.g. Cutler and Richardson (1997, p. 253)). In this case, agents possess a probability distribution on their own health status and preferences become "belief-dependent:" Changes to this subjective probability distribution (may) lead to changes in the utility and (marginal utility) of income or consumption.

Similarly, the recent literature in macroeconomics and finance has focused on state-dependent utilities to explain the behavior of individual consumers/investors and of financial variables. Examples include the works on habit formation (e.g. Constantinides (1990), Abel (1990), Campbell and Cochrane (1999)), relative social standing (Bakshi and Chen (1996)), stochastic subsistence consumption levels (e.g. Campbell and Viceira (2001)) and loss aversion (e.g. Barberis, Huang and Santos (2000)). In this literature the "state" is always assumed perfectly observable although in many cases it is more realistic to assume that it is only partially observable: For example, when preferences depend on the relative performance with respect to a reference class of agents (as in *external* habit formation models), the lack of knowledge of other agents' income/consumption levels naturally leads to a belief-dependent representation of agents' preferences. Clearly, state-dependent utilities are then recovered as a special case in which agents have a degenerate probability distribution.

In this article, I first use standard axioms in the decision theory literature to show that "belief-dependent" utility functions can be obtained by re-interpreting the representation results about state-dependent utility function. I argue that this notion of belief-dependent utility functions naturally induces the concept of "*aversion to state-uncertainty*," that is, the aversion to a more diffuse distribution on the unknown state of nature. Intuitively, consider again the case of health-dependent utility and suppose that an agent assigns equal probabilities to the three scenarios that a non life-threatening illness is "serious," "mild" or "does not exist." If upon a doctor visit the agent learns that the illness is only "mild" and he is happy because he can now enjoy current consumption more, then we may think of him as being averse to state-uncertainty.[1] In other words, state-dependency may affect the utility in a non-linear fashion so that a change in the *dispersion* of the probability distribution on the "state" would bring about a change in the agent's utility. If such changes in the dispersion of the probability distribution do not affect the agent's utility, then he/she is neutral to state-uncertainty. I also characterize belief-dependent utility functions with constant coefficient of absolute or relative risk aversion or with belief-dependent coefficient of risk aversion.

I then apply this type of preferences to a pure exchange economy with incomplete information, where for generality I first leave unspecified the nature and the process of the unobservable "state" affecting investors' preferences. Using a new discretization approach, I obtain analytical expressions for prices and returns for both bonds and stocks and find conditions under which aversion to state uncertainty yields higher expected returns and volatility, and lower interest rates. Indeed, these effects take place for example when the "state" is a "procyclical" variable but inversely related to the marginal utility of consumption, a situation that occurs typically in external habit formation models (see e.g. Campbell and Cochrane (1999)). The results of this paper then provide additional insights on standard habit formation models, with the additional interpretation in terms of aversion to state uncertainty.

The empirical section pushes the latter interpretation further, where I assume for simplicity that the "state" and the dividend drift rate are perfectly correlated. In addition, I assume that the economy is hit by *structural breaks*, a situation that generates time-variation in the dispersion of the agents posterior density on the underlying drift rate of dividends. The

[1] Clearly, we should assume that the three health states have "equal distance" under some metric.

latter assumption is of particular interest because it leads to an additional interpretation of the concept of "aversion to state uncertainty," namely, "*aversion to long run risk*". In fact, since the drift rate of consumption determines the average path of *future* consumption, "aversion to state uncertainty" can be interpreted as "aversion to the dispersion of the long-run average consumption paths," which differs from the standard notion of (local) risk aversion in its temporal dimension. One interesting implication is that it is possible to separate the intertemporal elasticity of substitution from this "aversion to long-run risk" while retaining the time-separability of the utility function. This in turn makes the model highly tractable.

In this set up, I show that when calibrated to consumption data the model is able to explain many of the empirical features in the asset pricing literature. Specifically, *aversion to state uncertainty* (or long-run risk) has the effect of increasing the risk premium, lowering the risk free rate and increasing return volatility. In addition, accordingly with previous literature on learning (see e.g. Timmerman (1993), David (1997), Veronesi (1999, 2000)), I also find time-varying expected returns and stochastic volatility. Intuitively, aversion to state-uncertainty (long run risk) generates a high equity premium and a high return volatility because it increases the sensitivity of the marginal utility of consumption to news. In addition, it also lowers the interest rate because it increases the demand for bonds from investors who are concerned about the long-run mean of their consumption. Indeed, under the interpretation put forward in the paper, it is not only the "volatility" of consumption that matters in generating a risk premium, but the uncertainty on its long run drift as well. Hence, the consumption process can even be rather smooth (as in the data) and yet carry a somewhat high risk due to the uncertainty on its long-term average dispersion (which depends on the drift). Finally, since I am not constrained in assuming a high coefficient of (local) *risk* aversion to generate a high equity risk premium, interest rates result low also because I can calibrate the model to reasonable values for the elasticity of intertemporal substitution.

I also estimate the structural break model using quarterly consumption data for the post-war period and I find that the time series of fitted posterior distributions on the drift rate of consumption tend to display a relatively large and time-varying dispersion. This empirical finding further strengthens the notion that investors may be averse to this "long run risk" besides the local risk aversion stemming from the (low) volatility of consumption. Besides matching the unconditional first and second moments of returns and the level and the volatility of the interest rate, I also obtain a time series of fitted return volatility that matches well the realized volatility of stock returns in the post-war period (unfortunately, quarterly data for consumption are not available before 1946). In fact, it turns out that burst of high return volatility coincide with periods where the posterior density on the drift rate of consumption is more dispersed, which in turn increases the volatility of stock returns. I also show that the model implied price-dividend ratio is reasonable, although it matches the realized one to a less degree.

As a methodological contribution, this article also proposes a general methodology to obtain pricing and return implications in the presence of a complex filtering model and highly non-linear dynamics of hidden state variables. Based on the observation that even in the *simplest cases*, the effective computation of asset prices still requires the use of numerical integration methods, and hence implicitly using an approximation of the state space, this article shows that if we approximate the state space to begin with, asset prices can be solved for analytically even under very complex filtering problems. In essence, the discretization of the state space enables me to rewrite the posterior *density* as a posterior *probability distribution*, which can be shown to follow a vector linear process with stochastic volatility. This approximation turns out to be extremely tractable, allowing me to obtain analytical formulas for the conditional expectation of future payoffs even in the presence of a highly nonlinear model for the underlying hidden variable and a utility function that shows "belief-dependency." Closed form solutions for bonds and stock prices and returns are obtained, making it simple to interpret the forces that affect the dynamics of asset returns and interest rates.

The paper is related to a number of recent articles: First, recent literature uses utility functions with external habit formation to describe a preference over consumption relative to other agents' consumption (see e.g. Abel (1990) and Campbell and Cochrane (1999)). In these models, the marginal utility of consumption becomes "state-dependent," where the state is the habit level, because it depends on the distance between current consumption and habit level. This in turn affects asset prices by yielding for example time-varying risk aversion. Second, recent literature has also put forward the concept of "recursive" utility in order to generate a preference for resolution of uncertainty and disentangle the risk aversion from the elasticity of intertemporal substitution (see e.g. Epstein and Zin (1989), Weil (1989), Campbell (1996)). Finally, a number of studies have concentrated on the concept of "Knightian uncertainty" to explain asset prices (see e.g. Epstein and Wang (1995), Maenheut (1999), Hansen et. al (1999), Cagetti et al (2000)).

Belief-dependent utility functions as obtained in this paper have the same intuitive motivations as those in these recent approaches, but it substantially differs from them in many respects. For example, in my empirical application I also find that belief-dependent utility functions yield a marginal utility of consumption that varies over time in response to past innovations in consumption. However, the main difference from a standard habit formation model is that the marginal utility now changes due to changes in the *dispersion* of the belief distribution on the relevant state (say, the habit level). This yields an additional dimension along which asset prices fluctuate, together with a new interpretation in terms of aversion to state uncertainty.

Similarly, by incorporating the overall distribution of beliefs in the utility function I effectively endow investors with a preference for resolution of their own uncertainty on the underlying true state of nature. Although this is different from preferences over early/late resolution of uncertainty as understood in the recent literature, it yields nonetheless similar implications. For example, the elasticity of intertemporal substitution in my approach is still equal to the inverse of the coefficient of relative risk aversion. However, since in my set up investors are also averse to "long run risk," it is still possible to obtain a high equity premium and high volatility of returns without affecting the intertemporal elasticity of substitution.

Finally, although state-uncertainty is rather different from the "Knightian uncertainty" a la' Gilboa and Schmeidler (1993), whereby agents are endowed with families of prior distributions on a given state and then use the max-min rule to take decisions, this paper retains the intuitive appeal that "uncertainty" is *bad* and that agents prefer certainty to uncertainty. In addition, in my set up investors' (unique) posterior distribution can be estimated from data on, say, consumption growth, making it simple to quantify the size of the effects induced by aversion to uncertainty on asset prices.

The article proceeds as follows: Next section introduces the concept of belief-dependent utility functions and aversion to state uncertainty. Section 3 introduces the asset pricing model and proposes a new discretization approach to obtain closed form solutions for asset prices. Section 4 uses the tools developed in the previous section to obtain closed form solutions for asset prices for a class of belief-dependent utility functions. Section 5 specializes the analysis to the case of Constant Relative Risk Aversion and obtains stock returns implications. Section 6 takes the model to the data: After introducing a statistical model for dividends characterized by random jumps in their drift rate (structural breaks), it also takes the model to the data and calibrate the utility parameters to match the unconditional moments for returns. Section 7 concludes.

2 Belief-Dependent Utility Functions

In this first section, I set out the minimum notation necessary to understand the nature of the axiomatic representation of belief-dependent utility functions. The discussion is taken from Myerson (1991), whose axioms and representation theorems are contained in Appendix A. Let $\mathcal{C}$ be a set of prizes and Θ a set of states. A *lottery* $f : \Theta \longrightarrow \Delta(\mathcal{C})$ is a function assigning a probability distribution $\Delta(\mathcal{C})$ on $\mathcal{C}$ to each state $\theta \in \Theta$. For every event $S \subseteq \Theta$, let us denote by $\succeq_S$ a *conditional preference relation* on the set of lotteries on $\mathcal{C}$. Assuming that $\succeq_S$ satisfies the (standard) axioms listed in Appendix A, we then have that there exists a *state-dependent* utility function $u : \mathcal{C} \times \Theta \longrightarrow R$ and a subjective conditional probability function $\pi\left(.|S\right)$ on Θ such that for all lotteries f and g ,

$$f \succeq_S g \iff \sum_{\theta \in S} \pi\left(\theta|S\right) \sum_{c \in \mathcal{C}} f\left(c|\theta\right) u\left(c|\theta\right) \geq \sum_{\theta \in S} \pi\left(\theta|S\right) \sum_{c \in \mathcal{C}} g\left(c|\theta\right) u\left(c|\theta\right) \tag{1}$$

Here, $\pi\left(.|S\right)$ is simply a conditional probability distribution on Θ with unit mass on the event S that satisfies Bayes law.

To better interpret the representation of preferences in (1), consider its specialization to the "constant" lotteries, that is such that $f\left(c|\theta\right) = 1$ for every $\theta \in \Theta$. If we denote such a lottery by $[c]$, then the application of the representation result (1) implies that

$$[c] \succeq_S [c'] \iff \sum_{\theta \in S} \pi(\theta|S)\, u(c|\theta) \geq \sum_{\theta \in S} \pi(\theta|S)\, u(c'|\theta)$$

Since $[c]$ is a constant act and agents "know" $u(c|\theta)$ for all θ, this representation is simply saying that even if an agent obtains a prize c, his/her "subjective" utility from "consuming" c is "belief dependent," in the sense that it depends on the whole subjective distribution $\pi(\theta|S)$ over θ. In other words, since the "uncertainty" over the lottery f may be resolved *before* the uncertainty over the underlying state of nature θ, this approach implies that agents have belief-dependent utility functions.

Examples:

(*i*) *Health-dependent utility functions* (see e.g. Zeckhauser (1970, 1973), Arrow (1974), Viscusi and Evans (1990)): Utility functions have been shown empirically to depend on agents health status, especially in the case of severe health effects. As mentioned in the introduction, for a large set of diseases and risk factors it is more reasonable to assume that agents have only partial information on their own health status, because diagnostic technologies are imprecise and check-ups infrequent. This generates a belief-dependent utility function as in (1) where θ is the health status and $\pi(\theta)$ the subjective probability on it.

(*ii*) *Relative Performance: External Habits and Relative Social Standing*: Recent literature in macroeconomics and finance has explored preferences that explicitly incorporate relative performance. This could be at the consumption level (e.g. Abel (1990), Campbell and Cochrane (1999)), income level (see e.g. Chan and Kogan (2001)) and wealth level (Bakshi and Chen (1996)). In all cases, agents are unlikely to have a precise information about the consumption/wealth level of all the other agents within a reference class, but most likely they only have a probability distribution on it. Again, in this case the utility function becomes belief-dependent utility.

(*iii*) *Money Illusion* (see e.g. Shafir et al. (1997)): The perception of the (marginal) utility of *real* consumption is affected by the level of current and future inflation. If agents have a distribution on the inflation rate (or drift rate of inflation), then preferences are belief-dependent.

Definition: A *belief dependent utility function* over an act f is given by

$$U(f,\pi) = \sum_{\theta\in\Theta} \pi(\theta) \sum_{c\in\mathcal{C}} f(c|\theta)\, u(c|\theta) \tag{2}$$

In particular, a belief dependent utility function over a prize c is

$$U(c,\pi) = \sum_{\theta\in\Theta} \pi(\theta)\, u(c|\theta) \tag{3}$$

2.1 Aversion to State-Uncertainty

Characterization (3) naturally leads to a definition of aversion to "state-uncertainty." In fact, given a prize c we can vary the distribution π over Θ and obtain various levels of "utility." Of interest to us are the changes in the "dispersion" of the probability π while keeping its expected value constant. To this end, it is often used the concept of "mean-preserving spread" (see Ingersoll (1987)) to do comparative statics exercises. Assuming that Θ is a metric space, so that mean-preserving spreads are well defined, I define aversion to state uncertainty as follows:

Definition: *(a)* Let π and c be given. A belief-dependent utility function $U(c,\pi)$ displays *aversion to state-uncertainty given* c if a mean preserving spread $\widehat{\pi}$ on the distribution π yields

$$U(c,\widehat{\pi}) < U(c,\pi) \tag{4}$$

(b) A belief dependent utility function displays *aversion to state uncertainty* if (4) holds for all c.
(c) Similarly, a belief dependent utility function displays *neutrality to state uncertainty* if $U(c,\widehat{\pi}) = U(c,\pi)$ for all c.

Examples revisited: In the context of example (i), consider an agent who has been diagnosed a possibly bad, but not life threatening, disease: if news that rule out both a bad health and a good health status makes the agent happy, then he/she is averse to "health-state" uncertainty. In the context of the relative-performance example (ii) an agent that is happy to learn that everyone else is as good as she is when she only believed to be average displays aversion to state-uncertainty. Finally, in example (iii) if the utility from real consumption is affected non-linearly by the current inflation rate, the belief-dependent utility may display aversion to state-uncertainty.[2]

2.2 Absolute and Relative Risk Aversion of Belief Dependent Utility Functions

In the context of belief-dependent utility functions we can define the usual notions of absolute risk aversion $A(\pi, c)$ and relative risk aversion $\gamma(\pi, c)$ respectively as

$$A(\pi, c) = -\frac{\partial^2 U(c, \pi)/\partial c^2}{\partial U(c, \pi)/\partial c} \text{ and } \gamma(\pi, c) = -\frac{c\partial^2 U(c, \pi)/\partial c^2}{\partial U(c, \pi)/\partial c} \tag{5}$$

These are the analogous notions of absolute and relative risk aversion as in the case of state-independent utility function. Since for given distribution π, the utility function $U(c) = U(c, \pi)$ is a standard Von-Neuman Morgenstern utility function with respect to state-independent lotteries (of which the constant lotteries are a special case), formulas in (5) reflect the *local* curvatures of the utility function that are necessary and sufficient to generate "aversions" to fair bets (either in absolute or in relative terms).

Given their importance in finance applications, I now characterize the belief dependent utility function for the case of constant absolute or relative risk aversion:

Proposition 1: *(a)* $A(c, \pi) = A$ constant if and only if

$$U(c, \pi) = E_t[k_1(\theta)] - E_t[k_2(\theta)]\, e^{-Ac} \tag{6}$$

where $k_2(\theta) > 0$ for all $\theta \in \Theta$ and $E_t[k_i(\theta)] = \sum_{j=1}^{n} \pi_j k_i(\theta_j)$.

(b) $\gamma(c, \pi) = \gamma$ constant if and only if

$$U(c, \pi) = E_t[k_1(\theta)] + E_t[k_2(\theta)]\, \frac{c^{1-\gamma}}{1-\gamma} \tag{7}$$

with $k_2(\theta) > 0$ for all $\theta \in \Theta$.

[2] The reader may wonder at this point whether there exists any independent evidence that agents indeed display aversion to state-uncertainty, such as from experimental data or surveys. To the best of my knowledge, no study on this subject has ever been undertaken. On the other hand, there is evidence supporting state-dependent utilities (see citations in the introduction). Hence, any non-linearity in the state-dependent utility paired with a lack of knowledge of the state would implicitly generate preferences for state-uncertainty.

Proof: See Appendix. ■

It is immediate to see that from Jensen's inequality, the characteristics of the functions $k_1(.)$ and $k_2(.)$ determines whether the utility function displays aversion, loving or neutrality to state-uncertainty. If $k_i(.)$'s are linear, for example, then the utility function are indeed neutral to state-uncertainty: A change in the dispersion of the distribution π would not affect the characteristics of the utility function. In any other case, they will display (possibly locally) state-uncertainty aversion or loving.

A final point is about the more general "belief-dependent" risk aversion. Even though the agent does not have full information about the underlying state of nature, according to Myerson's axiomatic approach he/she may nonetheless establish the level of his/her relative risk aversion conditional on each state being true. That is

$$\gamma(\theta) = -\frac{c\partial^2 u(c|\theta)}{\partial u(c|\theta)}$$

In this case, the representation of utility function is as follows:

Proposition 2: The belief-dependent utility function with belief-dependent risk aversion is given by

$$U(c,\pi) = E_t[k_1(\theta)] + \sum_{i=1}^{n} \pi_i k_2(\theta_i) \frac{c^{1-\gamma(\theta_i)}}{1-\gamma(\theta_i)} \tag{8}$$

Proof: See Appendix. ■

3 A Pure Exchange Economy

I now apply the belief-dependent utility functions to a Lucas (1978) pure exchange economy and derive asset pricing implications under a rather general set-up for the processes for dividends and the "state." To fix ideas, one can think of the latter as an external habit or inflation level as in examples (*ii*) and (*iii*) above (see e.g. Campbell and Cochrane (1999), Brandt and Wang (1999)), or as the state of medical technology, which in turn affects the population health level and hence their utility, as in example (*i*). The empirical section will focus on one particular application to the case of habit formation when the economy is hit by structural breaks. Let $\mathbf{W}_t = (W_{1,t}, W_{2,t})$ be a 2-dimensional Wiener process defined on a complete probability space

$(\Omega, \mathcal{P}^0, \mathcal{F}^0)$. The usual regularity conditions are assumed throughout (see e.g. Duffie (1996), Karatzas and Shreve (1998)). I make the following assumptions about the economy:

Assumption 1: *Real* log-dividends $\delta_t = \log(D_t)$ evolve according to the stochastic differential equation

$$d\delta_t = g_t dt + \boldsymbol{\sigma} d\mathbf{W}_t \tag{9}$$

where g_t is unobservable and its dynamics is described below, and $\boldsymbol{\sigma}$ is a 1×2 constant vector.

Assumption 2: As in section 2, the representative investor utility function is belief-dependent and it has the following form

$$U(c_t, p_{\theta,t}, t) = \int_{\mathcal{R}} p_\theta(\theta|\mathcal{F}_t)\, u(c_t, t|\theta)\, d\theta \tag{10}$$

where θ is an unobservable state whose dynamics is described below, $p_{\theta,t} = p_\theta(\theta|\mathcal{F}_t)$ is the posterior marginal density on θ conditional on investors' information at time t, $\mathcal{F}_t$, and[3]

$$u(c_t, t|\theta) = e^{-\phi t} k(\theta) \frac{c_t^{1-\gamma(\theta)}}{1-\gamma(\theta)} \tag{11}$$

where ϕ is the subjective discount rate.

Assumption 3: $\boldsymbol{\nu}_t = (g_t, \theta_t)$ follows any continuous time, stationary Markov process, instantaneously independent of $\mathbf{W}_t$. Let $p(\boldsymbol{\nu}, t; \boldsymbol{\nu}', t')$ denote the transition probability density that characterizes its law of motion, with $t' > t$.

Assumption 4: Agents observe a noisy signal on θ_t, given by

$$ds_t = \theta_t dt + \boldsymbol{\sigma}_s d\mathbf{W}_t \tag{12}$$

where $\boldsymbol{\sigma}_s$ is a 1×2 constant vector.

Finally, I need an assumption on how investors update their beliefs given the observation of past dividends:

Assumption 5: Given a prior belief at time $t = 0$, $p(\boldsymbol{\nu}|\mathcal{F}_0) = p_0(\boldsymbol{\nu})$, investors rationally update their posterior distribution on $\boldsymbol{\nu}_t$, $p(\boldsymbol{\nu}|\mathcal{F}_t)$, by using Bayes law.

[3] We could use the more general form as in (8) but it turns out it has no consequences on the price of assets. In order to limit the amount of notation, we set $k_1(\theta) = 0$.

Remarks:

(i) The form of the belief-dependent utility function nests the cases of belief-dependent relative risk aversion, constant relative risk aversion ($\gamma(\theta) = \gamma$ for all θ) and belief-independent utility ($k(\theta) = 1$ and $\gamma(\theta) = \gamma$ for all θ).

(ii) Standard state-dependent utility functions can be obtained in the limiting case as $\boldsymbol{\sigma}_s \to 0$, so that θ_t is essentially revealed through the signal (12).

(iii) Assumption 3 on the stochastic behavior of the state variables $\boldsymbol{\nu}_t = (g_t, \theta_t)$ is very general. In particular, I do not require them to follow continuous path processes. In addition, although $\boldsymbol{\nu}_t$ is assumed instantaneously independent of $\mathbf{W}_t$, in the filtered probability space (see Lemma 1) the correlation with (δ_t, s_t) will be high, because innovations on the latter affect the estimate of $\boldsymbol{\nu}_t$.

