FEDERAL RESERVE BANK OF SAN FRANCISCO

WORKING PAPER SERIES

# Doubts and Variability: A Robust Perspective on Exotic Consumption Series

Rhys Bidder
Federal Reserve Bank of San Francisco

Matthew E. Smith
Hutchin Hill Capital

September 2015

Working Paper 2013-28
http://www.frbsf.org/publications/economics/papers/2013/wp2013-28.pdf

**Suggested citation:**

Bidder, Rhys, Matthew E. Smith. 2015. "Doubts and Variability: A Robust Perspective on Exotic Consumption Series." Federal Reserve Bank of San Francisco Working Paper 2013-28. http://www.frbsf.org/economic-research/publications/working-papers/wp2013-28.pdf

**Doubts and Variability: A Robust Perspective on Exotic Consumption Series**

Rhys Bidder*
Federal Reserve Bank of San Francisco

Matthew E. Smith†
Hutchin Hill Capital

This draft: September 23, 2015

**Abstract**

Consumption-based asset-pricing models have experienced success in recent years by augmenting the consumption process in 'exotic' ways. Two notable examples are the Long-Run Risk and rare disaster frameworks. Such models are difficult to characterize from consumption data alone. Accordingly, concerns have been raised regarding their specification. Acknowledging that both phenomena are naturally subject to ambiguity, we show that an ambiguity-averse agent may behave as if Long-Run Risk and disasters exist even if they do not or exaggerate them if they do. Consequently, prices may be misleading in characterizing these phenomena since they encode a pessimistic perspective of the data-generating process.

---

*Economic Research Department, Federal Reserve Bank of San Francisco, 101 Market Street, San Francisco, CA 94115; Email: rhys.bidder@sf.frb.org

†Disclaimer: The views expressed in this paper, and all errors and omissions, should be regarded as those solely of the authors and not necessarily those of the Federal Reserve Bank of San Francisco, the Federal Reserve Board of Governors or the Federal Reserve System. Disclaimer: This paper represents the personal work product and views of Matthew Smith and have neither been reviewed nor endorsed by Hutchin Hill Capital, LP, its affiliates, any funds advised by them, or any of their officers, directors, employees, agents, partners, principals, owners, or shareholders (collectively, "Hutchin Hill Parties"). No representation or warranty is made by Matthew Smith or any Hutchin Hill Party about the completeness or accuracy of any information provided herein, and such parties expressly disclaim all liability for any of the information contained in these materials. These materials have been prepared for educational use only and may not be relied upon for any other purpose (including without limitation any commercial purpose). These materials do not constitute an offer, or the solicitation of an offer, to buy or sell any security or financial instrument or investment advisory services. No Hutchin Hill Party shall be liable or have any obligation with respect to this paper or any reliance thereon.

## 1 Introduction

A principal challenge for the early generations of consumption-based asset pricing models was to generate sufficiently high risk premia and prices of risk while respecting the apparent smoothness in consumption growth from period to period (Mehra and Prescott (1985) and Hansen and Jagannathan (1991)). One response to this challenge was been to introduce 'exotic' elements in consumption dynamics. Notably, the Long-Run Risk (LRR) model of Bansal and Yaron (2004) asserts a small but persistent component in consumption growth that allows the process to exhibit considerable risk over longer horizons without introducing counterfactual volatility at high frequencies. An alternative strategy is to appeal to rare disasters in consumption growth (Rietz (1988), Barro (2006) and Wachter (2013)).[1]

Despite the successes achieved using these exotic processes, concerns have been raised as to their specification, or even existence. By definition, direct evidence of the LRR component is hard to detect in post-war consumption data, leading to the question of whether or not the component actually exists (Hansen, Heaton, and Li (2008), Marakani (2009) and Beeler and Campbell (2012)). With regard to disasters, as noted in Dolmas (2013), the rarity of the phenomenon in question inevitably undermines empirical analysis. Unlike in the LRR case there is perhaps no debate as to whether disasters *exist* but there certainly is substantial debate over their calibration.

In this paper, we take a different approach. Rather than positing the existence of Long-run Risk or taking a firm stance on the calibration of rare disasters we show that an ambiguity averse agent's fear of model misspecification can nevertheless generate or exaggerate these phenomena endogenously in the mind of the agent.[2] Consequently, one explanation for why these exotic properties appear to be encoded in prices, but are simultaneously difficult to identify in the consumption data directly, is that prices reflect not only the true model of consumption but also the agent's fear of misspecification.

Our agent does not fully trust her 'benchmark' model of consumption growth, as captured by the implied joint distribution over sequences of its various components. She acknowledges that the benchmark is an approximation to the true data generating process and fears it is misspecified in some unknown way. She expresses these fears by envisaging alternative probability distributions, which she thinks may plausibly describe consumption growth. Formally, we endow the agent with a desire for robustness to model misspecification, as captured by the multiplier preferences of Hansen and Sargent (2008). As a way of constructing a robust

[1]See also Barro (2006), Barro and Ursua (2008), Gourio (2012), Gabaix (2012) and Nakamura, Steinsson, Barro, and Ursa (2013)

[2]For an interesting alternative approach to generating a phenomenon interpretable as subjective long-run consumption risks, see Collin-Dufresne, Johannes, and Lochstoer (2013).

evaluation of random payoffs the agent envisages adverse distortions, balancing the damage they could cause against their plausibility. A particular 'worst case' distribution emerges from the agent's optimization problem, allowing insight into the sort of misspecifications she desires robustness against. The agent then evaluates risky payoffs *as if* this worst case is generating the data. We show that that this worst case will naturally encode phenomena akin to LRR and disasters.