(iv) Assumption 5 is needed to ensure that the results in the paper do not arise because of an increase in degrees of freedom: Assuming that beliefs must be rational and, in the empirical section, derived from the observation of past data (dividends) makes sure that these are taken exogenously and are not chosen to match the moments of asset returns as we want.

3.1 Equilibrium Pricing

Although the economy is characterized by many sources of risk, namely $\mathbf{W}_t$ and the shocks characterizing $\boldsymbol{\nu}_t$, when we *condition* on the information set $\mathcal{F}_t = \{\delta_\tau, s_\tau : 0 \leq \tau \leq t\}$ we find that all "shocks" to the economy are captured by the innovation process

$$d\widetilde{\mathbf{W}}_t = \mathbf{D} \left[\begin{pmatrix} d\delta_t \\ ds_t \end{pmatrix} - E \begin{pmatrix} d\delta_t \\ ds_t \end{pmatrix} |\mathcal{F}_t \right) dt \right] \tag{13}$$

where $\mathbf{D} = \left((\boldsymbol{\sigma}', \boldsymbol{\sigma}_s')' \right)^{-1}$. It turns out that $\widetilde{\mathbf{W}}_t$ is a standard Wiener process with respect to the filtration $\{\mathcal{F}_t\}$ generated by $\{\delta_\tau, s_\tau : 0 \leq \tau \leq t\}$ (see Lemma 1 below). We shall denote by $(\Omega, \mathcal{P}, \mathcal{F}, \{\mathcal{F}_t\})$ the filtered probability space induced by the $d\widetilde{\mathbf{W}}_t$ on the original probability space. Taking the latter as the primitive probability space and redefining the processes for fundamentals from (13) as

$$\begin{aligned} d\delta_t &= E_t [g|\mathcal{F}_t] \, dt + \boldsymbol{\sigma} d\widetilde{\mathbf{W}}_t \\ ds_t &= E_t [\theta|\mathcal{F}_t] \, dt + \boldsymbol{\sigma}_s d\widetilde{\mathbf{W}}_t \end{aligned}$$

we obtain a standard pure-exchange economy set-up (see Duffie (1996), Karatzas and Shreve (2000)). In this case, it is known that we can always assume the existence of a sufficient number of assets (at least two) in zero net supply to make the markets complete. It follows that any asset paying the stochastic stream $\{q_\tau\}$ of consumption good has a price given by

$$P_t = E_t\left[\int_t^\infty \frac{U_c(c_\tau, p_{\theta,\tau}, \tau)}{U_c(c_t, p_{\theta,t}, t)} q_\tau d\tau\right] = E_t\left[\int_t^\infty e^{-\phi(\tau-t)} \frac{\int_{\mathcal{R}} p_{\theta,\tau}(\theta)\, k(\theta)\, c_\tau^{-\gamma(\theta)} d\theta}{\int_{\mathcal{R}} p_{\theta,t}(\theta)\, k(\theta)\, c_t^{-\gamma(\theta)} d\theta} q_\tau d\tau\right] \tag{14}$$

Unfortunately, solving for this expectation is rather a daunting exercise, even for simple processes for θ_t and g_t and belief-independent utility functions, and numerical techniques must be employed (see e.g. Yan (2000) and Brennan and Xia (2001)). Next section provides a viable analytical alternative to the numerical approach.

3.2 A Discretization Approach

I solve directly for the expectation in (14) by relying on a fine approximation of $\mathcal{R}^2$. Besides obtaining an analytical expression for prices, this approach also yields an analytical solution for returns, which can then be fully characterized theoretically without relying on simulations.

Consider a discretization of the state space $\mathcal{V} = \{\mathbf{v}^1, ..., \mathbf{v}^n\}$ where n is "large" (but finite) and each $\mathbf{v}^i = (g^i, \theta^i)$ is a point on $\mathcal{R}^2$. Hence, by definition, we may have $\mathbf{v}^i$ and $\mathbf{v}^j$ such that either $g^i = g^j$ or $\theta^i = \theta^j$. Since I assumed $\boldsymbol{\nu}_t$ to be stationary, we can always choose the set $\mathcal{V}$ such that $\Pr(\boldsymbol{\nu}_t < \min_i(\mathbf{v}^i))$ and $\Pr(\boldsymbol{\nu}_t > \max_i(\mathbf{v}^i))$ are negligible. $\mathcal{V}$ will be the state space on which I define a continuous time Markov chain process $\{\mathbf{v}_t\}$ which *approximates* the original $\{\boldsymbol{\nu}_t\}$. The process $\{\mathbf{v}_t\}$ can be fully described by its infinitesimal generator $\boldsymbol{\Lambda}$, where $[\boldsymbol{\Lambda}]_{ij} = \lambda_{ij} \geq 0$ for $i \neq j$ and $[\boldsymbol{\Lambda}]_{ii} = \lambda_{ii} = -\sum_{j \neq i} \lambda_{ij}$ (see Karlin and Taylor (1975)). Standard results then show that the transition probability to move from $\mathbf{v}^i$ to $\mathbf{v}^j$ during the time interval τ is given by

$$\Pr(\mathbf{v}_{t+\tau} = \mathbf{v}^j | \mathbf{v}_t = \mathbf{v}^i) = e_i' \cdot \mathbf{U} \cdot \exp(\mathbf{A} \times \tau) \cdot \mathbf{U}^{-1} \cdot e_j \tag{15}$$

where $\mathbf{A}$ is the diagonal matrix with the eigenvalues of $\boldsymbol{\Lambda}$ on its principal diagonal, $\mathbf{U}$ is the matrix of the associated eigenvectors (see Lemma 2 below), e_i is the i-th column of the identity matrix and "$\exp(\mathbf{A} \times \tau)$" denotes the diagonal matrix with ii-th element given by $\exp(\mathbf{A}_{ii} \times \tau)$. Equation (15) prompts the following definition:

Definition: Let $\mathcal{V} = \{\mathbf{v}^1, .., \mathbf{v}^n\}$ be given with interval size equal to h and let Δ be a (small) time interval. I say that the continuous time, discrete-state process $\{\mathbf{v}_t\}$ with infinitesimal generator $\mathbf{\Lambda}$ *approximates* the continuous time, continuous state Markov process $\{\boldsymbol{\nu}_t\}$ defined by its transition density $p(\boldsymbol{\nu}_t, t; \boldsymbol{\nu}_{t'}, t')$ if

$$\Phi_p\left(\mathbf{v}^i, t; \mathbf{E}\left(\mathbf{v}^j\right), t+\Delta\right) \approx e_i' \cdot \mathbf{U} \cdot \exp\left(\mathbf{A} \times \Delta\right) \cdot \mathbf{U}^{-1} \cdot e_j \tag{16}$$

where $\Phi_p\left(\mathbf{v}^i, t; ., t+\Delta\right)$ is the probability measure induced on $\mathcal{R}^2$ from density $p\left(\mathbf{v}^i, t; ., t+\Delta\right)$ and $\mathbf{E}\left(\mathbf{v}^j\right) \subset \mathbf{R}^2$ is a rectangular interval of size h^2 centered in $\mathbf{v}^j$.

In other words, on the discretized grid $\mathcal{V} \times [\Delta, 2\Delta, ...]$ the transition probabilities induced by the original processes and the one approximated through the infinitesimal generator $\mathbf{\Lambda}$ are "close." By choosing n sufficiently large and Δ sufficiently small, for all practical purposes we have that the two processes coincide.[4]

Given this approximation, we can now make use of the following lemma by Liptser and Shyriayev (1977). Let us denote investors' subjective probability that the state is $\mathbf{v}^i$ at time t given their information $\mathcal{F}_t$ by

$$\pi_t^i = \Pr\left(\mathbf{v}_t = \mathbf{v}^i | \mathcal{F}_t\right)$$

This is the discretized version of the density $p(\boldsymbol{\nu}|\mathcal{F}_t)$ introduced earlier. Differently from the latter, the process for the vector $\boldsymbol{\pi}_t = \left(\pi_t^1, ..., \pi_t^n\right)$ can be easily described as a vector diffusion.

[4] I refer the reader to Kushner and Dupuis (2001) for a detailed treatment on the approximation of general continuous time, continuous state processes by continuous time, discrete state processes (see Ch. 4.3, 5.6). Their definition of approximation relies on the notion of "*local consistency*," that is, over the time inverval Δ, the approximating process should have the same local properties as the original process (see pages 71 and 129). The final conclusion, however, is that on the grid $\mathcal{V} \times [\Delta, 2\Delta, ...]$ the transition probabilites coincide. I take the latter directly as a condition of "approximation," since their construction is rather lengthy and not particular useful for the conclusions of this paper.

One can also consult Israel et al. (2001) on issues related to existence, approximations, computations and references. In particular, it turns out that often an *exact* infinitesimal generator for a continuous time process that replicates a discrete-time transition matrix does not exist (see Theorem 3.1 in Israel et al. (2001)). Nonetheless, an *approximate* generator can generally be computed.

Lemma 1: Let the prior distribution $\boldsymbol{\pi}_0 = \widehat{\boldsymbol{\pi}}$, with $\sum_{i=1}^{n} \widehat{\pi}^i = 1$, be given. Then the vector $\boldsymbol{\pi}_t = \left(\pi_t^1, ..., \pi_t^n\right)'$ evolves according to the system of n stochastic differential equations

$$d\boldsymbol{\pi}_t = \boldsymbol{\Lambda}' \cdot \boldsymbol{\pi}_t \; dt + \boldsymbol{\pi}_t \odot \boldsymbol{\sigma}\left(\boldsymbol{\pi}_t\right) \; d\widetilde{\mathbf{W}}_t \tag{17}$$

where $\odot$ indicates the element-by-element multiplication, $\boldsymbol{\sigma}\left(\boldsymbol{\pi}_t\right)$ is a $n \times 2$ vector whose $i-th$ component is

$$\boldsymbol{\sigma}_i\left(\boldsymbol{\pi}_t\right) = \left(\mathbf{v}^i - \overline{\mathbf{v}}_t\right)' \; \mathbf{D}'$$

and where $\overline{\mathbf{v}}_t = \sum_{j=1}^{n} \pi_t^j \mathbf{v}^j$ and $\mathbf{D} = \left(\left(\boldsymbol{\sigma}', \boldsymbol{\sigma}'_s\right)'\right)^{-1}$. In addition, $\widetilde{\mathbf{W}}_t$ defined in (13) is a Wiener process defined on the filtered probability space $(\Omega, \mathcal{P}, \mathcal{F}, \{\mathcal{F}_t\})$ where $\{\mathcal{F}_t\}$ is the filtration generated by (δ_τ, s_τ).

Proof: See Liptser and Shyriayev (1977). ■

Remark: Equation (17) underscores the gain from the discretization approach: Although the original process for the original hidden state $\boldsymbol{\nu}_t = (g_t, \theta_t)$ could be very complex and highly non-linear, the process for the discretized posterior probability distribution $\boldsymbol{\pi}_t$ is simple, being it a vector linear process with stochastic volatility. As next Lemma shows, this property yields analytical formulas for conditional expectations of variables that are relevant for pricing:

Lemma 2: Let β be a constant and define the matrix $\overline{\boldsymbol{\Lambda}}_\beta = \boldsymbol{\Lambda} + \beta \times diag\left(g^1, ..., g^n\right) + \frac{1}{2}\beta^2 \boldsymbol{\sigma}\boldsymbol{\sigma}'$ $\mathbf{I}$. If $\overline{\boldsymbol{\Lambda}}_\beta$ admits n distinct real eigenvalues, then for any $\tau > 0$ and for any $i = 1, ..., n$ we have

$$E\left[c_{t+\tau}^{\beta} \pi_{t+\tau}^{i} | \mathcal{F}_t\right] = c_t^{\beta} \boldsymbol{\pi}'_t \cdot \mathbf{G}_i(\beta, \tau) \tag{18}$$

where $\mathbf{G}_i(\beta, \tau)$ is the $n-$dimensional vector given by $\mathbf{G}_i(\beta, \tau)= \mathbf{U}_\beta \cdot e^{\mathbf{A}_\beta \; \tau} \cdot \mathbf{U}_\beta^{-1} \cdot e'_i$, where $\mathbf{A}_\beta$ is the diagonal matrix with the eigenvalues of $\overline{\boldsymbol{\Lambda}}_\beta$ on its principal diagonal and $\mathbf{U}_\beta$ is the matrix of the associated eigenvectors, and e_i is the $i-$th column of the identity matrix.

Proof (*Sketch*): Defining the $\mathbf{n}_t = \boldsymbol{\pi}_t c_t^{\beta}$ and using Ito's Lemma, one finds that the vector $d\mathbf{n}_t$ also follows a vector linear process as (17), with infinitesimal generator $\overline{\boldsymbol{\Lambda}}_\beta$. The claim follows from standard results on linear vector processes. See Appendix for details.■

Remarks: (i) If $\overline{\boldsymbol{\Lambda}}_\beta$ has either complex or multiple eigenvalues, a solution to the expectation (18) can still be found and it is still linear in the current $\boldsymbol{\pi}_t$. The main difference is that the coefficients of the $\boldsymbol{\pi}_t$ will be horizon dependent and possibly oscillatory.

(ii) The case $\beta = 0$ implies $\overline{\mathbf{\Lambda}}_\beta = \mathbf{\Lambda}$ and the result in (15) is recovered as a special case.

Given that now the probability density $p(\boldsymbol{\nu}|\mathcal{F}_t)$ is approximated by the vector of probabilities $\boldsymbol{\pi}_t$, I will denote the utility function as dependent on $\boldsymbol{\pi}_t$ rather than p_t. For notational convenience, let $\gamma_i = \gamma\left(\mathbf{v}^i\right) = \gamma\left(\theta^i\right)$ and $k_i = k\left(\mathbf{v}^i\right) = k\left(\theta^i\right)$. Let also $\mathbf{k} = (k_1, .., k_n)$. The belief-dependent utility function is now given by $U\left(c_t, \boldsymbol{\pi}_t, t\right) = e^{-\phi t}\sum_{i=1}^n \pi_t^i k_i \frac{c_t^{1-\gamma_i}}{1-\gamma_i}$. It is also convenient to define the marginal utility of consumption in state i (if it was observable) by

$$m^i\left(c_t\right) = k^i c_t^{-\gamma_i} \tag{19}$$

and then the vector $\mathbf{m}\left(c_t\right) = \left(m^1\left(c_t\right), ..., m^n\left(c_t\right)\right)$.

4 Bond and Stock Prices

I now apply Lemmas 1 and 2 to obtain closed form solutions for bonds and stocks. To fully gauge the usefulness of Lemma 2, it is instructive to go through the steps to obtain the price of a zero-coupon bond paying one unit of consumption good at time $t+\tau$. From the pricing formula (14) we see that

$$\begin{aligned} Q_t(\tau) \equiv\ & Q\left(\tau, \boldsymbol{\pi}_t, c_t\right) = E_t\left[\frac{U_c\left(c_{t+\tau}, \boldsymbol{\pi}_{t+\tau}, t+\tau\right)}{U_c\left(c_t, \boldsymbol{\pi}_t, t\right)}\right] = \frac{1}{e^{-\phi t}\cdot \boldsymbol{\pi}_t' \cdot \mathbf{m}\left(c_t\right)} E_t\left[e^{-\phi(t+\tau)}\sum_{i=1}^n k_i \pi_{t+\tau}^i c_{t+\tau}^{-\gamma_i}\right] \\ =\ & \frac{e^{-\phi\tau}}{\boldsymbol{\pi}_t'\cdot \mathbf{m}\left(c_t\right)}\sum_{i=1}^n k_i E_t\left[\pi_{t+\tau}^i c_{t+\tau}^{-\gamma_i}\right] = \frac{e^{-\phi\tau}\boldsymbol{\pi}_t'\cdot \mathbf{G}(-\boldsymbol{\gamma},\tau)\cdot \mathbf{m}\left(c_t\right)}{\boldsymbol{\pi}_t'\cdot \mathbf{m}\left(c_t\right)} \end{aligned} \tag{20}$$

where $\mathbf{G}(-\boldsymbol{\gamma},\tau)$ is the $n \times n$ matrix whose i-th column is $\mathbf{G}_i(-\gamma_i, \tau)$ as defined in Lemma 2. By evaluating $Q\left(\tau, \boldsymbol{\pi}_t, c_t\right)$ at $\boldsymbol{\pi}_t = e_i$, we obtain the price of the bond *conditional* on $\mathbf{v}^i$ being the true state, which then yields the more interpretable formula

$$Q_t(\tau) = \overline{\boldsymbol{\pi}}_t\left(c_t\right)' \cdot \mathbf{Q}(\tau) \tag{21}$$

where $Q_i(\tau) = Q\left(\tau, e_i, c_t\right)$ and

$$\overline{\pi}^i\left(c_t\right) = \frac{\pi_t^i k^i c_t^{-\gamma_i}}{\boldsymbol{\pi}_t'\cdot\mathbf{m}\left(c_t\right)} \tag{22}$$

is a probability distribution on $\mathcal{V}$ that is weighted by the marginal utility across the states $\mathbf{v}^i$'s. Formula (21) generalizes earlier results by Yared (1999) and Veronesi and Yared (2000). Hence, with belief-dependent utility functions the value of one unit of consumption in the future is a weighted average of its discounted value conditional on each of the $\mathbf{v}^i$, where the weights are probabilities $\overline{\pi}^i(c_t)$ that are *adjusted* to reflect the different levels of marginal utility across the unobservable states $\mathbf{v}^i$.[5]

Turning now to the stock prices and interest rates, the appendix shows:

Proposition 3: For each $\ell = 1, ..., n$ and for given c_t let

$$B_\ell(c_t) = \frac{1}{k_\ell c_t^{-\gamma_\ell}} \sum_{i=1}^{n} k_i c_t^{-\gamma_i} e'_\ell \cdot \left(\phi \mathbf{I} - \overline{\mathbf{\Lambda}}_{\beta_i}\right)^{-1} e_i \tag{23}$$

where $\overline{\mathbf{\Lambda}}_{\beta_i}$ is defined in Lemma 2 with $\beta_i = 1 - \gamma_i$. Then:

(a) The price of the risky asset is

$$P_t = D_t \; \left(\overline{\boldsymbol{\pi}}(c_t)' \cdot \mathbf{B}(c_t)\right) \tag{24}$$

(b) The real rate of interest is

$$r_t = \overline{\boldsymbol{\pi}}(c_t)' \cdot \mathbf{K}(c_t) \tag{25}$$

where $\mathbf{K}(c_t)$ is a n dimensional vector, with element j given by

$$K_j(c_t) = \phi + \gamma_j g^j - \frac{1}{2}\gamma_j^2 \boldsymbol{\sigma\sigma'} - C_j^*(c_t)$$

where $C_j^*(c_t) = \sum_{i=1}^{n} \lambda_{ji} \frac{k_i}{k_j} c_t^{\gamma_j - \gamma_i}$.

Proof: See Appendix ■

These asset pricing formulas have a number of properties that I discuss in the next few pages. In the following, I will refer to the functions $B_\ell(c_t)$'s appearing in (23) as *conditional price-dividend ratios*, because each of them is the price-dividend ratio that would occur at time t if there was perfect certainty on the underlying state. Indeed, from (22) we see immediately that if $\pi_t^\ell = 1$ for some i, then $\overline{\pi}^\ell(c_t) = 1$ and hence $P_t/D_t = B_\ell(c_t)$.

I remark that the case of belief-*in*dependent utility functions is obtained as a special case where $k(\theta) = k = 1$ and $\gamma_i = \gamma_j = \gamma$. In this case, we obtain the following result, that generalizes a result in Veronesi (2000) and Veronesi and Yared (2000):

[5] Clearly, the distribution in formula (22) can be interpreted as a state-price density, where the states are the $\mathbf{v}^i$'s in $\mathcal{V}$.

Corollary 1: In the case of state-independent utilities, let $\overline{\mathbf{B}} = \left(\phi \mathbf{I} - \overline{\mathbf{\Lambda}}_{1-\gamma}\right)^{-1} \cdot \mathbf{1}_n$ and $Q_i^B(\tau) = e^{-\phi\tau} \mathbf{G}_i'(-\gamma, \tau) \cdot \mathbf{1}_n$.Then:

$$P_t^B = D_t \left(\boldsymbol{\pi}_t' \cdot \overline{\mathbf{B}}\right) \tag{26}$$

$$r_t^B = \phi + \gamma \ \boldsymbol{\pi}_t' \cdot \mathbf{g} - \frac{1}{2}\gamma^2 \boldsymbol{\sigma}\boldsymbol{\sigma}' \tag{27}$$

$$Q_t^B(\tau) = \boldsymbol{\pi}_t{}' \cdot \mathbf{Q}^B(\tau) \tag{28}$$

I will use (26) and (27) as "benchmark" cases to gauge the effect of the belief-dependency. In particular, the following remarks are worth pointing out:

Remarks: (*i*) For $\gamma > 1$ the conditional price-dividend ratios $\overline{B}_i$'s in Corollary 1 are typically *decreasing* with g^i, i.e. a *higher* growth rate of the economy is associated with a *lower* price-dividend ratio, due to the low elasticity of intertemporal substitution (see e.g. Campbell (1999), Veronesi (2000) for a discussion).[6]

(*ii*) As emphasized in Veronesi (2000), this effects makes it impossible to generate a high risk premium under learning, because the market return tends to be a good hedge against bad news in consumption.

(*iii*) Since this problem is due to the low elasticity of intertemporal substitution (EIS), Epstein-Zin-Weil preferences that disentangle risk aversion from EIS do not provide help unless one is ready to assume that EIS>1. Macroeconomic studies seem to agree that EIS<1 (see e.g. Campbell (1999) for a discussion).

5 Asset Prices and Returns under CRRA

In this paper I only concentrate on the case where the relative risk aversion is constant (see Proposition 1). Expressions for returns similar to the ones below can be obtained in the more general case at the expense of more notation and little gain from an intuitive point of view.

[6] This is not absolutely true and depends on the transition probabilities λ_{ij}. However, this holds under the assumptions made in section 6, where the model is taken to the data. See also Figure 1.

5.1 Asset Prices

In the case of constant relative risk aversion, we have $\gamma_i = \gamma_j = \gamma$. Recall the convenient notation $\mathbf{k} = \left(k^1, ..., k^n\right)$ and $\mathbf{g} = \left(g^1, ..., g^n\right)$. We obtain the following corollary to Proposition 3.

Corollary 2: Let $\overline{\mathbf{\Lambda}}_{1-\gamma}$ be defined as in Lemma 2. Define the $n-$dimensional vectors

$$\overline{\boldsymbol{\pi}}_t = \frac{\mathbf{k} \odot \boldsymbol{\pi}_t}{\mathbf{k}' \cdot \boldsymbol{\pi}_t} \tag{29}$$

$$\mathbf{B} = \mathbf{k}^{-1} \odot \left(\phi \mathbf{I} - \overline{\mathbf{\Lambda}}_{1-\gamma}\right)^{-1} \mathbf{k} \tag{30}$$

with $\mathbf{k}^{-1} = \left(k_1^{-1}, .., k_n^{-1}\right)'$. Then: *(a)* The stock price is given by:

$$P_t = D_t \, \overline{\boldsymbol{\pi}}_t' \cdot \mathbf{B} \tag{31}$$

(b) The zero-coupon bond price with maturity $t + \tau$ is

$$Q_t(\tau) = \overline{\boldsymbol{\pi}}_t' \cdot \mathbf{Q}(\tau) \tag{32}$$

where $\mathbf{Q}(\tau) = e^{-\phi\tau} \, \mathbf{k}^{-1} \odot \mathbf{G}(-\gamma, \tau) \cdot \mathbf{k}$ and $\mathbf{G}(-\gamma, \tau)$ is as defined in Lemma 2.

Comparing the pricing formulas obtained here with the corresponding ones in Corollary 1 for the benchmark case, we see that the effect of a belief-dependent utility function shows itself in two terms in the pricing functions (31) - (32) compared to the benchmark cases (26) and (28). Concentrating on the stock price (a similar discussion holds for the bond) we have: First, belief-dependent utilities affects the "conditional price-dividend ratios" B_i as it can be seen by comparing (30) with the same expression in Corollary 1. Intuitively, the marginal utility of consumption is now belief-dependent. Hence, even when we condition on a particular state (i.e. we set $\pi_t^i = 1$ for some i), the state-dependent marginal utility affects the comparison between current and future marginal utilities, thereby affecting the conditional price-dividend ratio B_i. Second, when $\boldsymbol{\pi}_t$ is a non-degenerate distribution, the conditional price-dividend ratios B_i are weighted by the probabilities $\overline{\pi}_t^i$ rather than the original π_t^i. Intuitively, the probabilities $\overline{\pi}_t^i$ are now adjusted for the impact that each state θ^i has on the marginal utility of consumption, namely for k_i. In other words, if $k_i > k_j$, then $\overline{\pi}_t^i$ becomes relatively bigger than $\overline{\pi}_t^j$: Stock prices now reflect more those states characterized by higher marginal utility of consumption. Similar comments hold for the bond price.

5.2 Stock Returns

Before discussing the results about stock returns and interest rates, it is convenient to introduce the following quantities: For $x = g, \theta$ define:

$$V_t^{k,x} = \frac{Cov_t\left(k\left(\theta\right), x\right)}{E_t\left[k\left(\theta\right)\right]} \tag{33}$$

$$V_t^{kB,x} = \frac{Cov_t\left(k\left(\theta\right) B(\theta, g), x\right)}{E_t\left[k\left(\theta\right) B(\theta, g)\right]} \tag{34}$$

where for clarity, I emphasize now that the conditional price-dividend ratios B_i implicitly depend on the underlying values of θ^i and g^i, that is, $B_i = B\left(\theta^i, g^i\right)$. Notice that given $\boldsymbol{\pi}_t$, $\mathbf{k}$ and $\mathbf{B}$ from Corollary 2, $V_t^{k,x}$ and $V_t^{kB,x}$ have analytical formulas. To provide an intuition for $V_t^{k,x}$ and $V_t^{kB,x}$, consider the case where $x = \theta$, for example. Then, $V_t^{k,\theta}$ can be interpreted to measure two important components of the belief-dependent utility function formulation: First, the degree of "state-dependency" as measured by the behavior of the function $k\left(\theta\right)$: If $k\left(\theta\right) = k = \text{constant}$, then $V_t^{k,\theta} = 0$. Second, the degree of uncertainty on θ_t. If the distribution $\boldsymbol{\pi}_t$ gives probability one to one particular state θ^ℓ, then again $V_t^{k,\theta} = 0$. Similarly, for $x = g$, we have that $V_t^{k,g} = 0$ if either there exists ℓ such that $\pi_t^\ell = 1$, or $k\left(\theta\right) = k = \text{constant}$, or θ_t and g_t are uncorrelated. It is easy to see that $V_t^{k,x}$ is positive or negative depending on $Cov_t\left(k\left(\theta\right), x\right)$. In particular, $V_t^{k,\theta} < 0$ if $k\left(\theta\right)$ is decreasing in θ (and $\boldsymbol{\pi}_t$ is non-degenerate). A similar intuitive argument holds also for $V_t^{kB,x}$, with the additional caveat that it will depend also on the behavior of the *equilibrium* conditional price-dividend ratio.

Corollary 3: The interest rate is given by

$$r_t = r_t^B + \gamma V_t^{k,g} - \overline{\boldsymbol{\pi}}_t' \cdot \mathbf{C}^* \tag{35}$$

where r_t^B is given in (27), $\mathbf{C}^* = \mathbf{k}^{-1} \odot \left(\boldsymbol{\Lambda} \cdot \mathbf{k}\right)$ and, again, $\mathbf{k}^{-1} = \left(k_1^{-1}, ..., k_n^{-1}\right)$.

Hence, the interest rate is affected by the state-dependent utility function formulation through the two additional terms, namely, $V_t^{k,g}$ and $\overline{\boldsymbol{\pi}}_t' \cdot \mathbf{C}^*$. Notice that $V_t^{k,g} < 0$ if g_t and $k\left(\theta_t\right)$ tend to be negatively correlated. This is intuitive. When g_t and $k\left(\theta_t\right)$ are negatively correlated, we have that a negative shock to consumption growth occurs when the marginal utility of consumption increases because $k\left(\theta_t\right)$ increases. This increase the demand for bonds, which in turn decreases the interest rate. Notice that a corollary of this argument shows that if $g_t = g$ is a known constant (which is a special case of the set up in Section 3), then $V_t^{k,g} = 0$. Even if $V_t^{k,g} = 0$, the second term $\overline{\boldsymbol{\pi}}_t' \cdot \mathbf{C}^*$ is generally different from zero. This measures the predictable changes in state (through the transition matrix $\boldsymbol{\Lambda}$), which yield a predictable variation in future marginal utilities (through the $k(\theta)$ function) which again will generally call for an increase in the demand for bonds.

Turning now to stock returns, let us denote the *excess* return as

$$dR_t = \frac{dP_t + D_t dt}{P_t} - r_t dt$$

I then obtain the following result:

Proposition 5: The excess return dR_t follows the process

$$dR_t = \mu_{R,t} dt + \boldsymbol{\sigma}_{R,t} d\widetilde{\mathbf{W}}_t$$

with

$$\mu_{R,t} = \gamma \left(\boldsymbol{\sigma}\boldsymbol{\sigma}' + \Delta V_t^g \right) - V_t^{k,g} - \left(V_t^{k,g}, V_t^{k.\theta} \right) \mathbf{D}'\mathbf{D} \left(\Delta V_t^g, \Delta V_t^\theta \right)' \tag{36}$$

$$\boldsymbol{\sigma}_{R,t} = \boldsymbol{\sigma} + \left(\Delta V_t^g, \Delta V_t^\theta \right) \mathbf{D}' \tag{37}$$

where, for $x = g, \theta$ we have $\Delta V_t^x = V_t^{kB,x} - V_t^{k,x}$

Proof: See Appendix. ■

From the formulas (36) and (37), it is clear that stock returns are characterize mainly by the quantities $V_t^{k,x}$ and ΔV_t^x, for $x = g, \theta$. As I mentioned already, the former is positive or negative depending on whether $k(\theta)$ and "x" are positively or negatively related. The following Lemma characterizes ΔV_t^x: To state the lemma, it is convenient to return to the notation introduced in Section 3. That is, let $p(\theta, g|\mathcal{F}_t)$ denote the posterior density on (θ, g), and $B(\theta, g)$ denote the conditional price dividend ratio at the value (θ, g).

Lemma 3: Assume $p(\theta, g|\mathcal{F}_t)$ is non-degenerate. Then, if $B(\theta, g)$ is monotonically non-decreasing in (θ, g), then both $\Delta V_t^g \geq 0$ and $\Delta V_t^\theta \geq 0$.

Proof: See Appendix. ■

Remarks:

(*i*) In the benchmark case with state-*in*dependent utilities, $k(\theta)$ is constant, which yields $V_t^{k,x} = 0$ for $x = g, \theta$. Hence, the second and third term in (36) vanish. In addition, $\Delta V_t^g = V_t^{B,g} < 0$ because, as explained, $B(g)$ is typically decreasing for $\gamma > 1$ (see Figure 1 and Veronesi (2000)). Hence, high returns cannot be obtained in this case.

(*ii*) From Lemma 3, if a high drift rate of dividends g and a high state θ generates a high price-dividend ratio, then both $\Delta V_t^g \geq 0$ and $\Delta V_t^\theta \geq 0$. If in addition, $k(\theta)$ is decreasing and $Cov_t(\theta, g) > 0$, then also $V_t^{k,g} \leq$ and $V_t^{k,\theta} \leq 0$ which yields unambiguously higher returns and volatility than in the benchmark case discussed in point (*i*).

(*iii*) In the case of perfect certainty on (θ, g), we have $V_t^{y,x} = 0$ for $y = k, kB$ and $x = g, \theta$ yielding an expected return equal to $\gamma \boldsymbol{\sigma}\boldsymbol{\sigma}'$ and volatility $\boldsymbol{\sigma}$. In this case, we see that expected returns are constant and no excess/stochastic volatility obtains.

Finally, I can obtain the dynamics for the risk-free rate:

Proposition 6: The risk-free rate evolves according to the process

$$dr_t = \mu_r(\boldsymbol{\pi}_t)\,dt + \sigma_r(\boldsymbol{\pi}_t)\,d\widetilde{\mathbf{W}}_t \tag{38}$$

where μ_r is given in the appendix and

$$\sigma_r(\boldsymbol{\pi}_t) = \left(r - \phi - \frac{1}{2}\gamma^2\sigma^2\right)\left[V_t^{r,g} - V_t^{k,g}, V_t^{r,\theta} - V_t^{k,\theta}\right]\mathbf{D}' \tag{39}$$

where for $x = g, \theta$ $V_t^{r,x} = Cov_{\pi_t}(C^r, x)/E^{\pi_t}[C^r]$ and $C_i^r = \gamma k_i \theta_i - k_i C_i^*$.

5.3 Habit Formation

It is useful to interpret the results above in light of the popular external habit formation model. Consider the utility index

$$u(c_t, t|X_t) = e^{-\phi t}\frac{(c_t - X_t)^{1-\gamma}}{1-\gamma} \tag{40}$$

where X_t is a slow-moving external habit, as in e.g. Abel (1990) and Campbell and Cochrane (1999). As discussed in Section 2, since X_t is an *external* index on the "Joneses" consumption (see Abel (1990)) it is likely that each investor does not actually observe X_t, but only has distribution $p_t(X_t)$ on it obtained from signals. From Section 2, preferences must then be described by $E_p[u(c_t,t|X_t)]$ yielding "belief-dependency". In addition, the curvature of the utility function (40), implies that a mean preserving spread $\widetilde{p}(X)$ on the density $p(X)$ decreases the agent utility. That is, for given consumption level c_t at time t, $E_{\widetilde{p}}[u(c_t,t|X_t)] < E_p[u(c_t,t|X_t)]$. This implies that preferences with habit formation naturally generate an *aversion to state-uncertainty*. Intuitively, this is because with a higher dispersion of beliefs, there is marginally more probability to be close to X_t, which is the point of maximum disutility.

In addition, using the same argument as in Campbell and Cochrane (1999), we can also see that economic expansions will tend to increase the relative consumption $S_t = (c_t - X_t)/c_t$ while the opposite happens in contractions, because X_t is a slow-moving index and hence c_t moves faster than X_t. Noticing that uncertainty on X_t then implies uncertainty on S_t, we can follow Campbell and Cochrane (1999) and rewrite the marginal utility implied by (40) as $u_c(c_t,t|X_t) = e^{-\phi t} S_t^{-\gamma} c_t^{-\gamma}$. Identifying now S_t with θ_t in the previous notation, we obtain that indeed S_t and g_t are positively related, while $k(S_t) = S_t^{-\gamma}$ is decreasing in S_t.[7] This implies that with state uncertainty, $V_t^{k,\theta} < 0$ and $V_t^{k,g} < 0$. Campbell and Cochrane (1999) assume g constant, which if known implies $V_t^{k,g} = \Delta V_t^g = 0$, but show that $B(S_t,g)$ is increasing in S_t, which implies $\Delta V_t^\theta > 0$. From (36) and (37) this yields an expected return and a volatility that are higher than the ones one would obtain without state (or belief) dependency. Wachter (2000) explores the implications of Campbell and Cochrane (1999) model when g_t is time varying (as in assumption 3 in this paper) and finds that $B(S_t,g_t)$ is still increasing in S_t but slightly decreasing in g_t (although it is almost unaffected by it). In this case the sign of ΔV_t^g is ambiguous but still $\Delta V_t^\theta \geq 0$. As I show below in the context of a simple example, if g_t and S_t are perfectly correlated, then still $\Delta V_t^g \geq 0$ yielding again a high equity premium and a high volatility.

[7] For the results in the previous sections to hold, we really need that the *marginal* utility of consumption is written in a multiplicative form, not the *level* of utility. Under the Campbell and Cochrane (1999) formulation, the relevant function $k(S)$ is $k(S) = S^{-\gamma}$ and not $k(S) = S^{1-\gamma}$. Either way, the results follow.

6 Empirical Implications

In this section I take the model to the data. I will focus on a very special case of the more general set up developed above, namely, the case where the state affecting individual preferences is a monotonic function of the drift rate of dividends. To fix ideas, one could interpret this set up as an extreme case of the habit formation model discussed in the previous section: In fact, the surplus $S_t = (c_t - X_t)/c_t$ introduced by Campbell and Cochrane (1999) is a procyclical variable, which increases when consumption growth is high. Since in the model described earlier the economic activity is determined by the dividend drift rate g_t, we can keep the interpretation of the habit formation model by assuming $S_t = G(g_t)$ for some function $G(.)$ with $G'(g) > 0$.[8] Since investors are learning about g over time, this also implies that the expected surplus $E_t[S_t]$ is perfectly correlated with news in consumption growth, as the habit formation specification requires.

6.1 Structural Breaks and Long-Term Risk Aversion

I also assume that the economy is hit by structural breaks, that is, random changes of the drift rate of dividends.[9] This assumption yields a stochastic movement in the dispersion of investors' beliefs on the dividend drift rate g_t, a movement over time that provides us with an alternative explanation for the changes of the conditional moments of stock returns. In fact, if g_t changes infrequently, then its value determines the average consumption path for a rather long period of time. Recall now that because the unknown drift rate g_t affects investors preferences and they display aversion to "state-uncertainty" (see section 5.3), the structural breaks assumption yields what can be termed "aversion to dispersion of the long run average consumption," or, "*aversion to long-run risk.*" I will return on this interpretation while commenting the empirical results, where I will also compare this model to the case where g_t follows a slow moving mean reverting process.

Turning to the modeling, let the drift rate of dividends (consumption) follow the pure jump

[8] The assumption is extreme, because ideally we would like to have S_t change smoothly with changes in g_t. However, the latter assumption loses in tractability. In addition, the expected surplus $E_t[S_t]$ (which is what investors observe given their ignorance of X_t) is much smoother than g_t itself.

[9] See Veronesi (2000) and Timmerman (2001) for similar models applied to stock prices.

process

$$dg_t = (J_t - g_t)\, dQ_t^{\alpha(g_t)} \tag{41}$$

where $dQ_t^{\alpha(g_t)}$ denotes the increment of a Poisson process with intensity $\alpha(g_t)$ and J_t is a random variable with any density $f(g)$. As explained, I can transform the process (41) into a continuous time, discrete-state process as follows: Select first a fine grid $\mathcal{V} = \{g^1, .., g^n\}$ with boundaries g^1 and g^n chosen such that $\Pr(J_t < g^1) = P(J_t > g^n) \approx 0$. Given the grid $\mathcal{V}$, define $\alpha_i = p(g^i)$ and $f_i = F(g^i + h/2) - F(g^i - h/2)$, where h is the interval size of the grid and $F(.)$ is the cumulative distribution on $f(.)$. Finally, obtain the $n \times n$ infinitesimal matrix $\mathbf{\Lambda}$ as $\lambda_{ij} = \alpha_i f_j$ for $j \neq i$ and $\lambda_{ii} = \alpha_i f_i - \alpha_i$. In this case we have the following corollary:

Corollary 5: Let $\lambda_{ij} = \alpha_i f_j$ for $j \neq i$ and $\lambda_{ii} = \alpha_i f_i - \alpha_i$. Define the two constants

$$H_1 = \sum_{i=1}^{n} \frac{f_i k_i}{\phi + \alpha_i - (1-\gamma) g^i - \frac{1}{2}(1-\gamma)^2 \boldsymbol{\sigma\sigma'}} \text{ and } H_2 = \sum_{i=1}^{n} \frac{f_i \alpha_i}{\phi + \alpha_i - (1-\gamma) g^i - \frac{1}{2}(1-\gamma)^2 \boldsymbol{\sigma\sigma'}}$$

Then, for all $i = 1, ..., n$:

$$B_i = \frac{1 - H_2 + \frac{\alpha_i}{k_i} H_1}{\left(\phi + \alpha_i - (1-\gamma) g^i - \frac{1}{2}(1-\gamma)^2 \boldsymbol{\sigma\sigma'}\right)(1 - H_2)}$$

Proof: See Appendix. ■

Notice that for $\gamma > 1$, if $k_i = 1$ for all i (i.e., no belief dependent utility), then the price dividend ratio conditional on state i is decreasing with g^i unless also α_i is increasing with i. As mentioned in Remark (i) after proposition 3, this is due to the low elasticity of intertemporal substitution. As a consequence, the function k_i that enters into the utility function $U(c_t, \boldsymbol{\pi}_t, t) = e^{-\phi t} \sum_{i=1}^{n} \pi_t^i k_i c_t^{1-\gamma} / (1-\gamma)$ has the effect of increasing the conditional price-dividend ratio B_i as g^i increases as long as $\alpha_i > 0$.

To describe further the properties of the conditional price-dividend ratios B_i, I need more structure on the function $k(g)$ and the density $f(g)$ of the jump variable J_t. I assume $J_t \sim \mathcal{N}(\mu, \sigma_J)$ and

$$k(g) = e^{-\rho(g - g^n)}$$

This form of the function ensures the restrictions imposed by the habit formation interpretation of the model are met (that is, that the implied $S_t^{-\gamma} = k(g_t) > 1$). The parameter ρ defines the curvature of $k(g)$ and can be interpreted as a measure of *aversion to state uncertainty.* A high level of ρ implies that a mean preserving spread on the posterior $p(g|\mathcal{F}_t)$ will have a higher impact on the representative agent's utility (and marginal utility). The case $\rho = 0$ corresponds to the state-independent utility function investigated in Veronesi (2000) which will help making comparisons. Figure 1 plots the conditional price-dividend ratios B_i under the parameter estimates in Table 1 A, $\gamma = 2$ and for three values of $\rho = 0, 20, 40$. We can see that as the aversion to state uncertainty increases, the conditional price-dividend ratio moves from being negatively sloped with respect to g^i to being positively sloped. Hence, increases in consumption growth would reduce the price-dividend ratio in the base case $\rho = 0$ but it would increase it for $\rho = 40$.

6.2 Fitting Real Consumption Data

In this section I estimate the simple model in the previous section using quarterly data on real consumption growth from 1946-1999. Consumption data are from the NIPA tables and include only non-durables and services. Nominal per-capita data have been deflated using the CPI index.

The approximation approach described in section 4.2 suggests also a two step methodology to estimate the parameters of the model: (i) Given the approximating grid $\mathcal{V}$ and the time step of observations Δ (=1 quarter), estimate the parameters of the discrete-time, discrete-state model. This can be accomplished by using the (approximate) Maximum Likelihood methodology put forward by Kitagawa (1987) or, as a generalized case, by Hamilton (1989). (ii) Given the estimated parameters, use the estimated transition probability $\mathbf{\Lambda}^*$ with $\lambda^*_{ij} = \Pr(g_{t+\Delta} = g^j | g_t = g^i)$ and look for the infinitesimal generator $\mathbf{\Lambda}$ that yields the same transition probability $\mathbf{\Lambda}^*$, that is, such that

$$\mathbf{U} \exp(\mathbf{A} \times \Delta) \mathbf{U}^{-1} = \mathbf{\Lambda}^*$$

where $\mathbf{A}$ and $\mathbf{U}$ are the $n \times n$ matrices with eigenvalues on the principal diagonal and column eigenvectors, respectively. (see Section 4.1). Israel et al. (2001) discuss issues related to existence and an algorithm to obtain the infinitesimal generator.

I use the methodology above to estimate the parameters of the discretized version of the pure jump model with constant jump probability and an Ornstein-Uhlenbeck process. The results are reported in Tables 1 and 2, respectively. I defer the discussion of the Ornstein-Uhlenbeck process to Section 6.4.

Panel A of Table 1 reports the estimation results for the pure jump process (41), where the mean μ of the jump distribution $f(g) = \mathcal{N}\left(\mu, \sigma_J^2\right)$ is restricted to equal the long-run mean consumption growth. In fact, Panel B of the same table shows that the estimated value of $\mu = -0.0006$, implying a negative long-term growth rate of consumption. Since this point estimate is non-significant, it seems a better approach to fix the long-term consumption growth to the sample value $\mu = .0049$.[10] The estimates shows that the (quarterly) jump probability $\alpha = 0.0454$, implying a "jump" approximately every five years. The (quarterly) volatility of consumption growth has been estimated at $\sigma = 0.0067$ slightly below to the sample standard deviation of consumption growth, which is equal to 0.0082. Of course, this is to be expected when there are jumps in the drift as part of the variability of consumption is now captured by the jumps themselves. Finally, we see from Table 1 that if a "jump" occurs, the new drift is chosen according with the distribution $f(g) = \mathcal{N}\left(\mu, \sigma_J^2\right)$ with $\sigma_J = 0.0133$. This is a rather large value which implies that given the average growth rate of consumption equal to $\mu = 0.0049$ the new drift is in the range $[-.0218, 0.0352]$ with probability 95%. Figure 2 plots the density distribution. Although such a wide range in the stationary distribution of the drift parameters is due to the first few years in the sample (see Panel C), they still indicate that a somewhat large "uncertainty" may ensue due to "jumps" in the drift rate of consumption. Notice also that the parameter σ_J is significant at the 5% level, even after correcting the standard errors for heteroskedasticity and autocorrelation (and even more strongly so if we let μ to be estimated from data). Finally, it is known that consumption growth was much more volatile for the pre-war data sample, which unfortunately is not available at the quarterly frequency. However, if agents used past data to determine the jump distribution $f(g)$, they would have probably estimated a distribution even more dispersed than the one I obtained in

[10] The likelihood ratio statistics for the null that μ equals in fact the latter value is 1.602, which entails that $\mu = .0049$ cannot be rejected in favor of the unrestricted model. In any case, the same implications about stock returns and prices are obtained as in Table 3 by assuming the higher discount rate.

Table 1A. For these reasons, I will mainly refer to the estimates in Panel A to discuss the asset pricing implications of the model.

Figure 3 reports the evolution of the posterior distribution $\boldsymbol{\pi}_t = \left(\pi_t^1, ..., \pi_t^n\right)$ from 1946-1999.[11] It is rather interesting to note the wide fluctuations in the dispersion of the posterior distribution over time. The intuition is rather simple: When a jump occurs (or agents believe it occurred) it takes some time to learn the new drift. During this period of time the posterior distribution widens. This effect is even more evident in Figure 4, which plots the time series of mean drift rate $E_t(g) = \sum_{i=1}^n \pi_t^i g^i$ (Panel A) and the root mean square error of agents distribution $RMSE_t = \sqrt{E_t(g^2) - E_t(g)^2}$ (Panel (B)). The figure clearly shows a high variability of expected drift in the first two years of the sample accompanied by a high root MSE. The "uncertainty" decreased in the 50s and 60s but recovered in the 70s and especially the beginning of the 80s. Higher uncertainty was again realized in the 90s. It is to this uncertainty that the agents in the economy are averse to.

Finally, Panel C of Table 1 reports the estimates of the same model, but for the shorter sample 1952-1999. As it can be seen, once the highly volatile consumption at the end of the 40's is left out, the standard deviation of the jump distribution is the much smaller $\sigma_J = .0033$. On the other hand, now the probability of a jump raised to $\alpha = .1837$, that is, about a jump every 6 quarters in average. I will comment further on the asset pricing implications of these other estimates at the end of next section.

6.3 Implications for Asset Prices

Given the estimated parameters for the underlying consumption process in Panel A of Table 1, we need to assess the size of the equity risk premium, volatility, risk-free rate and its volatility. Panel A of Table 3 reports the average ex-post mean values for stock excess returns,

[11] Given the discretization methodology used to estimate the model, the posterior distribution can be computed easily by Bayes law (see e.g. Hamilton (1989)):

$$\pi_i(t+1) = \frac{e^{-\frac{1}{2\sigma^2}(\Delta c(t+1)-\theta_i)^2}\left[\pi(t)\boldsymbol{\Lambda}^*\right]_i}{\sum_{j=1}^n e^{-\frac{1}{2\sigma^2}\left(\Delta c(t+1)-\theta_j\right)^2}\left[\pi(t)\boldsymbol{\Lambda}^*\right]_j}$$

where $\Delta c(t) = \ln(c(t+1)) - \ln(c(t))$ is the real consumption growth and $\boldsymbol{\Lambda}^*$ is now the transition probability matrix for the discrete time case.

the volatility of returns and the volatility of the real rate (the latter computed by estimating future inflation by one quarter ahead predictive regression). The level of the real rate is set to (annualized) 2% but it is not estimated from data because estimates of the inflation risk premium are not available.

Panel B reports the values for the same quantities for the pure jump process described in (41). Since all the conditional moments are stochastic in this case and depend on $\boldsymbol{\pi}_t$, the values in the table correspond to the case where $\boldsymbol{\pi}_t \sim N(\mu, \sigma_\pi)$, where $\mu = 0.0049$ is the long-term average of consumption growth and σ_π corresponds to the time-average Root Mean Square Error of the empirical posterior distribution.

Several effects should be noticed. First, the case $\rho = 0$ corresponds to the case of belief-independent utilities, also studied in Veronesi (2000), which we may take as a benchmark. Notice that $E[dR] < 0$ and that an increase in the coefficient of risk aversion γ *decreases* the excess return. As explained in Remark (*ii*) after Proposition 4 and in Veronesi (2000), learning on the drift rate of consumption induces a negative covariance between returns and consumption growth, due to the hedging demand for $\gamma > 1$. The natural hedge that results yields a low (and even negative) equity premium.

Second, an increase in the coefficient of "long-run risk aversion", ρ, yields an increase in the risk premium and in the volatility of stock returns and a decrease in the risk-free rate. The volatility of the risk-free rate becomes "negative" indicating a negative covariance between changes in the risk-free rate and the innovations in consumption growth. Its absolute value denotes the level of volatility, which then it is shown to increase as ρ increases. Interestingly, an increase in ρ also increase the average price-dividend ratio: This is due to the increase in the convexity of $k(g)$ that yields an increase in the convexity of B_i in (31) and hence of the price-dividend ratio.

Third, a close match with the data can be obtained by assuming that $\gamma = 2$, ρ slightly above 50 and ϕ about 0.04. It is worth pointing out that a coefficient $\gamma = 2$ implies a value of the intertemporal elasticity of substitution equal to 0.5, which is well within the acceptable region. The latter comment also emphasizes one distinction of the present approach with respect to Epstein-Zin preferences: Under E-Z preferences and a time-varying consumption growth process, the price-dividend ratio results *countercyclical* unless one assumes an elasticity of intertemporal substitution greater than 1 (see e.g. Bansal and Yaron (2000)). Although arguments can be made to convince that indeed the elasticity of intertemporal substitution is greater than 1, much of the macroeconomic evidence seem to suggest that a value around 0.5 is more appropriate (See e.g. Campbell (1999)). The current model does not suffer from this problem because the extra "kick" to generate high equity premium and return volatility is obtained through the aversion to long-term risk parameter ρ.

To emphasize the model ability to match also conditional moments, Figures 5 plots the actual ex-post stock return volatility from 1946 to 1999 and the one implied by the model from consumption data. The ex-post volatility is computed as the standard deviation of daily stock returns in the relevant quarter while the model volatility is the one implied by formula (37) for the posterior probability $\boldsymbol{\pi}_t$ in the same quarter. Clearly, the model overpredicted the volatility at the beginning of the sample and underpredicted the volatility at the end of the sample. However, over the 50 years of sample, the volatility moved in a similar fashion as the one in the data. Figure 6 plots the time series of price-dividend ratios implied by the model. Although the match is not as good as in the case of volatility, I notice that the model generated price-dividend ratio has the same dispersion (at least until the mid 90s) of the realized one as it fluctuate between 15 and 42.

Turning finally to the implications of the estimates for the smaller sample 1952-1999, Table 4 B reports the results of a calibration similar to the one in Table 3 B, although $\rho = 80$ and 100. In this case we see that it is harder for the model to match all the moments at the same time. In particular, we have the following findings: First, there is trade-off between volatility of stock returns and equity premium: a volatility higher than the one actually realized is necessary to obtain the same level of equity premium. In other words, the Sharpe ratio is too low for any ρ. Secondly, the level of the real rate is also slightly higher than the one in the "acceptable region" although, instead, its volatility is remarkably low. At the conditional level, as in Figure 4 the model still presents a time-varying volatility, although less correlated to with the one observed empirically (the relative Figure is omitted for brevity). Although in this case the model does not reproduce exactly all the moments, it still goes a long way compared to the belief-independent utilities or the case of perfect certainty. I will comment below on the implications for the Ornstein-Uhlembeck process.

A final remark is in order: Given that the results above could be interpreted within a habit formation model such as the one put forward by Campbell and Cochrane (1999), it is worth commenting on the differences in more detail. First, my model has *fewer* degrees of freedom than the one of Campbell and Cochrane (1999). In fact, the statistical properties of $E_t\left[S_t^{-\gamma}\right] = E_t\left[k(g)\right]$ which affect the marginal utility of consumption are obtained from data, with the only degree of freedom being the parameter ρ which regulates the aversion to state uncertainty. Second, the interpretation of the results is different: In Campbell and Cochrane (1999), for example, the conditional volatility of stock returns is dictated by an exogenously defined sensitivity function $\lambda\left(S_t\right)$, which is assumed to be decreasing in S_t. This sensitivity function $\lambda\left(S_t\right)$, although useful to fix the interest rate in their model, does not have a clear interpretation. This implies that effectively they are assuming that during bad times (S_t low), $\lambda\left(S_t\right)$ is high and hence return volatility is high. In my model it is the *estimated* conditional distribution $p\left(g_t|\mathcal{F}_t\right)$ on the unobservable state variable g_t that affects volatility. Since the filtered distribution $p\left(g_t|\mathcal{F}_t\right)$ obtained from consumption data turns out to be more dispersed during periods where the market volatility was also higher, the explanation about stochastic volatility has to do with an increase in *aversion to long-run risk*. This is obtained endogenously and not exogenously. Third, the fewer degrees of freedom in my model makes it more difficult to match all the moments in the financial data. However, the close match obtained in Table 3 (and to a less extent, Table 4) is therefore to be considered a success of the model.

6.4 Pure Jumps versus Slow Mean Reversion

I conclude this empirical section with some remarks on the choice of the pure jump model (41) for the drift of the consumption process. The key feature of this assumption is that it entails a stochastic process for the Root Mean Square Error of the posterior probability distribution. This in turn is important to generate the changes in "uncertainty" and hence much of the action described in the previous section. An important question is whether such a model describes well the data. For comparison purposes Table 2 reports the estimates of a standard

Ornstein-Uhlenbeck process:

$$dg_t = (\mu - bg_t)\, dt + \sigma_g dW_{2,t} \tag{42}$$

with $dW_{2,t}dW_t = 0$ and $b = 1 - a$. This process has been widely used in the past to model the (hidden or not) drift rate of consumption or dividends (see e.g. Campbell (1999), Brennan and Xia (2001), Yan (2000), Wachter (2001)). The model has been estimated by Maximum Likelihood using the same technique as to estimate the pure jump process (see Kitagawa (1987)). In this case, the estimation naturally reduces to the use of the Kalman Filter. The question is whether consumption data are described better by the pure jump process (41) or the Ornstein-Uhlenbeck process (42).

Given that the models are non-nested, standard tests cannot be performed. Nonetheless, the Akaike Information Criterion (Akaike (1973)) can be used to give a first assessment of the two models: Since the number of parameters is the same, we see that the higher value of the likelihood function in the case of the pure jump model indicates a lower value of the Akaike index, yielding some support for the pure jump process.[12] In addition, we can also look at the forecasting errors in the two models. In the case of the OU process, the Gaussian assumption (necessary for the Kalman filter to work) entails that residuals should be normally distributed. Instead, they show a rather high negative skewness and an extremely high kurtosis, which would tend to reject the Gaussian model. Similar numbers are obtained for the pure-jump process, but in this case, the one-step ahead observation of Δc_t is given by a mixture of normals, which would indeed entail a higher kurtosis. Interestingly, the skewness of residuals of the pure jump process is closer to zero, as the symmetry of the model would suggest.

The last two columns of Tables 1 and 2 report the in-sample Root Mean Square Prediction Error (RMSE) and the Mean Absolute Prediction Error (MAE) in the two models, respectively. Under this metric we see that the Kalman filter (42) performs marginally better than the pure jump model. This is to be expected, given that the jumps are *unpredictable* in the pure jump

[12] Of course, comparing the values of Likelihood functions is not appropriate in general without the interpretation of Akaike (1973) in terms of relative entropy of two distributions. In addition, it is known that the Akaike Information Criterios (AIC) is affected by some biases. However, see Kitagawa (1973a,b) for a defense of the AIC for model selection, supported by MC experiments, in a set up similar to the one of this paper (Gaussian process against non-Gaussian process).

model. Any small autocorrelation in the data will be captured better by the Kalman filter than by a pure jump model. In any case, the difference in prediction errors does not seem to be high, with the ratios of RMSEs and MAEs in the two model being 1.0827 and 1.0671, respectively. Similar comments apply for the results relative to the shorter sample 1952-1999, contained in Panel C of Tables 1 and 2 for the two models.

In terms of the ability of the Ornstein-Uhlenbeck process to yield equilibrium asset prices and returns that match the unconditional moments in the data, Panel C of Table 3 reports the calibration results. In this calibration, the infinitesimal generator $\mathbf{\Lambda}$ has been chosen to approximate the discrete-time transition density $p(g, t; g', t')$ implicit in the estimates of the model (see e.g. Israel et al (2001)). We see that compared to the pure jump model we need a higher coefficient of aversion to long-run risk to match the data ($\rho = 100$), although in this case the return volatility results slightly too high. In addition, the model would imply very little variation in the conditional moments of stock returns: This is because under the Ornstein-Uhlenbeck process (42) and the Kalman-Bucy filter, the Root Mean Square Error of investors' posterior distribution is a deterministic function that converges quickly to its long term value. This implies that both the expected returns and the volatility of returns are basically fixed at their unconditional values (the relative figures have been omitted for brevity).

Finally, for the smaller sample 1952-1999 we see from Table 4, Panel C, that the results for the unconditional moments of returns are very similar to the ones obtained under the Pure-Jump process (Panel B). However, conditionally, the O-U process still implies a basically constant volatility, while the Pure-Jump process still generates a time-varying conditional volatility as the empirical evidence suggests.

7 Conclusions and Extensions

In this article I re-interpreted standard axioms in choice theory delivering state-dependent utility functions to introduce the notion of a "*belief*-dependent" utility function, that is, a utility function that depends on the subjective beliefs on an underlying partially observable state of nature. I show that this interpretation naturally leads to a notion of "*aversion to state-uncertainty*," that is, aversion to a wider subjective distribution on the underlying (unobservable) state.

I then apply this type of preferences to a standard pure exchange economy under a rather general process for the dividend drift and the "state," whose economic meaning is left unspecified at this stage for generality. Using a new discretization approach, I obtain analytical expressions for prices and returns for both bonds and stocks and find conditions under which aversion to state uncertainty yields higher expected returns and volatility, and lower interest rates. Indeed, these effects take place, for example, if the state is positively correlated with the drift rate of consumption and negatively correlated with the agents' marginal utility, a situation that occurs for example in external habit formation models.

In the empirical section, I specialize the model to the case where the "state" and the drift rate of dividends are perfectly correlated, a situation that is consistent with standard external habit formation models where agents are unaware of other agents' consumption. In addition, I assume that the economy is hit by structural breaks, which yield a time-variation in the dispersion of the agents posterior density on the underlying drift rate of dividends. The latter assumption is of particular interest because it leads to an additional interpretation of the concept of "aversion to state uncertainty," namely, *"aversion to long run risk"*. This is because the drift rate of consumption is responsible for the long-term average consumption path. As a consequence, an "uncertainty averse" investor is averse to the dispersion of long-term average consumption paths. In other words, while "risk aversion" applies to the local volatility of consumption, "aversion to long-run risk" applies also to the dispersion of drifts, which by their own nature have a longer term connotation.

In this set up, when fitted to real consumption data, I show that posterior distributions contain a good deal of uncertainty on the current drift of consumption, a finding which strengthen the notion that investors may be "averse" to this dispersion. Beside matching the unconditional moments of stock returns and the interest rate process, I also show that the time-series of model-generated conditional volatility and price-dividend ratios are broadly consistent with the ones observed in the financial data.

The model and the methodology developed in this paper have a number of potentially interesting applications: First, the specification of the "state" could be different from the drift rate of dividends (consumption), as assumed in the empirical part of this paper. Asset pricing implications of "money illusion" for example could be derived by assuming that the state θ_t is instead an unobservable drift rate of inflation, which investors learn over time. Similarly, if θ_t is instead a latent health status (or risk factor), implications for insurance and portfolio decisions could be obtained which would depend on the belief on the underlying state. (How much are we willing to pay to know our own health status? If we are uncertain about it, what is our exposure to other risks, such us stock holding?) Second, tests could be developed to separate out time-varying risk aversion ($\gamma(\theta)$ non constant) from time varying marginal utility ($k(\theta)$ non constant, as in this paper), all within the same framework of belief-dependent utility functions. Indeed, as already mentioned, one can obtain formulas for asset returns that similar to the ones developed in Section 5 for CRRA even with non-constant $\gamma(\theta)$. Third, the analytical formulas make it simple to estimate the model from financial data using for example GMM estimation methods. This can be accomplished by stacking the "scores" of the Log Likelihood function obtained from Kitagawa (1987) or Hamilton (1989) (see Section 6), together with other moments for financial variables.[13] Tests of the applications just suggested can then be devised in a relatively simple manner. Fourth, the model developed in this paper used the pricing kernel implied by the marginal utility of a representative agent. However, the same methodology can be applied by just assuming directly its implied (belief-dependent) stochastic discount factor. The richness of the dynamics and yet the tractability of the stochastic discount factor with belief-dependency could prove useful to fit cross-sectional bond data using analytical formulas similar to the ones developed in the text, thereby providing an alternative model for the term structure to the one currently available.

[13] A previous version of the paper contained estimates of the pure-jump process obtained using this methodology. For brevity, that part has been eliminated from this version.

8 References

Abel, Andrew (1990) "Asset prices under habit formation and catching up with the Joneses," *American Economic Review Papers and Proceedings*, 80, 38-42.

Akaike H. (1973) "Information Theory and an Extension of the Maximum Likelihood Principle," in *Second International Symposium on Information Theory*, Eds. B.N. Petrov and F. Caski, Budapest.

Arrow, Kenneth J. (1974) "Optimal insurance and generalized deductible" *Scandinavian Actuarial Journal*, 1-42. Reprinted in *Collected Papers of Kenneth J. Arrow, Volume 3: Individual Choice under Certainty and Uncertainty,* Cambridge University Press, 1984, Cambridge: MA.

Bakshi, Gurdip and Zhiwu Chen (1996) "The spirit of capitalism and stock market prices" *American Economic Review,* 86, 1, 133-157.

Bansal R, and A. Yaron (2000) "Risk for the Long Run: A Potential Resolution of Asset Pricing Puzzles" Working Paper, Duke University.

Barberis, Nicholas, Ming Huang and Tano Santos (2000) "Prospect theory and asset prices," forthcoming in the *Quarterly Journal of Economics.*

Brandt, Michael W. and Kevin Q. Wang (1999) "Time-varying risk aversion, unexpected inflation and the dynamics of the term structure" Wharton Working Paper, University of Philadelphia.

Brennan, Michael and Yihong Xia (2001) "Stock price volatility and the equity premium" *Journal of Monetary Economics, Forthcoming.*

Cagetti Marco, Lars P. Hansen, Thomas Sargent and Noha Williamn (2000), "Robustness and pricing with uncertain growth," Working Paper, University of Chicago.

Campbell, John Y. (1996) "Understanding Risk and Returns," *Journal of Political Economy*, 104, 2 298-345.

Campbell, John Y. (1999) "Asset prices, consumption and the business cycle" in John Taylor and Michael Woodford eds. *Handbook of Macroeconomics*, Vol. 1 (North Holland, Amsterdam).

Campbell John Y. and John Cochrane (1999) "By force of habit: A consumption-based explanation of aggregate stock market behavior," *Journal of Political Economy*, 107, 205-251.

Campbell John Y. and N. Gregory Makiew (1990) "Permanent Income, Current Income, and Consumption," *Journal of Business and Economic Statistics*, 8,3,265-180.

Campbell John Y. and Luis Viceira (2001) "Strategic Asset Allocation: Portfolio Choice for Long Term Investors," Manuscript.

Constantinides, George (1991) "Habit formation: A resolution of the equity premium puzzle," *Journal of Political Economy*, 98, 842-862.

Cutler, David M. and Elizabeth Richardson (1997) "Measuring the Health of the U.S. Population," *Brooking Papers on Economic Activity*, Microeconomics, Volume 1997, 217-271.

David, Alexander (1997) "Fluctuating confidence in stock markets: Implications for returns and volatility," *Journal of Financial and Quantitative Analysis*, 32, 4, 427-462.

Duffie, Darrell (1996), *Dynamic Asset Pricing Theory*, Princeton University Press, Princeton, New Jersey.

Epstein, Lawrence and Tan Wang (1994) "Intertemporal Asset Pricing Under Knightian Uncertainty," *Econometrica*, 62, 2, 283-322.

Epstein, Lawrence and Stanley Zin (1989) "Substitution, risk aversion, and the temporal behavior of consumption and asset returns," *Econometrica*, 57, 937-969.

Gilboa Itzack and David Schmeidler (1989) "Maxmin Expected Utility with Non-Unique Prior," *Journal of Mathematical Economics*, 18, 141-153.

Gilboa Itzack and David Schmeidler (1993) "Updating Ambiguous Beliefs," *Journal of Economic Theory*, 59, 33-49.

Hamilton, James D. (1989), "A new approach to the analysis of nonstationary returns and the business cycle." *Econometrica*, 57, 357-384.

Hansen, Lars P., Thomas Sargent and Thomas Tallarini Jr. (1999) "Robust Permanent Income and Pricing," *Review of Economic Studies*, 66, 873-907.

Ingersoll, Jonathan E., (1987), *Theory of Financial Decision Making* (Rowman & Littlefield).

Israel Robert B., Jeffrey S. Rosenthal and Jason Z. Wei (2001) "Finding Generators for Markov Chains via Empirical Transition Matrices, with Applications to Credit Ratings," *Math-*

ematical Finance, 11,2, 245-265.

Karatzas, Ioannis and Steven E. Shreve (1998), *Methods of Mathematical Finance*, Springer-Verlag, New York.

Karlin S., and H.M. Taylor (1975), *A First Course in Stochastic Processes*, Academic Press, New York.

Karni Edi, (1985) *Decision Making under Uncertainty: The Case of State-Dependent Preferences*, Harvard University Press, Cambridge, MA.

Kitagawa, Genshiro (1987a) "Non-Gaussian State-Space Modeling of Nonstationary Time Series" *Journal of the American Statistical Association: Theory and Methods*, 82, 400, 1032 - 1041

Kitagawa, Genshiro (1987b) "Non-Gaussian State-Space Modeling of Nonstationary Time Series: A Rejoinder" *Journal of the American Statistical Association: Theory and Methods*, 82, 400, 1060 - 1063

Kreps, David M. and E.L. Porteus (1978) "Temporal resolution of uncertainty and dynamic choice theory," *Econometrica*, 46, 185-200.

Kushner Harold J. and Paul Dupuis (2001), *Numerical Methods for Stochastic Control Problems in Continuous Time*, Second Edition, Springer, New York.

Liptser, Robert.S. and Albert N. Shiryayev (1977), *Statistics of Random Processes: I, II*, Springer-Verlag, New York.

Lucas, Robert E. (1978) "Asset prices in an exchange economy," *Econometrica*, 46, 1429-1445.

Maenhout, Pascal (1999) "Robust portfolio choice and asset pricing," Working paper, Harvard University.

Myerson, Roger B (1991) *Game Theory: Analysis of Conflicts*, Harvard University Press, Cambridge, Massachusetts.

Shafir, Eldar, Peter Diamond and Amos Tversky (1997) "Money Illusion", *Quarterly Journal of Economics*, 112, 2, 341-374.

Skiadas, Costis (1997) "Subjective Probability under Additive Aggregation of Conditional Preferences" *Journal of Economic Theory*, 76, 2, 242-271.

Timmerman, Allan (1993) "How learning in financial markets generates excess volatility and predictability of stock returns" *Quarterly Journal of Economics*, 108, 1135-1145.

Timmerman, Allan (2001) "Structural Breaks, Incomplete Information and Stock Prices" *Journal of Business and Economic Statistics*, (forthcoming).

Veronesi, Pietro (1999) "Stock market overreaction to bad news in good times: A rational expectations equilibrium model," *Review of Financial Studies,* Vol. 12, 5.

Veronesi, Pietro (2000) "How does information quality affect stock returns?" *Journal of Finance.*

Veronesi, Pietro and Francis Yared (1999) "Short and Long Horizon Term and Inflation Risk Premia in the US Term Structure: Evidence from an Integrated Model for Nominal and Real Bond Prices under Regime Shifts," CRSP Working Paper, University of Chicago - GSB.

Wachter, Jessica A. (2001) "Habit Formation and Returns on Bonds and Stocks," Working Paper, New York University, Stern School of Business.

Weil, Philippe (1989) "The Equity Premium Puzzle and the Risk-free Rate Puzzle," *Journal of Monetary Economics*, 24, 401-421.

Yan, Hong (2000) "Uncertainty growth prospects, learning and asset prices," Working Paper, University of Texas at Austin..

Yared, Francis (1999) "Path-Dependence in Expected Inflation: Evidence from a New Term Structure Model," Ph.D. dissertation, University of Chicago.

Zechauser, Richard J. (1970) "Medical insurance: A case study of the trade off between risk spreading and appropriate incentives" *Journal of Economic Theory*, 2, 1, 10-26.

Zechauser, Richard J. (1973) "Coverage for Catastrophic Illness" *Public Policy*, 21, 149-172.

9 Appendix A: Axiomatic Foundation of State-dependent Utility

There are a number of axiomatic approches to state-dependent preferences (see e.g Karni (1985)). I will refer to the Myerson (1991) set of axioms. I start by introducing some notation:

For every finite set Z, let $\Delta(Z)$ be the set of probability distributions over Z, that is

$$\Delta(Z) = \left\{ q : Z \to R^{|Z|} | \sum_{y \in Z} q(y) = 1 \text{ and } q(y) \geq 0 \right\}$$

Let $\mathcal{C}$ be the set of possible prizes (= consequences = consumption) the decision maker could get. Let Θ be the set of possible states. I define a lottery to be any function f that specifies a nonnegative real number $f(c|\theta)$ for every $c \in \mathcal{C}$ and for every $\theta \in \Theta$, such that $\sum_{c \in \mathcal{C}} f(c|\theta) = 1$. That is

$$L = \{f : \Theta \to \Delta(\mathcal{C})\}$$

I will denote by $[c]$ the lotteries giving probability one to the prize $c \in \mathcal{C}$.

I will assume that conditional on each event $S \in \Theta$, the agent will be able to rank lotteries *conditional* on the event S being true. That is, given any two lotteries f and $g \in L$, I will denote $f \succ_S g$ to mean that the agent strictly prefers the lottery f to the lottery g if the event S were true. Similarly, $f \succeq_S g$ denotes weak preference. I will denote Ξ the set of all events in Θ.

Finally, for every two lotteries f and g and scalar $\alpha \in [0,1]$ I will denote by $f\alpha g = \alpha f + (1-\alpha) g$ the lottery assigning probability $\alpha f(c|\theta) + (1-\alpha) g(c|\theta)$ to every $c \in \mathcal{C}$ and for every $\theta \in \Theta$.

The following are Myerson (1991) axioms

Axiom 1.1: (a - *Completeness*) $f \succeq_S g$ or $g \succeq_S f$ and (b - *Transitivity*) $f \succeq_S g$ and $g \succeq_S h$ then $f \succeq_S h$.

Axiom 1.2: (*Relevance*) If $f(.|\theta) = g(.|\theta)$ for all $\theta \in S$, then $f \sim_S g$.

Axiom 1.3: (*Monotonicity*) If $f \succ_S h$ and $0 \leq \beta < \alpha \leq 1$, then $f\alpha h \succ_S f\beta h$.

Axiom 1.4: (*Continuity*) If $f \succeq_S g$ and $g \succeq_S h$, then there exists $\alpha \in [0,1]$ such that $f\alpha h \sim_S g$.

Axiom 1.5: (a - *Objective Substitution*) If $e \succeq_S f$ and $g \succeq_S h$ and $\alpha \in [0,1]$, then $e\alpha g \succeq_S f\alpha h$. (b - *Strict Objective Substitution*) If $e \succ_S f$ and $g \succeq_S h$ and $\alpha \in (0,1]$, then $e\alpha g \succ_S f\alpha h$.

Axiom 1.6: (a - *Subjective Substitution*) If $f \succeq_S g$ and $f \succeq_T g$ and $S \cap T = \varnothing$, then $f \succeq_{S \cup T} g$; (b - Strict Subjective Substitution) If $f \succ_S g$ and $f \succeq_T g$ and $S \cap T = \varnothing$, then $f \succ_{S \cup T} g$;

Axiom 1.7: (*Interest*) For every state in $\theta \in \Theta$, there exist prizes y and z such that $[y] \succ_\theta [z]$

Before stating the representation theorem, I need the following definition:

Definition: A *Conditional Probability Function* on Θ is any function $\pi : \Xi \to \Delta(\Theta)$ such that for every $S \in \Xi$, $\pi(.|S)$ is a well defined probability function, such that $\pi(\theta|S) = 0$ if $\theta \notin S$ and $\sum_{\theta \in S} \pi(\theta|S) = 1$.

The following representation theorem is proved by Myerson (1991), among others.

Theorem 1: Axioms 1.1 - 1.7 are satisfied if and only if there exists a utility function $u : \mathcal{C} \times \Theta \to R$ and a conditional probability function $\pi : \Xi \longrightarrow \Delta(\Theta)$ such that

(I) $\max_{c \in \mathcal{C}} u(c,\theta) = 1$ and $\min_{c \in \mathcal{C}} u(c,\theta) = 0$

(II) For all R, S, T such that $R \subseteq S \subseteq T \subseteq \Theta$ and $S \neq \varnothing$ we have

$$\pi(R|T) = \pi(R|S)\,\pi(S|T)$$

(III) For all $f, g \in L$ and for all $S \in \Xi$ we have

$$f \succeq_S g \Longleftrightarrow \sum_{\theta \in S} \pi(\theta|S) \sum_{c \in \mathcal{C}} f(c|\theta) u(c|\theta) > \sum_{\theta \in S} \pi(\theta|S) \sum_{c \in \mathcal{C}} g(c|\theta) u(c|\theta) \tag{43}$$

Proof: See Myerson (1991). ■

For completeness, I also state the axiom that provides state-independent utility functions and the representation theorem:

Axiom 1.8: (*State Neutrality*) For every two states θ and θ', if $f(.|\theta) = f(.|\theta')$, $g(.|\theta) = g(.|\theta')$ and $f \succeq_\theta g$ then $f \succeq_{\theta'} g$.

In this case, we have the following:

Theorem 2: Axioms 1.1 - 1.8 are satisfies if an only if there exists a utility function $u : \mathcal{C} \to R$ and a conditional probability function $\pi : \Xi \longrightarrow \Delta(\Theta)$ such that (I) and (II) in Theorem 1 are satisfied and in addition

(IV) For all $f, g \in L$ and for all $S \in \Xi$ we have

$$f \succeq_S g \Longleftrightarrow \sum_{\theta \in S} \pi(\theta|S) \sum_{c \in \mathcal{C}} f(c|\theta) u(c) \geq \sum_{\theta \in S} \pi(\theta|S) \sum_{c \in \mathcal{C}} g(c|\theta) u(c)$$

Proof: See Myerson (1991). ■

One important caveat is that the representation (43) is not unique in the sense that also a conditional probability system $\widehat{\pi}(.|.)$ and a state dependent utility function $\widehat{u}(.|.)$ represents the same conditional preferences over lotteries if (and only if) there exists a positive number A and a function $B : S \to R$ such that

$$\widehat{\pi}(\theta|S) \widehat{u}(c|\theta) = A\pi(\theta|S) u(c|\theta) + B(\theta)$$

(see Myerson (1991, Theorem 1.2)). However, Skiadas (1997) provides a set of axioms able to uniquely identify the conditional probability and the state-dependent utility function (see his Theorem 1, point (a) for a representation as in (43) and point (b) for the uniqueness of the probability and utility representation).

10 Appendix B: Proofs of Propositions

Proof of Proposition 1: (a) By definition

$$\frac{\partial^2 U(c,\pi)}{\partial c^2} = -A \frac{\partial U(c,\pi)}{\partial c} \tag{44}$$

Since $U(c,\pi) = \sum_{\theta \in \Theta} U(c|\theta) \pi(\theta)$ we have $\partial U(c,\pi)/\partial c = \sum_\theta \pi(\theta) \partial u(c|\theta)/\partial c$ and $\partial^2 U(c,\pi)/\partial c^2 = \sum_\theta \pi(\theta) \partial^2 u(c|\theta)/\partial c^2$. Hence, we can rewrite (44)

$$\sum_\theta \pi(\theta) \left(\frac{\partial^2 u(c|\theta)}{\partial c^2} + A \frac{\partial u(c|\theta)}{\partial c} \right) = 0$$

This is true for all $\pi(\theta)$ if and only if

$$\frac{\partial^2 u(c|\theta)}{\partial c^2} = -A\frac{\partial u(c|\theta)}{\partial c}$$

This is a simple second order differential equation, whose solution is

$$u(c|\theta) = k_1(\theta) + k_2(\theta)e^{-Ac}$$

(b) Using the definition we have

$$-c\frac{\partial^2 U(c,\pi)}{\partial c^2} = \gamma\frac{\partial U(c,\pi)}{\partial c} \tag{45}$$

Hence, (45) can be rewritten

$$\sum_\theta \pi(\theta)\left(\frac{\partial u(c|\theta)}{\partial c}\gamma + c\frac{\partial^2 u(c|\theta)}{\partial c^2}\right) = 0$$

Since this must hold for all θ, we must have

$$\frac{\partial u(c|\theta)}{\partial c}\gamma + c\frac{\partial^2 u(c|\theta)}{\partial c^2} = 0$$

Let $V(c|\theta) = \frac{\partial U(c|\theta)}{\partial c}$, so that

$$\frac{\partial V(c|\theta)}{\partial c} = -\frac{V(c|\theta)}{c}\gamma$$

The solution to this differential equation is

$$V(c|\theta) = k_2(\theta)c^{-\gamma}$$

Hence, integrating $V(c|\theta)$ over c we obtain

$$U(c|\theta) = \begin{cases} k_1(\theta) + k_2(\theta)\frac{c^{1-\gamma}}{1-\gamma} & \text{if} \quad \gamma \neq 1 \\ k_1(\theta) + k_2(\theta)\ln(c) & \text{if} \quad \gamma = 1 \end{cases}$$

■

Proof of Proposition 2: From the proof of proposition 1 (b) it is immediate to see that we could let γ be function of θ, i.e. $\gamma = \gamma(\theta)$, obtaining the representation

$$U(c|\theta) = \begin{cases} k_1(\theta) + k_2(\theta)\frac{c^{1-\gamma(\theta)}}{1-\gamma(\theta)} & \text{if} \quad \gamma(\theta) \neq 1 \\ k_1(\theta) + k_2(\theta)\ln(c) & \text{if} \quad \gamma(\theta) = 1 \end{cases}$$

■

Proof of Proposition 3. To prove this proposition, I need the following lemma first (Lemma 2 in the text):

Lemma 2: For all $i = 1, .., n$, denote $n_t^{\beta,i} = c_t^{\beta}\pi_t^i$ where β is a constant. Define also the matrix $\overline{\mathbf{\Lambda}}_\beta = \mathbf{\Lambda} + \beta \times diag\left(g^1, ..., g^n\right) + \frac{1}{2}\beta^2 \boldsymbol{\sigma\sigma'}\ \mathbf{I}$. Then for any $u > t$ we have

$$E\left[n_u^{\beta,i}|\mathcal{F}_t\right] = \sum_{\ell=1}^{N} n_t^{\beta,\ell} \sum_{j=1}^{N} w\left(\beta\right)_{jk}^{-1} w_{ij}\left(\beta\right) e^{\omega_j(\beta)(u-t)}$$

where $\omega_j\left(\beta\right)$ are the eigenvalues of $\overline{\mathbf{\Lambda}}'_\beta$ and $w_{ij}\left(\beta\right)$ are associated eigenvectors and $w\left(\beta\right)_{ij}^{-1} = [\mathbf{W}^{-1}]_{ij}$.

Proof of Lemma 2: Since $c_t = D_t = \exp\left(\delta_t\right)$, we have

$$\frac{dc_t}{c_t} = \left(\sum_{j=1}^{n} \pi_t^j g^j + \frac{1}{2}\boldsymbol{\sigma\sigma'}\right) dt + \boldsymbol{\sigma} d\widetilde{\mathbf{W}}_t = \mu_c\left(\boldsymbol{\pi}_t\right) dt + \boldsymbol{\sigma} d\widetilde{\mathbf{W}}_t$$

It is convenient to work with the *row* vector $\mathbf{n}_t^\beta = \left(n_t^{\beta,1}, .., n_t^{\beta,n}\right)$. By Ito's lemma

$$\begin{aligned} dn_t^{\beta,i} &= \left\{\left[\mathbf{n}_t^\beta \cdot \mathbf{\Lambda}\right]_i + \beta n_t^{\beta,i}\mu_c\left(\boldsymbol{\pi}_t\right) + \frac{1}{2}\beta\left(\beta-1\right) n_t^{\beta,i}\boldsymbol{\sigma\sigma'}\right\} dt \\ &\quad + n_t^{\beta,i}\left(\mathbf{v}^i - \overline{\mathbf{v}}_t\right)' \mathbf{D}' d\widetilde{\mathbf{W}}_t + \beta n_t^{\beta,i}\boldsymbol{\sigma} d\widetilde{\mathbf{W}}_t + \beta n_t^{\beta,i}\left(\mathbf{v}^i - \overline{\mathbf{v}}_t\right) \mathbf{D}'\boldsymbol{\sigma}' dt \\ &= \left\{\left[\mathbf{n}_t^\beta \cdot \mathbf{\Lambda}\right]_i + \beta n_t^{\beta,i}\mu_c\left(\boldsymbol{\pi}_t\right) + \frac{1}{2}\beta\left(\beta-1\right) n_t^{\beta,i}\boldsymbol{\sigma\sigma'}\right\} dt \\ &\quad + n_t^{\beta,i}\left(\mathbf{v}^i - \overline{\mathbf{v}}_t\right)' \mathbf{D}' d\widetilde{\mathbf{W}}_t + \beta n_t^{\beta,i}\boldsymbol{\sigma} d\widetilde{\mathbf{W}}_t + \beta n_t^{\beta,i}\left(\mathbf{v}^i - \overline{\mathbf{v}}_t\right)\left(1,0\right)' dt \end{aligned}$$

where the last line stems from the definition of $\mathbf{D} = \left(\left(\boldsymbol{\sigma}', \boldsymbol{\sigma}_s'\right)'\right)^{-1}$ so that $\mathbf{D}'\boldsymbol{\sigma}' = \left(\left(\boldsymbol{\sigma}', \boldsymbol{\sigma}_s'\right)\right)^{-1}\left(\boldsymbol{\sigma}', \boldsymbol{\sigma}_s'\right)\left(1,0\right) = \mathbf{I}\times\left(1,0\right) = \left(1,0\right)$. From $\mu_c\left(\boldsymbol{\pi}_t\right) = \sum_{j=1}^{n}\pi_t^j g^j + \frac{1}{2}\boldsymbol{\sigma\sigma'}$ and $\left(\mathbf{v}^i - \overline{\mathbf{v}}_t\right)\left(1,0\right)' = g^i - \sum_{j=1}^{n}\pi_t^j g^j$ we obtain

$$dn_t^{\beta,i} = \left\{\left[\mathbf{n}_t^\beta \cdot \mathbf{\Lambda}\right]_i + \frac{1}{2}\beta^2 n_t^{\beta,i}\boldsymbol{\sigma\sigma'} + \beta n_t^{\beta,i} g^i\right\} dt + n_t^{\beta,i}\left(\boldsymbol{\sigma} + \left(\mathbf{v}^i - \overline{\mathbf{v}}_t\right)' \mathbf{D}'\right) d\widetilde{\mathbf{W}}_t \qquad (46)$$

We can write this in vector form:

$$d\mathbf{n}_t^{\beta\,\prime} = \left(\mathbf{n}_t^\beta \overline{\mathbf{\Lambda}}_\beta\right)' dt + \mathbf{n}_t^{\beta\,\prime} \odot \mathbf{\Sigma}\left(\boldsymbol{\pi}_t\right) d\widetilde{\mathbf{W}}_t \qquad (47)$$

where $\overline{\mathbf{\Lambda}}_\beta$ is given in the claim of the Lemma and $\mathbf{\Sigma}\left(\boldsymbol{\pi}_t\right)$ is some bounded $n \times 2$ matrix. Let

$$\widetilde{\mathbf{n}}_u^\beta = E\left[\mathbf{n}_u^\beta|\mathcal{F}\left(t\right)\right]$$

We can write (47) in integral form

$$\mathbf{n}_u^{\beta\,\prime} = \mathbf{n}_t^{\beta\,\prime} + \int_t^u \left(\mathbf{n}_s^\beta \overline{\mathbf{\Lambda}}_\beta\right)' ds + \int_t^u \mathbf{n}_s^{\beta\,\prime} \odot \mathbf{\Sigma}\left(\boldsymbol{\pi}_s\right) d\widetilde{\mathbf{W}}_s$$

Taking expectations on both sides and using the fact that the stochastic integral has zero expectations, we then have (by also transposing)

$$\widetilde{\mathbf{n}}_u^\beta = \mathbf{n}_t^\beta + \int_t^u \widetilde{\mathbf{n}}_s^\beta \overline{\mathbf{\Lambda}}_\beta ds$$

This can we rewritten as

$$d\widetilde{\mathbf{n}}_t^{\beta} = \widetilde{\mathbf{n}}_t^{\beta}\overline{\mathbf{\Lambda}}_{\beta}dt$$

Assuming the matrix $\overline{\mathbf{\Lambda}}_{\beta}$ has n distinct real eigenvalues, the solution to this system of ordinary differential equations with initial condition $\widetilde{\mathbf{n}}_t^{\beta} = \mathbf{n}_t^{\beta}$ is

$$\widetilde{\mathbf{n}}_u^{\beta} = \mathbf{n}_t^{\beta}\ \mathbf{U}_{\beta}e^{\mathbf{A}_{\beta}(u-t)}\mathbf{U}_{\beta}^{-1}$$

where $[\mathbf{A}_{\beta}]_{jj}$ are the eigenvalues of $\overline{\mathbf{\Lambda}}_{\beta}$ and $\mathbf{U}_{\beta}$ are associated column eigenvectors.[14] We then find

$$E\left[n_u^{\beta,i}|\mathcal{F}_t\right] = \sum_{\ell=1}^{N} n_t^{\beta,\ell}\sum_{j=1}^{N}[\mathbf{U}_{\beta}]_{\ell j}\left[\mathbf{U}_{\beta}^{-1}\right]_{ji}e^{[\mathbf{A}_{\beta}]_{jj}(u-t)}$$

This concludes the proof of the lemma.■

Proof of proposition 4: We shall need the following technical conditions:
Condition A1: For all $i = 1, .., n$

$$E\left[\int_0^{\infty} e^{-\phi\tau}|D_{\tau}^{1-\gamma_i}|d\tau\right] < \infty$$

Since the good is perishable, it is always suboptimal to consume less than D_t and consuming more is not feasible. Hence, we can impose the market clearing condition that $c_t = D_t$ for all $t > 0$. Usual arguments imply that we can use the marginal utility of consumption to discount future consumption. Hence, the price of an asset must satisfy

$$\begin{aligned} P_t &= E_t\left[\int_t^{\infty}\frac{U_c[c_s, s, \boldsymbol{\pi}_s]}{U_c[c_t, t, \boldsymbol{\pi}_t]}D_s ds\right] \\ &= \frac{1}{U_c[c_t, t, \boldsymbol{\pi}_t]}E_t\left[\int_t^{\infty} e^{-\phi s}\sum_i k_i c_s^{-\gamma_i}\pi_s^i D_s ds\right] \\ &= \frac{1}{U_c[c_t, t, \boldsymbol{\pi}_t]}E_t\left[\int_t^{\infty} e^{-\phi s}\sum_i k_i c_s^{1-\gamma_i}\pi_s^i ds\right] \end{aligned}$$

Noticing that for all i and s, $c_s^{1-\gamma_i}\pi_s^i \le c_s^{1-\gamma_i}$, invoking condition A1 and Fubini's theorem, we can use the result in Lemma 2 to obtain the value of P_t

$$P_t = \frac{1}{U_c[c_t, t, \boldsymbol{\pi}_t]}E_t\left[\int_t^{\infty} e^{-\phi s}\sum_i k_i c_s^{1-\gamma_i}\pi_s^i ds\right]$$

[14] For convenience, I provide here the usual argument for the solution of the a system of ODEs. I drop the argument β for simplicity. Given the system $d\widetilde{\mathbf{n}}/dt = \widetilde{\mathbf{n}}\cdot\overline{\mathbf{\Lambda}}$, let $\overline{\mathbf{\Lambda}} = \mathbf{UAU}^{-1}$ which implies $\mathbf{U}^{-1}\overline{\mathbf{\Lambda}}\mathbf{U} = \mathbf{A}$ and define $\mathbf{y} = \widetilde{\mathbf{n}}\mathbf{U}$. Then $d\mathbf{y}/dt = d\widetilde{\mathbf{n}}/dt\cdot\mathbf{U} = \widetilde{\mathbf{n}}\cdot\overline{\mathbf{\Lambda}}\mathbf{U} = \mathbf{y}\mathbf{U}^{-1}\overline{\mathbf{\Lambda}}\mathbf{U} = \mathbf{yA}$. Hence, the solution is $\mathbf{y}(u) = \mathbf{c}e^{\mathbf{A}u}$ which implies $\widetilde{\mathbf{n}}(u) = \mathbf{y}(u)\mathbf{U}^{-1} = \mathbf{c}e^{\mathbf{A}u}\mathbf{U}^{-1}$. The initial conditions $\widetilde{\mathbf{n}}(t) = \mathbf{n}(t) = \mathbf{c}e^{\mathbf{A}t}\mathbf{U}^{-1}$ implies $\mathbf{c} = \mathbf{n}(t)\mathbf{U}e^{-\mathbf{A}t}$ which then yields $\widetilde{\mathbf{n}}(u) = \mathbf{n}(t)\mathbf{U}e^{-\mathbf{A}t}e^{\mathbf{A}u}\mathbf{U}^{-1} = \mathbf{n}(t)\mathbf{U}e^{\mathbf{A}(u-t)}\mathbf{U}^{-1}$.

$$
\begin{aligned}
&= \frac{1}{e^{-\phi t}\sum_{i=1}^{n} k_i \pi_t^i c_t^{-\gamma_i}} \left\{ \int_t^{\infty} e^{-\phi s} \sum_i k_i E_t \left[c_s^{1-\gamma_i} \pi_s^i \right] ds \right\} \\
&= \frac{1}{e^{-\phi t}\sum_{i=1}^{n} k_i \pi_t^i c_t^{-\gamma_i}} \left\{ \int_t^{\infty} e^{-\phi s} \sum_i k_i \widetilde{n}_s^{\beta_i,i} ds \right\}
\end{aligned}
$$

where $\beta_i = 1 - \gamma_i$. I now apply the result of Lemma 2 and substitute for each

$$
\widetilde{n}_s^{\beta_i,i} = \sum_{\ell=1}^{N} n_t^{\beta_i,\ell} \sum_{j=1}^{N} \left[\mathbf{U}_{\beta_i} \right]_{\ell j} \left[\mathbf{U}_{\beta_i}^{-1} \right]_{ji} e^{\left[\mathbf{A}_{b_i} \right]_{jj}(s-t)}
$$

where $\mathbf{U}_{\beta_i}$ and $\mathbf{A}_{\beta_i}$ are the eigenvectors and eigenvalues of the matrix $\overline{\mathbf{\Lambda}}_{\beta_i} = \mathbf{\Lambda} + (1-\gamma_i)\, diag\left(g^1, ..., g^n\right) + \frac{1}{2}(1-\gamma_i)^2 \boldsymbol{\sigma\sigma}'\mathbf{I}$, where for convenience I set $\beta_i = 1 - \gamma_i$. This yields

$$
\widetilde{n}_s^{\beta_i,i} = \sum_{\ell=1}^{N} n_t^{\beta_i,\ell} \sum_{j=1}^{N} \left[\mathbf{U}_{\beta_i} \right]_{\ell j} \left[\mathbf{U}_{\beta_i}^{-1} \right]_{ji}^{-1} e^{\left[\mathbf{A}_{\beta_i} \right]_{jj}(s-t)}
$$

and hence (all steps provided for convenience):

$$
\begin{aligned}
P_t &= \frac{1}{e^{-\phi t}\sum_{i=1}^{n} k_i \pi_t^i c_t^{-\gamma_i}} \left\{ \int_t^{\infty} e^{-\phi s} \sum_{i=1}^{N} k_i \sum_{\ell=1}^{N} n_t^{\beta_i,\ell} \sum_{j=1}^{N} \left[\mathbf{U}_{\beta_i} \right]_{\ell j} \left[\mathbf{U}_{\beta_i}^{-1} \right]_{ji} e^{\left[\mathbf{A}_{\beta_i} \right]_{jj}(s-t)} ds \right\} \\
&= \frac{1}{\sum_{i=1}^{n} k_i \pi_t^i c_t^{-\gamma_i}} \left\{ \int_t^{\infty} \sum_{i=1}^{N} k_i \sum_{\ell=1}^{N} n_t^{\beta_i,\ell} \sum_{j=1}^{N} \left[\mathbf{U}_{\beta_i} \right]_{\ell j} \left[\mathbf{U}_{\beta_i}^{-1} \right]_{ji} e^{\left(\left[\mathbf{A}_{\beta_i} \right]_{jj} - \phi \right)(s-t)} ds \right\} \\
&= \frac{1}{\sum_{i=1}^{n} k_i \pi_t^i c_t^{-\gamma_i}} \left\{ \sum_{i=1}^{N} k_i \sum_{\ell=1}^{N} n_t^{\beta_i,\ell} \sum_{j=1}^{N} \left[\mathbf{U}_{\beta_i} \right]_{\ell j} \left[\mathbf{U}_{\beta_i}^{-1} \right]_{ji} \int_t^{\infty} e^{\left(\left[\mathbf{A}_{\beta_i} \right]_{jj} - \phi \right)(s-t)} ds \right\} \\
&= \frac{1}{\sum_{i=1}^{n} k_i \pi_t^i c_t^{-\gamma_i}} \left\{ \sum_{i=1}^{N} \sum_{\ell=1}^{N} k_i n_t^{\beta_i,\ell} \sum_{j=1}^{N} \left[\mathbf{U}_{\beta_i} \right]_{\ell j} \left[\mathbf{U}_{\beta_i}^{-1} \right]_{ji} \frac{1}{\phi - \left[\mathbf{A}_{\beta_i} \right]_{jj}} \right\} \\
&= \frac{1}{\sum_{i=1}^{n} k_i \pi_t^i c_t^{-\gamma_i}} \left\{ \sum_{i=1}^{N} \sum_{\ell=1}^{N} k_i c_t^{1-\gamma_i} \pi_t^{\ell} \sum_{j=1}^{N} \left[\mathbf{U}_{\beta_i} \right]_{\ell j} \left[\mathbf{U}_{\beta_i}^{-1} \right]_{ji} \frac{1}{\phi - \left[\mathbf{A}_{\beta_i} \right]_{jj}} \right\} \\
&= c_t \frac{\sum_{\ell=1}^{N} \pi_t^{\ell} c_t^{-\gamma_\ell} k_\ell}{\sum_{i=1}^{n} k_i \pi_t^i c_t^{-\gamma_i}} \left\{ \sum_{i=1}^{N} \frac{k_i}{k_\ell} c_t^{\gamma_\ell - \gamma_i} \sum_{j=1}^{N} \left[\mathbf{U}_{\beta_i} \right]_{\ell j} \left[\mathbf{U}_{\beta_i}^{-1} \right]_{ji} \frac{1}{\phi - \left[\mathbf{A}_{\beta_i} \right]_{jj}} \right\} \\
&= c_t \sum_{\ell=1}^{N} \overline{\pi}_t^{\ell}(c_t)\, B_\ell(c_t)
\end{aligned}
$$

where

$$
\overline{\pi}_t^{\ell}(c_t) = \frac{k_\ell \pi_t^{\ell} c_t^{-\gamma_\ell}}{\sum_{i=1}^{n} k_i \pi_t^i c_t^{-\gamma_i}}
$$

$$B_\ell(c_t) = \sum_{i=1}^{N} \frac{k_i}{k_\ell} c_t^{\gamma_\ell - \gamma_i} \sum_{j=1}^{N} \left[\mathbf{U}_{\beta_i}\right]_{\ell j} \left[\mathbf{U}_{\beta_i}^{-1}\right]_{ji} \frac{1}{\phi - \left[\mathbf{A}_{\beta_i}\right]_{jj}}$$

Then we can rewrite

$$P_t = D_t \sum_{\ell=1}^{n} \overline{\pi}_t^\ell (c_t) B_\ell(c_t)$$

I finally show that

$$\sum_{j=1}^{N} \left[\mathbf{U}_{\beta_i}\right]_{\ell j} \left[\mathbf{U}_{\beta_i}^{-1}\right]_{ji} \frac{1}{\phi - \left[\mathbf{A}_{\beta_i}\right]_{jj}} = e_\ell \left(\phi \mathbf{I} - \overline{\mathbf{\Lambda}}_{\beta_i}\right)^{-1} e_i$$

so that

$$B_\ell(c_t) = \sum_{i=1}^{N} \frac{k_i}{k_\ell} c_t^{\gamma_\ell - \gamma_i} e_\ell \left(\phi \mathbf{I} - \overline{\mathbf{\Lambda}}_{\beta_i}\right)^{-1} e_i$$

In fact, we know that $\left(\phi \mathbf{I} - \overline{\mathbf{\Lambda}}_{\beta_i}\right) = \mathbf{U}_{\beta_i} \left(\phi \mathbf{I} - \mathbf{A}_{\beta_i}\right) \mathbf{U}_{\beta_i}^{-1}$which implies

$$\left(\phi \mathbf{I} - \overline{\mathbf{\Lambda}}_{\beta_i}\right)^{-1} = \mathbf{U}_{\beta_i} \left(\phi \mathbf{I} - \mathbf{A}_{\beta_i}\right)^{-1} \mathbf{U}_{\beta_i}^{-1}$$

Hence

$$\begin{aligned} e_\ell' \left(\phi \mathbf{I} - \overline{\mathbf{\Lambda}}_{\beta_i}\right)^{-1} e_i &= e_\ell \mathbf{U}_{\beta_i} \left(\phi \mathbf{I} - \mathbf{A}_{\beta_i}\right)^{-1} \mathbf{U}_{\beta_i}^{-1} e_i \\ &= \sum_{j=1}^{n} \frac{\left[\mathbf{U}_{\beta_i}\right]_{\ell j} \left[\mathbf{U}_{\beta_i}^{-1}\right]_{ji}}{\phi - \left[\mathbf{A}_{\beta_i}\right]_{jj}} \end{aligned}$$

(b) We know that the real rate of interest rate is given by $r_t = -E_t\left[\frac{d\mathcal{U}_t}{\mathcal{U}_t}\right]$ where $\mathcal{U}_t$ is the real pricing kernel given by $\mathcal{U}_t = \partial U / \partial c = e^{-\phi t} \sum_{i=1}^n k_i \pi_t^i c_t^{-\gamma_i} = e^{-\phi t} \sum_{i=1}^n k_i n_t^{\beta_i, i}$ where $n_t^{\beta_i, i}$ is defined in Lemma 2 and $\beta_i = -\gamma_i$. Using Ito's lemma and equation (46), we find that $d\mathcal{U}/\mathcal{U} = -\mu_{\mathcal{U}} dt + \boldsymbol{\sigma}_{\mathcal{U}} d\widetilde{\mathbf{W}}_t$ with

$$\begin{aligned} \mu_{\mathcal{U}} &= \phi + \frac{1}{\sum_{i=1}^n k_i \pi_t^i c_t^{-\gamma_i}} \left(\sum_{i=1}^{n} k_i \gamma_i c_t^{-\gamma_i} \pi_t^i g^i \right. \\ &\quad \left. - \frac{1}{2} \sum_{i=1}^{n} k_i \pi_t^i \gamma_i^2 c_t^{-\gamma_i} \boldsymbol{\sigma}\boldsymbol{\sigma}' - \sum_{i=1}^{n} k_i c_t^{-\gamma_i} \left[\boldsymbol{\pi}_t \mathbf{\Lambda}\right]_i \right) \\ \boldsymbol{\sigma}_{\mathcal{U}} &= -\frac{\sum_{i=1}^n k_i \pi_t^i c_t^{-\gamma_i} \gamma_i}{\sum_{j=1}^n k_j \pi_t^j c_t^{-\gamma_j}} \boldsymbol{\sigma} + \frac{\sum_{i=1}^n k_i c_t^{-\gamma_i} \pi_t^i \left(\mathbf{v}_i - \overline{\mathbf{v}}_t\right)'}{\sum_{j=1}^n k_j \pi_t^j c_t^{-\gamma_j}} \mathbf{D}' \end{aligned} \tag{48}$$

Hence $r_t = -E_t\left[\frac{d\mathcal{U}_t}{\mathcal{U}_t}\right] = \mu_{\mathcal{U}}$. By redefining variables we obtain expression (25). ■

Proof of Corollary 1: (a) is immediate and (b) stems from the following manipulation:

$$\begin{aligned} r_t &= \phi + \gamma \sum_{i=1}^{n} \overline{\pi}_t^i g^i - \frac{1}{2}\gamma^2 \boldsymbol{\sigma}\boldsymbol{\sigma}' - \sum_{j=1}^{n} \overline{\pi}_t^j C_j^* \\ &= \phi + \gamma \sum_{i=1}^{n} \pi_t^i g^i + \gamma \left(\sum_{i=1}^{n} \overline{\pi}_t^i g^i - \sum_{i=1}^{n} \pi_t^i g^i \right) - \frac{1}{2}\gamma^2 \boldsymbol{\sigma}\boldsymbol{\sigma}' - \sum_{j=1}^{n} \overline{\pi}_t^j C_j^* \\ &= r_t^B + \gamma V_t^{k,g} - \sum_{j=1}^{n} \overline{\pi}_t^j C_j^* \end{aligned}$$

because

$$\begin{aligned} V_t^{k,g} &= \frac{Cov_\pi(k,g)}{\boldsymbol{\pi}_t' \cdot \mathbf{k}} = \frac{\sum_{i=1}^{n} \pi_t^i k^i \left(g^i - \sum_{j=1}^{n} \pi_t^j g^j \right)}{\boldsymbol{\pi}_t' \cdot \mathbf{k}} \\ &= \sum_{i=1}^{n} \overline{\pi}_t^i g^i - \sum_{i=1}^{n} \pi_t^i g^i \end{aligned}$$

■

Proof of Proposition 4: The price of the asset in the constant relative risk aversion case is

$$P_t = D_t \frac{\sum_{j=1}^{n} \pi_t^j k_j B_j}{\sum_{i=1}^{n} k_i \pi_t^i}$$

In this proof only, let $X_t = \sum_i k_i \pi_t^i$. Hence,

$$dX_t = \sum_i k_i d\pi_t^i = \sum_i k_i \left[\boldsymbol{\pi}_t \boldsymbol{\Lambda}\right]_i dt + \sum_i k_i \pi_t^i \left(\mathbf{v}_i - \overline{\mathbf{v}}_t\right)' \mathbf{D}' d\widetilde{\mathbf{W}}_t$$

or

$$\begin{aligned} \frac{dX_t}{X_t} &= \frac{\sum_i k_i \left[\boldsymbol{\pi}_t \boldsymbol{\Lambda}\right]_i}{\sum_i k_i \pi_t^i} dt + \frac{\sum_i k_i \pi_t^i \left(\mathbf{v}_i - \overline{\mathbf{v}}_t\right)'}{\sum_i k_i \pi_t^i} \mathbf{D}' d\widetilde{\mathbf{W}} \\ &= \mu_{X,t} dt + \boldsymbol{\sigma}_{X,t} d\widetilde{\mathbf{W}}_t \end{aligned} \tag{49}$$

with

$$\boldsymbol{\sigma}_{X,t} = \frac{\sum_i k_i \pi_t^i \left(\mathbf{v}_i - \overline{\mathbf{v}}_t\right)'}{\sum_i k_i \pi_t^i} \mathbf{D}'$$

It is convenient to define $\widetilde{P}_t = D_t \sum_{i=1}^{n} \pi_t^i k_i B_i$. From Ito's lemma we obtain

$$\begin{aligned} d\widetilde{P}_t &= D_t \sum_i k_i B_i d\pi_t^i + \sum_{i=1}^{n} \pi_t^i k_i B_i dD_t + \sum_{i=1}^{n} k_i B_i d\pi_t^i dD_t \\ &= D_t \sum_i k_i B_i \left[\boldsymbol{\pi}_t \boldsymbol{\Lambda}\right]_i dt + D_t \sum_i k_i B_i \pi_t^i \left(\mathbf{v}_i - \overline{\mathbf{v}}_t\right)' \mathbf{D}' d\widetilde{\mathbf{W}}_t \end{aligned}$$

$$+\sum_{i=1}^{n}\pi_t^i k_i B_i D_t \mu_{D,t} dt + \sum_{i=1}^{n}\pi_t^i k_i B_i \boldsymbol{\sigma} d\widetilde{\mathbf{W}}_t$$
$$+\sum_{i=1}^{n} k_i B_i D_t \pi_t^i \left(\mathbf{v}_i - \overline{\mathbf{v}}_t\right)' \mathbf{D}' \boldsymbol{\sigma}' dt$$

Notice now again that by definition $\mathbf{D}' = \left(\left(\boldsymbol{\sigma}', \boldsymbol{\sigma}_s'\right)\right)^{-1}$ which implies $\mathbf{D}'\boldsymbol{\sigma}' = \left(\left(\boldsymbol{\sigma}', \boldsymbol{\sigma}_s'\right)\right)^{-1}\left(\boldsymbol{\sigma}', \boldsymbol{\sigma}_s'\right)(1,0)' = (1,0)$ Hence, we have

$$\left(\mathbf{v}_i - \overline{\mathbf{v}}_t\right)' \mathbf{D}' \boldsymbol{\sigma}' = \left(g^i - \sum_{j=1}^{n}\pi_t^j g^j\right)$$

which yields

$$d\widetilde{P}_t = \widetilde{P}_t \widetilde{\mu}_{P,t} dt + \widetilde{P}_t \widetilde{\boldsymbol{\sigma}}_{P,t} d\widetilde{\mathbf{W}}_t$$

where

$$\begin{aligned}\widetilde{\mu}_{P,t} &= \mu_{D,t} + \frac{\sum_{i=1}^{n} k_i B_i \left[\boldsymbol{\pi}_t \boldsymbol{\Lambda}\right]_i}{\sum_{i=1}^{n}\pi_i k_i B_i} + V_t^B \\ \widetilde{\boldsymbol{\sigma}}_{P,t} &= \boldsymbol{\sigma} + \frac{\sum_i k_i B_i \pi_t^i \left(\mathbf{v}_i - \overline{\mathbf{v}}_t\right)'}{\sum_i k_i B_i \pi_t^i}\mathbf{D}'\end{aligned}$$

Notice that since $P_t = \widetilde{P}_t / X_t$ we also find

$$\frac{dP_t}{P_t} = \left(\widetilde{\mu}_{P,t} - \mu_{X,t} + \boldsymbol{\sigma}_{X,t}\boldsymbol{\sigma}_{X,t}' - \widetilde{\boldsymbol{\sigma}}_{P,t}\boldsymbol{\sigma}_{X,t}'\right) dt + \left(\widetilde{\boldsymbol{\sigma}}_{P,t} - \boldsymbol{\sigma}_{X,t}\right) d\widetilde{\mathbf{W}}_t$$

and hence

$$\begin{aligned}\boldsymbol{\sigma}_{P,t} &= \boldsymbol{\sigma} + \left(\frac{\sum_i k_i B_i \pi_t^i \left(\mathbf{v}_i - \overline{\mathbf{v}}_t\right)'}{\sum_i k_i B_i \pi_t^i} - \frac{\sum_i k_i \pi_t^i \left(\mathbf{v}_i - \overline{\mathbf{v}}_t\right)'}{\sum_i k_i \pi_t^i}\right)\mathbf{D}' \\ &= \boldsymbol{\sigma} + \left(V_t^{kB,g} - V_t^{k,g}, V_t^{kB,\theta} - V_t^{k,\theta}\right)\mathbf{D}' \\ &= \boldsymbol{\sigma} + \left(\Delta V_t^g, \Delta V_t^\theta\right)\mathbf{D}'\end{aligned}$$

The equilibrium condition requires that the excess return

$$dR_t = \frac{dP_t + D_t dt}{P_t} - r_t dt$$

is such that

$$E_t\left[dR_t\right] = -Cov\left(dR_t, \frac{d\mathcal{U}}{\mathcal{U}}\right) = \boldsymbol{\sigma}_P \boldsymbol{\sigma}_{\mathcal{U}}'$$

Using (48) specialized to this case

$$\begin{aligned}\boldsymbol{\sigma}_{\mathcal{U}} &= -\gamma\boldsymbol{\sigma} + \left(V_t^{k,g}, V_t^{k.\theta}\right)\mathbf{D}' \\ &= -\gamma\boldsymbol{\sigma} + \boldsymbol{\sigma}_{X,t}\end{aligned}$$

we finally find

$$
\begin{aligned}
E_t[dR_t] &= -\left(\widetilde{\boldsymbol{\sigma}}_{P,t}-\boldsymbol{\sigma}_{X,t}\right)\boldsymbol{\sigma}_{\mathcal{U},t}dt \\
&= -\left(\boldsymbol{\sigma}+\left(\Delta V_t^g,\Delta V_t^\theta\right)\mathbf{D}'\right)\left(-\gamma\boldsymbol{\sigma}+\left(V_t^{k,g},V_t^{k.\theta}\right)\mathbf{D}'\right)' \\
&= \gamma\left(\boldsymbol{\sigma}\boldsymbol{\sigma}'+\left(\Delta V_t^g,\Delta V_t^\theta\right)\mathbf{D}'\boldsymbol{\sigma}'\right) \\
&\quad -\boldsymbol{\sigma}\mathbf{D}\left(V_t^{k,g},V_t^{k.\theta}\right)'-\left(\Delta V_t^g,\Delta V_t^\theta\right)\mathbf{D}'\mathbf{D}\left(V_t^{k,g},V_t^{k.\theta}\right)' \\
&= \gamma\left(\boldsymbol{\sigma}\boldsymbol{\sigma}'+\Delta V_t^g\right)-V_t^{k,g}-\left(\Delta V_t^g,\Delta V_t^\theta\right)\mathbf{D}'\mathbf{D}\left(V_t^{k,g},V_t^{k.\theta}\right)'
\end{aligned}
$$

which yields (36) ■

Proof of Lemma 3: Consider $x=g$. For $x=\theta$ the proof is analogous. By definition

$$
\begin{aligned}
\Delta V_t^g &= V_t^{kB,g}-V_t^{k,g}=\frac{E_t\left[k\left(\theta\right)B(\theta,g)g\right]}{E_t\left[k\left(\theta\right)B(\theta,g)\right]}-\frac{E_t\left[k\left(\theta\right)g\right]}{E_t\left[k\left(\theta\right)\right]} \\
&= E_t^{kB}[g]-E_t^k[g]
\end{aligned}
$$

where the expectations $E_t^{kB}[.]$ and $E_t^k[.]$ are taken with respect the two new densities

$$
p_t^{kB}(\theta,g)=\frac{p_t(\theta,g)k(\theta)B(\theta,g)}{E_t\left[k\left(\theta\right)B(\theta,g)\right]}\text{ and }p_t^k(\theta,g)=\frac{p_t(\theta,g)k(\theta)}{E_t\left[k\left(\theta\right)\right]}
$$

Then, the point wise ratio is

$$
\frac{p_t^{kB}(\theta,g)}{p_t^k(\theta,g)}=B\left(\theta,g\right)\frac{E_t\left[k\left(\theta\right)\right]}{E_t\left[k\left(\theta\right)B(\theta,g)\right]}
$$

Since $B(\theta,g)$ is non decreasing in both arguments, also the ratio $\frac{p_t^{kB}(\theta,g)}{p_t^k(\theta,g)}$ is non-decreasing in both θ and g. Because both densities must integrate to 1, we have that for every θ, $p_t^{kB}(\theta,g)\leq p_t^k(\theta,g)$ for $g\leq g^*$ and $p_t^{kB}(\theta,g)\geq p_t^k(\theta,g)$ for $g>g^*$ for some cutoff value g^*. This immediately yields the result. Similarly for $x=\theta$. ■

Proof of Proposition 6: The interest rate can be written

$$
\begin{aligned}
r_t &= \phi-\frac{1}{2}\gamma^2\boldsymbol{\sigma}\boldsymbol{\sigma}'+\frac{\sum_{i=1}^n\pi_t^i\gamma k_ig^i-\sum_{i=1}^n\sum_j k_i\pi_t^j\lambda_{ji}}{\sum_{i=1}^n k_i\pi_t^i} \\
&= \phi-\frac{1}{2}\gamma^2\boldsymbol{\sigma}\boldsymbol{\sigma}'+\frac{\sum_{i=1}^n\pi_t^i\gamma k_ig^i-\sum_{i=1}^n\pi_t^ik_i\sum_j k_j/k_i\lambda_{ij}}{\sum_{i=1}^n k_i\pi_t^i} \\
&= \phi-\frac{1}{2}\gamma^2\boldsymbol{\sigma}\boldsymbol{\sigma}'+\frac{\widetilde{r}_t}{X_t} \\
&= \widetilde{\phi}+\frac{\widetilde{r}_t}{X_t}
\end{aligned}
$$

where $\widetilde{r}_t=\sum_{i=1}^n\pi_t^iC_i^r$, $X_t=\sum_{i=1}^n k_i\pi_t^i$ and $C_i^r=\gamma k_ig_i-k_iC_i^*$. By Ito's lemma:

$$
\begin{aligned}
d\widetilde{r}_t &= \sum_i C_i^r\left[\boldsymbol{\pi}_t\boldsymbol{\Lambda}\right]_i dt+\sum_i C_i^r\boldsymbol{\pi}_t^i\left(\mathbf{v}_i-\overline{\mathbf{v}}_t\right)'\mathbf{D}'d\widetilde{\mathbf{W}}_t \\
&= \widetilde{\mu}_r dt+\widetilde{\boldsymbol{\sigma}}_{r,t}d\widetilde{\mathbf{W}}_t
\end{aligned}
$$

Hence, using (49)

$$\begin{aligned} dr_t &= \frac{\widetilde{r}_t}{X_t}\left(\frac{d\widetilde{r}_t}{\widetilde{r}_t}\right) - \frac{\widetilde{r}_t}{X_t}\frac{dX_t}{X_t} + \frac{\widetilde{r}_t}{X_t}\left(\frac{dX_t}{X_t}\right)^2 - \frac{\widetilde{r}_t}{X_t}\frac{d\widetilde{r}_t}{\widetilde{r}_t}\frac{dX_t}{X_t} \\ &= \left(r_t - \widetilde{\phi}\right)\left(\widetilde{\mu}^*_{r,t}dt + \widetilde{\boldsymbol{\sigma}}^*_{r,t}d\widetilde{\mathbf{W}}_t - \mu_{X,t}dt - \boldsymbol{\sigma}_{X,t}d\widetilde{\mathbf{W}}_t + \boldsymbol{\sigma}_{X,t}\boldsymbol{\sigma}'_{X,t}dt - \widetilde{\boldsymbol{\sigma}}^*_{r,t}\boldsymbol{\sigma}'_{X,t}dt\right) \end{aligned}$$

where

$$\begin{aligned} \widetilde{\mu}^*_{r,t} &= \frac{\sum_i C_i^r \left[\boldsymbol{\pi}_t\boldsymbol{\Lambda}\right]_i}{\sum_{i=1}^n \pi_t^i C_i^r} \text{ and} \\ \widetilde{\boldsymbol{\sigma}}^*_{r,t} &= \frac{\sum_i C_i^r \pi_t^i \left(\mathbf{v}_i - \overline{\mathbf{v}}_t\right)'}{\sum_{i=1}^n \pi_i D_i^r}\mathbf{D}' = \left[V_t^{r,g}, V_t^{r,\theta}\right]\mathbf{D}' \end{aligned}$$

where for $x = g, \theta$,

$$V_t^{r,x} = \frac{\sum_i C_i^r \pi_i \left(x^i - \sum_j \pi_j x^i\right)'}{\sum_{i=1}^n \pi_i C_i^r}$$

Hence, the diffusion of the interest rate process is

$$\begin{aligned} \boldsymbol{\sigma}_{r,t} &= \left(r_t - \widetilde{\phi}\right)\left(\left[V_t^{r,g}, V_t^{r,\theta}\right]\mathbf{D}' - \boldsymbol{\sigma}_{X,t}\right) \\ &= \left(r_t - \widetilde{\phi}\right)\left[V_t^{r,g} - V_t^{k,g}, V_t^{r,\theta} - V_t^{k,\theta}\right]\mathbf{D}' \end{aligned}$$

■

Proof of Corollary 5: Define $k_i = k(g_i)$ and $k = k_1, .., k_n$. From the formula for B_k in (30) we find that

$$B_j = \frac{1}{k_j}\sum_{i=1}^n e_j' \left(\phi\mathbf{I} - \overline{\boldsymbol{\Lambda}}(1-\gamma)\right)^{-1} e_i \; k_i$$

implies

$$B_j k_j = e_j' \left(\phi\mathbf{I} - \overline{\boldsymbol{\Lambda}}(1-\gamma)\right)^{-1} \cdot \mathbf{k}$$

In vector notation

$$diag\left(\mathbf{k}\right) \times \mathbf{B} = \left(\phi\mathbf{I} - \overline{\boldsymbol{\Lambda}}(1-\gamma)\right)^{-1}\mathbf{k}$$

which yields

$$\left(\phi\mathbf{I} - \overline{\boldsymbol{\Lambda}}(1-\gamma)\right) diag\left(\mathbf{k}\right) \times \mathbf{B} = \mathbf{k}$$

Recall now $\overline{\boldsymbol{\Lambda}}(1-\gamma) = \boldsymbol{\Lambda} + diag\left(\widehat{g}_1, ..., \widehat{g}_n\right)$ with $\widehat{g}_j = (1-\gamma)g_j + \frac{1}{2}(1-\gamma)^2\boldsymbol{\sigma}\boldsymbol{\sigma}'$. Hence, we have

$$\phi \; diag\left(\mathbf{k}\right) \times \mathbf{B} - \boldsymbol{\Lambda} \; \times diag\left(\mathbf{k}\right) \times \mathbf{B} - diag\left(\widehat{\mathbf{g}}\right) \times diag\left(\mathbf{k}\right) \times \mathbf{B} = \mathbf{k}$$

or defining $\widehat{g}_i = (1-\gamma)g^i + \frac{1}{2}(1-\gamma)^2\boldsymbol{\sigma}\boldsymbol{\sigma}'$:

$$\phi k_i B_i - \widehat{g}_i k_i B_i = k_i + \sum_{j=1}^n \lambda_{ij} k_j B_j$$

Using the assumption on λ_{ij} we obtain

$$\begin{aligned}
(\phi - \widehat{g}_i)\, k_i B_i &= k_i + \sum_{j \neq i} \alpha_i f_j k_j B_j + (\alpha_i f_i - \alpha_i)\, k_i B_i \\
&= k_i + \sum_{j=1}^{n} \alpha_i f_j k_j B_j - \alpha_i k_i B_i
\end{aligned}$$

Taking the last term to the right hand side, we have

$$(\phi - \widehat{g}_i + \alpha_i)\, k_i B_i = k_i + \alpha_i \sum_{j=1}^{n} f_j k_j B_j$$

or

$$k_i B_i = \frac{k_i}{\phi - \widehat{g}_i + \alpha_i} + \frac{p_i}{\phi - \widehat{g}_i + \alpha_i} \left(\sum_{j=1}^{n} f_j k_j B_j \right) \tag{50}$$

Multiply each side by f_i and sum across i to obtain

$$\begin{aligned}
\sum_{i=1}^{n} f_i k_i B_i &= \sum_{i=1}^{n} \frac{f_i k_i}{\phi - \widehat{g}_i + \alpha_i} + \sum_{i=1}^{n} \frac{f_i p_i}{\phi - \widehat{g}_i + \alpha_i} \left(\sum_{j=1}^{n} f_j k_j B_j \right) \\
&= H_1 + H_2 \left(\sum_{j=1}^{n} f_j k_j B_j \right)
\end{aligned}$$

where H_1 and H_2 are defined accordingly. Hence, assuming $1 - H_2 \neq 0$ we obtain

$$\sum_{i=1}^{n} f_i k_i B_i = \frac{H_1}{1 - H_2}$$

We can plug this back into (50) to obtain

$$\begin{aligned}
B_i &= \frac{1}{\phi + \alpha_i - (1-\gamma)\, g_i - \frac{1}{2}(1-\gamma)^2 \boldsymbol{\sigma\sigma'}} \\
&\quad + \frac{\alpha_i}{k(\theta_i)\left(\phi + \alpha_i - (1-\gamma)\, g_i - \frac{1}{2}(1-\gamma)^2 \boldsymbol{\sigma\sigma'}\right)} \left(\frac{H_1}{1 - H_2} \right) \\
&= \frac{1 - H_2 + \frac{\alpha_i}{k(\theta_i)} H_1}{\left(\phi + \alpha_i - (1-\gamma)\, g_i - \frac{1}{2}(1-\gamma)^2 \boldsymbol{\sigma\sigma'}\right)(1 - H_2)}
\end{aligned}$$

■

Table 1: ML Estimates of Pure-Jump Model

Panel A: 1946-1999 - Restricted μ									
Parameter	α	σ	σ_J	μ	$\log(\mathcal{L})$	Skew	Kurt	RMSE	MAE
Estimate	0.0454	0.0067	0.0133	0.0049	736.989	-.7416	8.4437	.8732	.6125
t-stat.	1.368	9.267	2.056	-					
Panel B: 1946-1999 - Estimated μ									
Parameter	α	σ	σ_J	μ	$\log(\mathcal{L})$	Skew	Kurt	RMSE	MAE
Estimate	0.0449	0.0067	0.0126	-0.0006	737.790	-.6548	8.1636	.8716	.6142
t-stat.	1.381	9.696	2.604	-.125					
Panel C: 1952-1999 - Estimated μ									
Parameter	α	σ	σ_J	μ	$\log(\mathcal{L})$	Skew	Kurt	RMSE	MAE
Estimate	.1837	0.0056	.0031	.0058	698.4619	-0.6949	8.0518	0.6734	0.4694
t-stat.	3.0821	11.9364	4.4733	7.3240					

This table reports the Maximum Likelihood annaulized estimates of the Pure-Jump statistical models for θ_t : Specifically, $\Delta \log(c_{t+1}) = g_t + \sigma\, \varepsilon_{t+1}$ where g_t follows a pure jump model with

$$g_{t+1} = \begin{cases} g_t & \text{with prob.} \quad 1-\alpha \\ \xi_{t+1} & \text{with prob.} \quad \alpha \end{cases}$$

where $\xi_t \sim \mathcal{N}\left(\mu, \sigma_J^2\right)$. Estimates are obtained by discretizing the interval $[-0.4, 0.4]$ in $n = 200$ intervals and applying ML estimation methods for the implied approximated models as in Kitagawa (1987). Standard errors are (Newey-West) corrected for heteroskedasticity and autocorrelation. In Panel A the parameter μ has been fixed to the long-run average of consumption growth. The "Skew" and "Kurt" column refer to the skewness and kurtosis of the one-quarter ahead prediction errors. The RMSE and MAE report the one quarter ahead Root Mean Square Prediction Error and Mean Absolute Prediction Errors, respectively (numbers are multiplied by 100).

Table 2: ML Estimates of Autoregressive Mean Model

Panel A: 1946-1999

Parameter	a	σ	σ_g	μ	$\log(\mathcal{L})$	Skew	Kurt	RMSE	MAE
Estimate	0.8870	0.0074	0.0017	0.0005	732.952	-1.2329	9.1164	0.8050	0.5756
t-stat.	9.0923	6.6538	4.2778	.8739					

Panel B: 1952 -1999

Parameter	a	σ	σ_g	μ	$\log(\mathcal{L})$	Skew	Kurt	RMSE	MAE
Estimate	0.7986	0.0057	0.0018	0.0012	697.4137	-0.8103	8.1164	0.6691	0.4734
t-stat.	10.1362	11.6510	3.8370	2.3537					

This table reports the Maximum Likelihood annaulized estimates of the Autoregressive Mean statistical models for θ_t. Specifically, $\Delta \log(c_{t+1}) = g_t + \sigma\, \varepsilon_{t+1}$ where g_t follows the autoregressive process

$$g_{t+1} = \mu + a g_t + \sigma_g \varepsilon_{g,t+1}$$

with $E[\varepsilon_t, \varepsilon_{g,t}] = 0$. Estimates are obtained by discretizing the interval $[-0.4, 0.4]$ in $n = 200$ intervals and applying ML estimation methods for the implied approximated models as in Kitagawa (1987). Standard errors are (Newey-West) corrected for heteroskedasticity and autocorrelation. The "Skew" and "Kurt" column refer to the skewness and kurtosis of the one-quarter ahead prediction errors. The RMSE and MAE report the one quarter ahead Root Mean Square Prediction Error and Mean Absolute Prediction Errors, respectively (numbers are multiplied by 100).

Table 3: Calibration 1946 - 1999

Panel A: Sample Data

			$E[dR](\%)$	$\sigma_R(\%)$	$r(\%)$	$\sigma_r(\%)$	mean P/D
			6.87	15.97	2.00^a	4.5125	29.4

Panel B: Pure Jump Process

γ	ρ	ϕ	$E[dR](\%)$	$\sigma_R(\%)$	$r(\%)$	$\sigma_r(\%)$	mean P/D
1.5	0	0.01	-0.0814	4.0412	3.8964	0.0059	61.5178
1.5	0	0.04	-0.0672	3.3386	6.8964	0.0059	21.2617
1.5	50	0.01	6.1691	21.2689	-0.6558	-0.1071	82.9530
1.5	50	0.04	5.5818	19.2440	2.3442	-0.1071	27.4109
1.5	100	0.01	26.9966	48.2116	-19.3584	-2.6469	176.8303
1.5	100	0.04	25.6397	45.7885	-16.3584	-2.6469	54.3688
2	0	0.01	-0.2426	9.0333	4.8529	0.0105	74.7092
2	0	0.04	-0.2071	7.7122	7.8529	0.0105	20.9598
2	50	0.01	4.9760	16.7673	0.1194	-0.0550	113.6168
2	50	0.04	4.5286	15.2598	3.1194	-0.0550	29.2300
2	100	0.01	24.9252	43.9851	-18.7643	-2.5761	308.1002
2	100	0.04	24.0627	42.4630	-15.7643	-2.5761	67.5799

Panel C: Ornstein-Uhlenbeck Process

γ	ρ	ϕ	$E[dR](\%)$	$\sigma_R(\%)$	$r(\%)$	$\sigma_r(\%)$	mean P/D
1.5	0	0.01	0.0096	0.4317	3.7488	0.0026	52.3274
1.5	0	0.04	0.0110	0.4935	6.7488	0.0026	20.9598
1.5	50	0.01	1.8435	12.4822	2.9970	0.0582	14.3793
1.5	50	0.04	1.7480	11.8355	5.9970	0.0582	60.9910
1.5	100	0.01	6.7036	24.5389	1.2985	0.0192	24.1991
1.5	100	0.04	6.3356	23.1918	4.2985	0.0192	16.3007
2	0	0.01	-0.0171	0.5782	4.6541	0.0047	105.6699
2	0	0.04	-0.0136	0.4590	7.6541	0.0047	33.0940
2	50	0.01	1.7456	11.2554	3.8094	0.0828	21.5318
2	50	0.04	1.6581	10.6911	6.8094	0.0828	410.2048
2	100	0.01	6.4814	23.0998	2.0181	0.0614	67.5840
2	100	0.04	6.1337	21.8608	5.0181	0.0614	40.9973

Panel A reports the ex-post mean excess stock returns, its volatility, and the volatility of the real-interest rate (assuming a constant risk premium) for the 1946-1999 period. Panel B and C report the moments implied by the pure jump model and the Ornstein-Uhlenbeck process,respectively, fitted in Table 1 and 2 for several utility parameters ρ, ϕ and γ. In both models, the posterior distribution used to compute the unconditional expected returns is $\pi \sim N\left(\mu_\pi, \overline{\sigma}_\pi^2\right)$ where μ_π is the long-term average of consumption growth and $\overline{\sigma}_\pi$ is the average Root Mean Square Error in the data. All the moments have been the computed using the formulas in the text.

a: The value of the real rate has not been estimated. A value approximately of 2% is considered appropriate in the literature.

Table 4: Calibration 1952 - 1999

Panel A: Sample Data

			$E[dR]$ (%)	σ_R (%)	r (%)	σ_r (%)	mean P/D
			6.5384	16.218	2.00^a	3.9496	30.8097

Panel B: Pure-Jump Model

γ	ρ	ϕ	$E[dR]$ (%)	σ_R (%)	r (%)	σ_r (%)	mean P/D
1.5	80	0.01	4.1693	19.5548	3.4919	0.1021	47.2880
1.5	80	0.04	4.0207	18.8579	6.4919	0.1021	19.7341
1.5	100	0.01	6.3805	24.3269	3.0492	0.1047	47.8249
1.5	100	0.04	6.1515	23.4537	6.0492	0.1047	19.9497
2	80	0.01	4.0805	18.6450	4.5278	0.1454	30.9654
2	80	0.04	3.9376	17.9923	7.5278	0.1454	16.1816
2	100	0.01	6.2557	23.3490	4.0574	0.1569	31.3222
2	100	0.04	6.0349	22.5249	7.0574	0.1569	16.3611

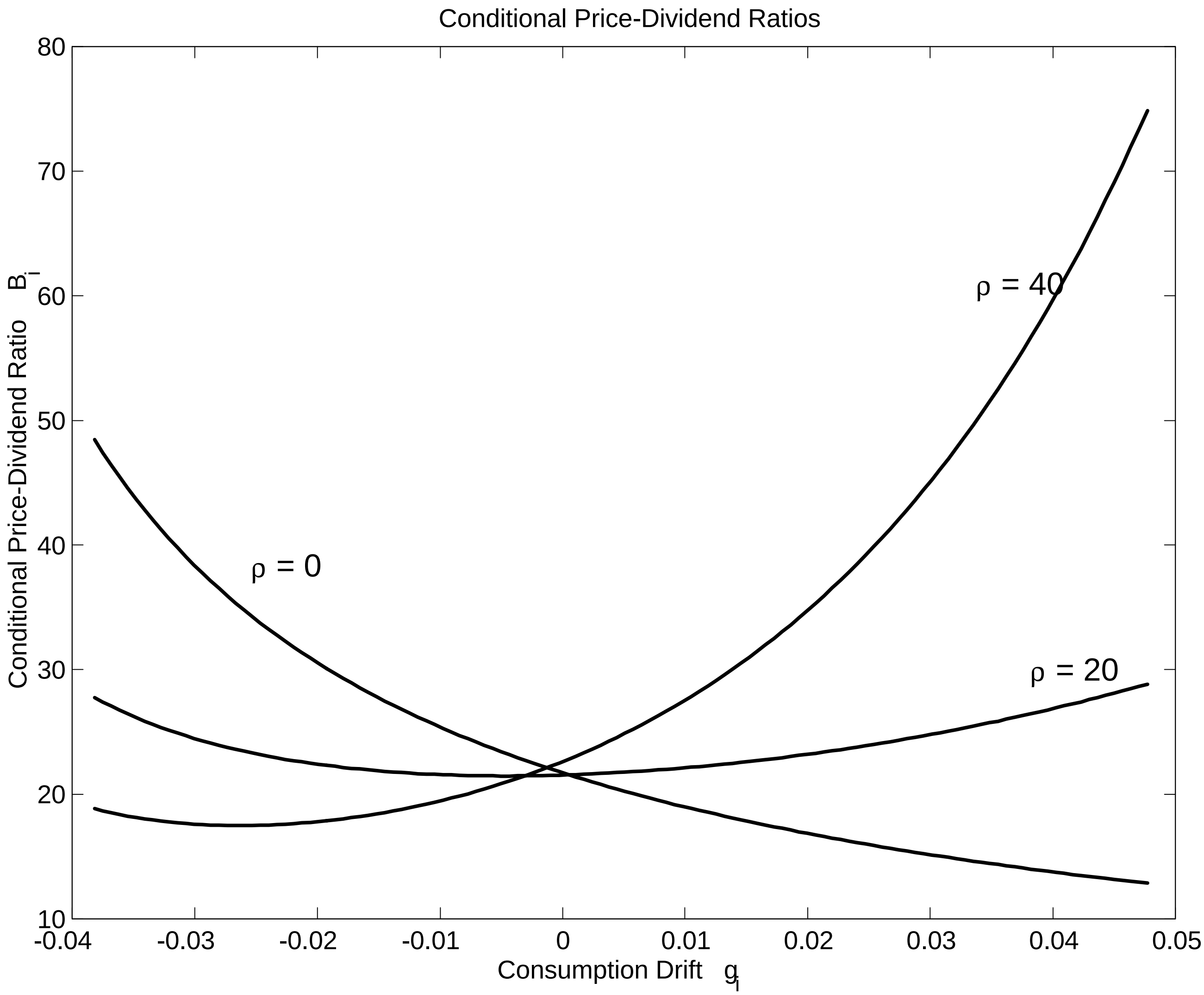
Panel C: Ornstein-Uhlenbeck Process

γ	ρ	ϕ	$E[dR]$ (%)	σ_R (%)	r (%)	σ_r (%)	mean P/D
1.5	80	0.01	4.0221	19.2598	3.5057	0.0979	47.6196
1.5	80	0.04	3.8922	18.6375	6.5057	0.0979	19.7468
1.5	100	0.01	6.1448	23.9321	3.1147	0.1031	48.0122
1.5	100	0.04	5.9440	23.1500	6.1147	0.1031	19.9036
2	80	0.01	3.9609	18.4618	4.5174	0.1382	31.2547
2	80	0.04	3.8349	17.8748	7.5174	0.1382	16.2283
2	100	0.01	6.0561	23.0736	4.0991	0.1517	31.5163
2	100	0.04	5.8612	22.3309	7.0991	0.1517	16.3592

Panel A reports the ex-post mean excess stock returns, its volatility, and the volatility of the real-interest rate (assuming a constant risk premium) for the 1952-1999 period. Panel B and C report the moments implied by the pure jump model and the Ornstein-Uhlenbeck process,respectively, fitted in Tables 1 and 2 for sevarl utility parameters ρ, ϕ and γ. In both models, the posterior distribution used to compute the unconditional expected returns is $\pi \sim \mathcal{N}\left(\mu_\pi, \overline{\sigma}_\pi^2\right)$ where μ is the long-term average of consumption growth and $\overline{\sigma}_\pi$ is the average Root Mean Square Error in the data. All the moments have been the computed using the formulas in the text.

a: The value of the real rate has not been estimated. A value approximately of 2% is considered appropriate in the literature.

Figure 1: Conditional Price-Dividend Ratios

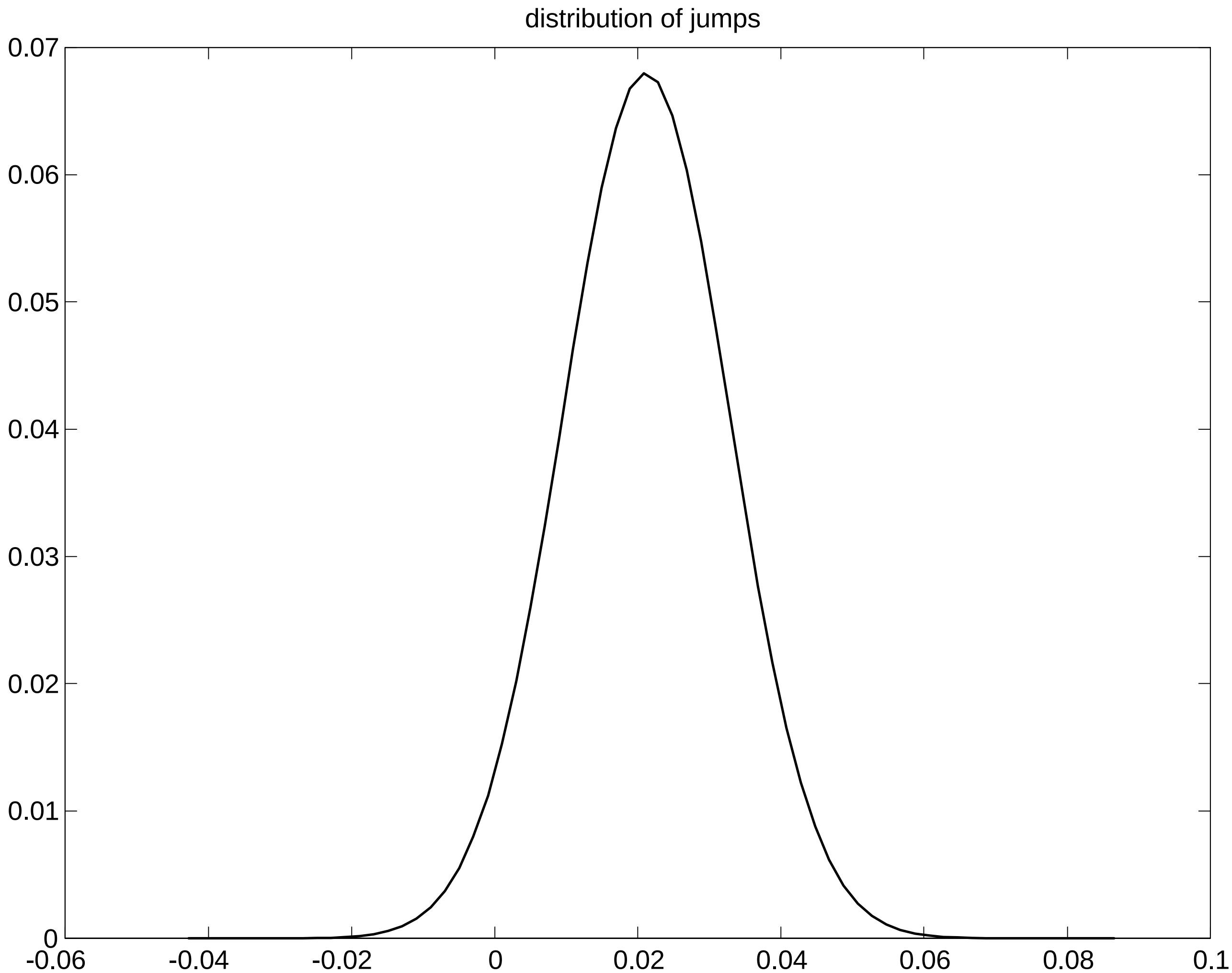


This figure plots the conditional price-dividend ratios obtained from the formula

$$B_i = \frac{1}{k\left(g_i\right)} \sum_{j=1}^{n} k\left(g_j\right) e_j \left(\phi I - \overline{\Lambda}\right) e_i'$$

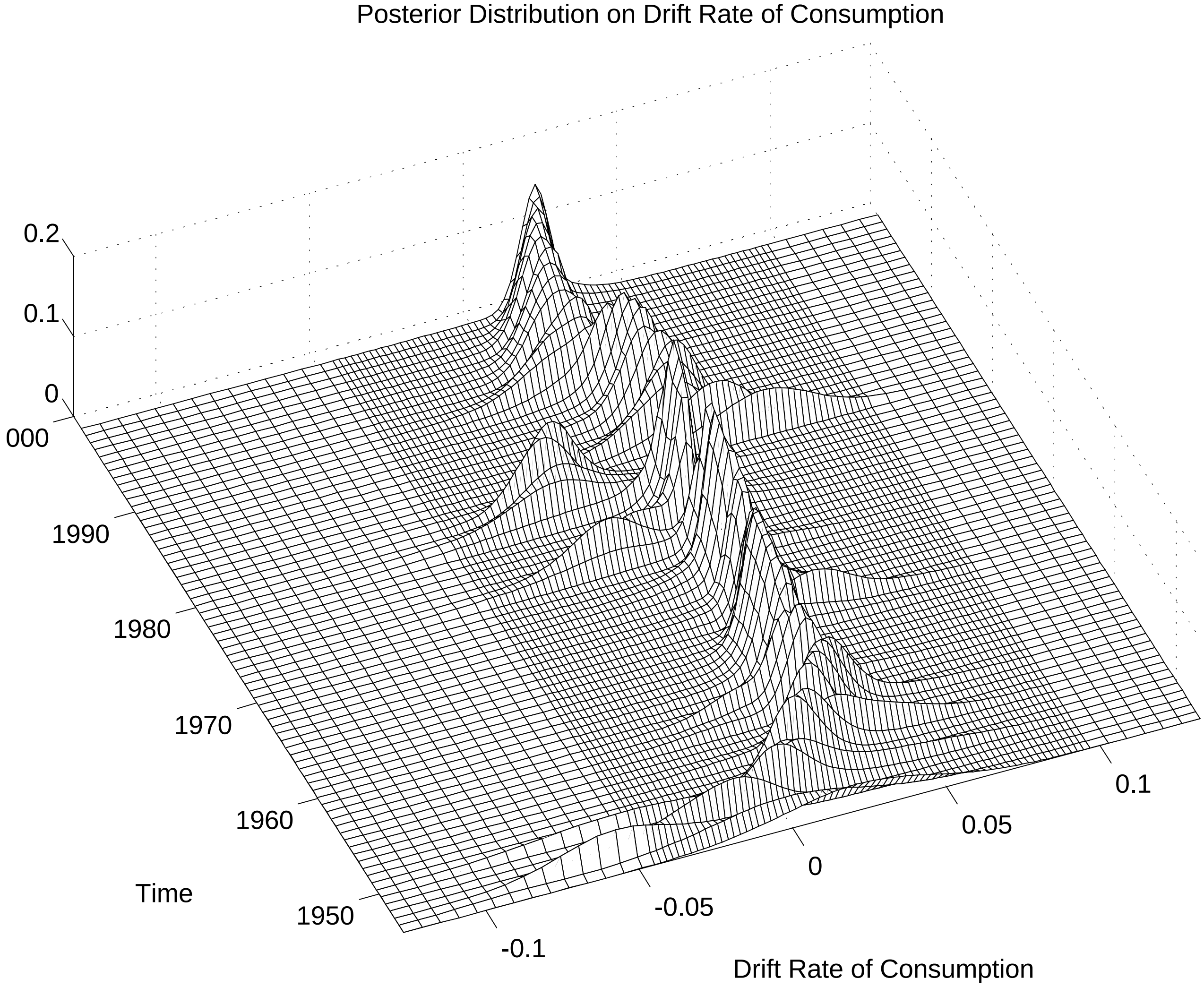
where $k\left(g_i\right) = \exp\left(-\rho\left(g_i - g_n\right)\right)$.

Figure 2: The Jump Distribution



The figure plots the estimated distribution of jumps in the drift rate of consumption $J_t \sim \mathcal{N}\left(\mu, \sigma_J^2\right)$ estimated from quarterly consumption from 1946-1999.

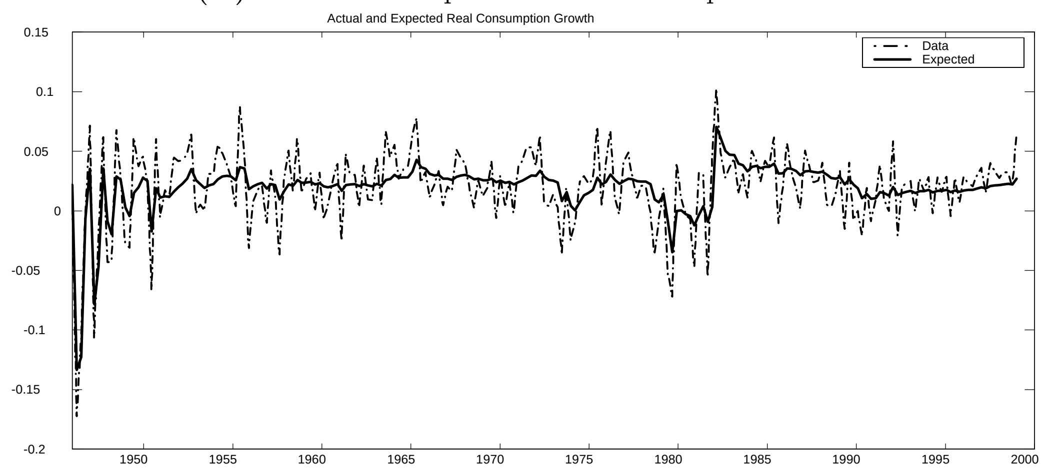
Figure 3: Time Series of Posterior Distribution on Consumption Drifts



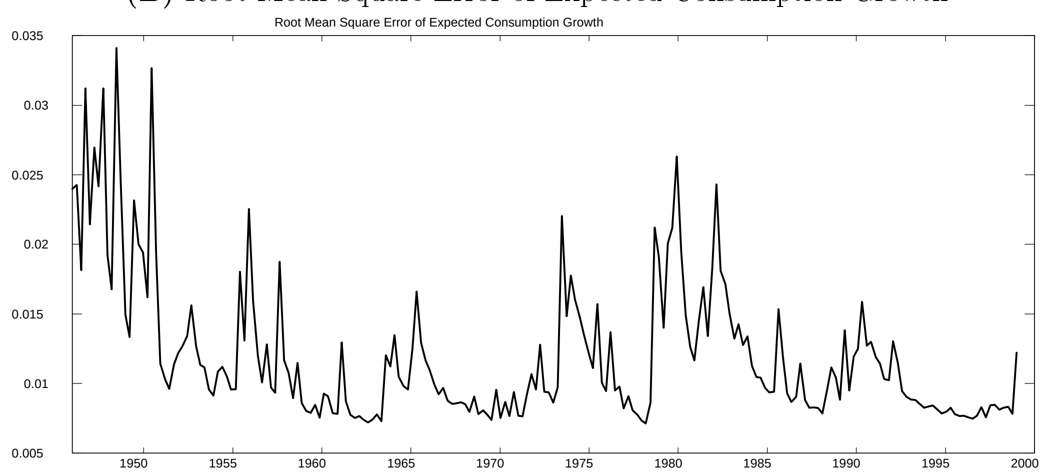
This figure plots the time series of the posterior distribution $\pi(t)$ on the (annualized) drift rate of consumption g_t computed using real consumption growth data from 1946 to 1999 and the MLE estimates in Table 2. Updating occurs by standard Bayes rule according to the formula $\pi_i(t+1) = \left\{ e^{-\frac{1}{2\sigma^2}(\Delta c(t+1)-g_i)^2} [\pi(t)\,\mathbf{\Lambda}]_i \right\} / \left\{ \sum_{j=1}^{n} e^{-\frac{1}{2\sigma^2}(\Delta c(t+1)-\theta_j)^2} [\pi(t)\,\mathbf{\Lambda}]_j \right\}$ where the transition matrix is given by $\mathbf{\Lambda} = \alpha f + (1-\alpha)\,\mathbf{I}$ and $f = (f_1, ..., f_n)$ is the discrete version of a normal distribution. The number of points in the discretized grid is $n = 200$.

Figure 4: Expected Growth Rate of Consumption and its Root Mean Square Error

(A) Actual and Expected Real Consumption Growht

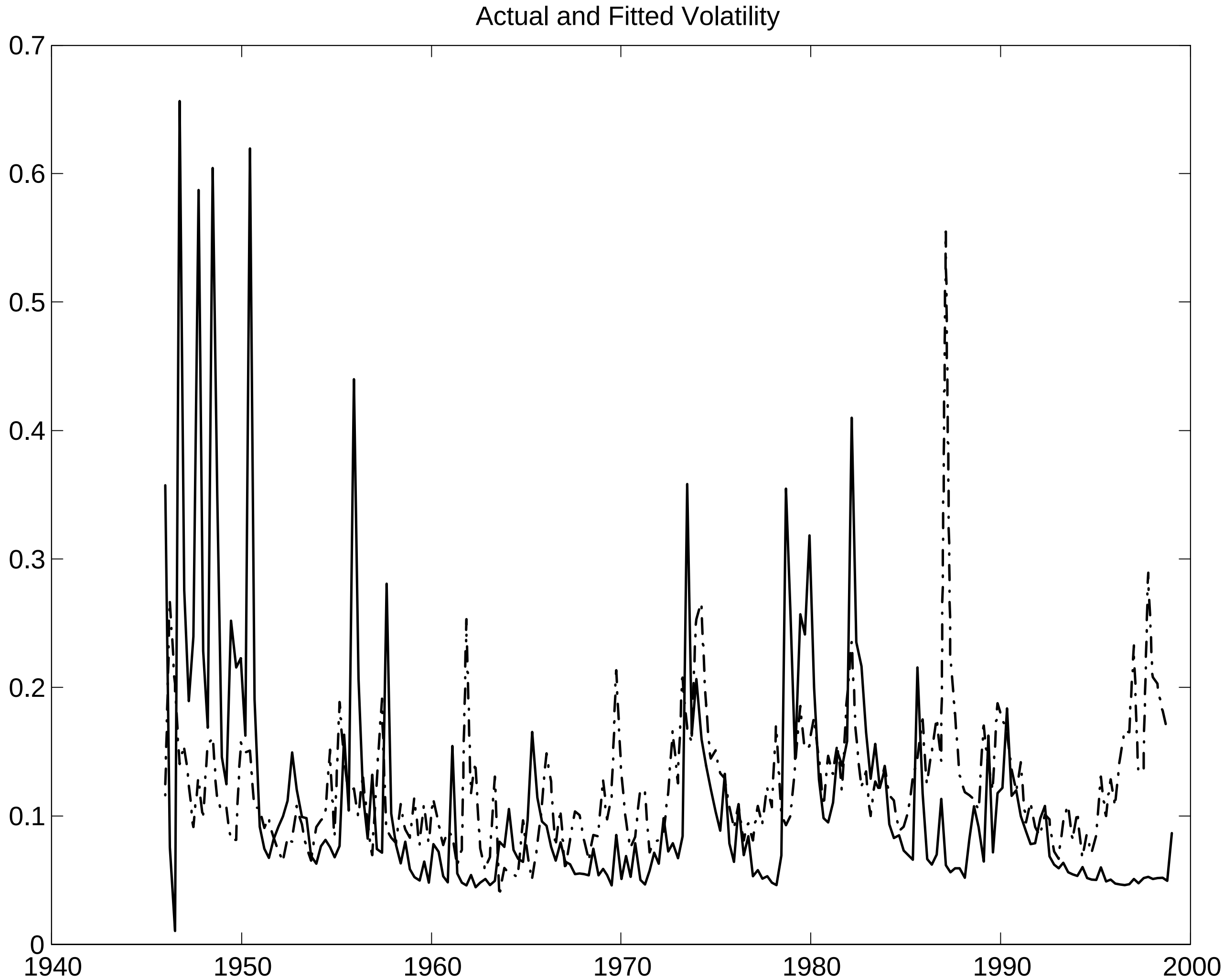


(B) Root Mean Square Error of Expected Consumption Growth



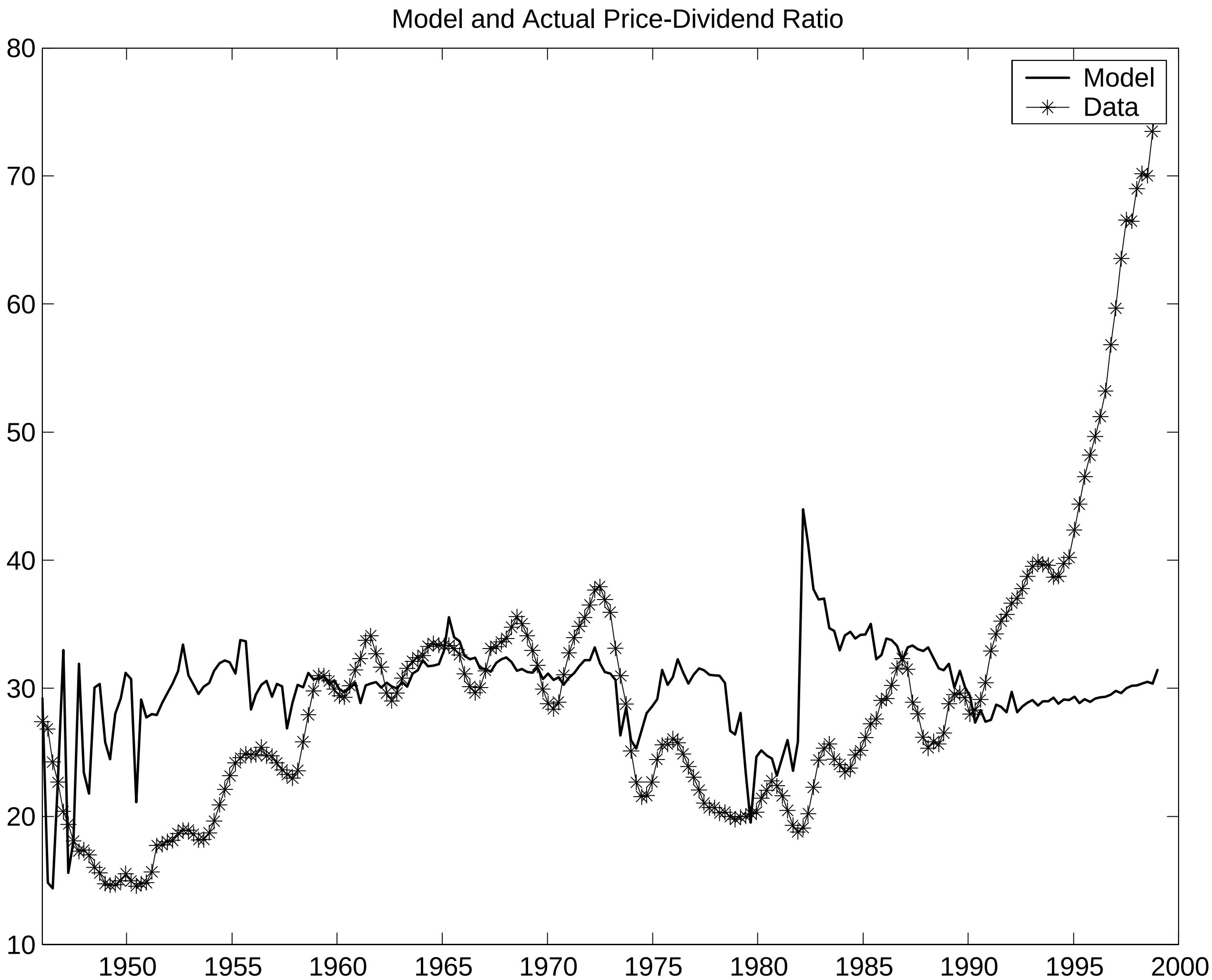
Panel (A) plots the time series of real consumption growth from 1946 to 1999 and the expected consumption growth obtained from the fitted posterior probabilities $\pi(t)$ as described in text and in Figure 2. Panel (B) plots the time series of the Root Mean Square Error of the expected consumption growth.

Figure 5: Actual and Model-Implied Volatility



This figure reports the plot of the conditional volatility implied by the model (solid line) and the ex-post integrated volatility computed from CRSP-data on stock returns (dash-dotted line). The conditional volatility implied by the model is computed applying the formula provided in the text and at every t it only depends on the posterior distribution $\pi(t)$ in Figure 2.

Figure 6: Actual and Model-Implied Price-Dividend Ratios



This figure plots the time series of the price-dividend ratio obtained from CRSP-data and the model-implied price dividend ratio. The latter is only a function of past consumption and it is given by the formula $P(t)/D(t) = \sum_{i=1}^{n} \overline{\pi}_i(t) B_i$.