In our simplest specification the benchmark model features white noise consumption growth with persistent variation in its conditional volatility. The worst case reflects the agent's fear of misspecifications that would imply lower growth and higher volatility, as represented by negative and positive mean shifts in the marginal distributions for endowment and volatility innovations, respectively. Most importantly, the agent's pessimism becomes more extreme when volatility is high, as captured by a greater negative distortion to the mean of the endowment innovation. Since the volatility process is persistent, the consumption growth process under the worst case therefore inherits this persistence. Consequently, the worst case exhibits the hallmark of LRR models - a small but persistent component in consumption growth. We generalize from the white noise benchmark to introduce a degree of LRR in the true data generating process but show how the amount of LRR can be reduced from a counterfactually high level, while still allowing asset pricing success, if one trades off LRR in the benchmark against LRR in the worst case model in the mind of the agent.

The association of high volatility with low growth under the worst case also induces negative skewness in the consumption growth process which (together with an overall unconditional negative mean shift) implies that dramatic declines in consumption occur significantly more frequently than under the benchmark model so that 'disasters' are more common, in terms of a longer left tail of the unconditional distribution of consumption growth. In order to allow for more a more standard, 'conditional' concept of disasters we go on to consider a benchmark model featuring a non-normal 'jump' component in consumption growth to allow for occasional dramatic declines in consumption. In this context we show that disasters are a more powerful phenomenon to interact with robustness, in the sense that a more reasonable calibration of ambiguity aversion can attain stylized asset pricing facts, such as elevated premia and prices of risk, along with predictability evidence, than in the heteroscedastic Gaussian case. It is entirely reasonable for an agent with a plausible degree of ambiguity aversion to behave as if disasters arrive with significantly greater frequency than implied by the model and, thus, not allowing for this fear of misspecification could undermine inference.

Our paper also contributes to the methodology of robust control analysis in that we show how to exploit the formalism of cumulant generating functions (CGFs) to obtain clear characterizations of the worst case distribution in discrete time. More importantly, we develop methods to draw from the worst case distribution using importance sampling and other Monte Carlo techniques that are more broadly applicable in models featuring robust agents.

The paper is organized as follows: Section 2 introduces our modeling framework, including our basic description of the benchmark consumption process and preferences. Section 3 solves the model and shows how we characterize the worst case distribution, analytically and using Monte Carlo methods. Section 4 contains our results on a robust interpretation of LRR. Section 5 contains our results on a robust reinterpretation of the rare disasters framework. Section 6 concludes.

## 2 Modeling framework

We here lay our basic modeling framework, constituting a description of an endowment economy and the preferences of an ambiguity averse representative agent.

### 2.1 The endowment process

We take as a starting point the Long-run Risk (LRR) model of Bansal, Kiku, and Yaron (2012), augmented by a non-Gaussian 'jump' process. In what follows we will examine special cases of this framework but we describe it here in its full generality.

$$
\begin{aligned}
\log g_{t+1} &= G_0 + x_t + w_{z,t+1} + v_t^{0.5} w_{g,t+1} \\
\log g_{d,t+1} &= G_{d,0} + \phi x_t + \phi_z w_{z,t+1} + \varphi_d v_t^{0.5} w_{d,t+1} + \tau_d v_t^{0.5} w_{g,t+1} \\
x_{t+1} &= \rho x_t + \varphi_x v_t^{0.5} w_{x,t+1} \\
v_{t+1} &= (1-\varphi_v)\,\bar{v} + \varphi_v v_t + \sigma_v w_{v,t+1} \\
h_{t+1} &= \delta_h c_h + \varphi_h h_t + w_{h,t+1}
\end{aligned}
$$

where $g_t \equiv \frac{C_t}{C_{t-1}}$ and $g_{d,t} \equiv \frac{D_t}{D_{t-1}}$ are consumption and dividend growth, respectively. $w_{i,t+1}$ for $i \in \{g, d, x, v\}$ are standard Normal iid innovations. $w_{z,t+1} \backsim PN(\theta, \delta; h_t)$, meaning that $w_{t+1}$ is a Poisson mixture of Normals, such that

$$
\begin{aligned}
w_{z,t+1}|j_{t+1} &\backsim N\left(j_{t+1}\theta_z, j_{t+1}\delta^2\right) \\
j_{t+1}|t &\backsim \frac{e^{-h_t} h_t^{j_{t+1}}}{j_{t+1}!}
\end{aligned}
$$

Finally, $w_{h,t+1}$ is a Martingale difference series that yields an Autoregressive Gamma (ARG) process for $h_t \backsim ARG(c_h, \varphi_h, \delta_h)$.[3] If one sets $c_h = \frac{\sigma_h^2}{2}$ and $\delta_h = (1-\varphi_h)\frac{\bar{h}}{c_h}$ then we have

$$h_t = (1-\varphi_h)\bar{h} + \varphi_h h_{t-1} + w_{h,t}$$

and we can interpret $\bar{h}$ as the steady state jump intensity.

We will refer to $x_t$ as the LRR component since it allows us to introduce time variation in the conditional mean of consumption growth. With appropriate calibration, $x_t$ allows the power of the consumption growth process to be focused at low frequencies - leaving consumption relatively smooth at high frequencies. This is at the core of the standard LRR mechanism of Bansal and Yaron (2004). The $w_{z,t}$ component allows us to introduce non-Normality and, in particular, conditional skewness in the consumption growth process. In the work of Backus, Chernov, and Martin (2011) and Backus, Chernov, and Zin (2014) this specification has been used as a flexible way of modeling disasters.

When simulating this system in the presence of stochastic volatility we censor the simulations of $v_t$ at a very small positive number but do not adjust the solution procedure to take account of this censoring. This approach is standard (see Bansal and Yaron (2004) and Beeler and Campbell (2012)) and, as discussed in Backus, Chernov, and Zin (2014), for reasonable calibrations of the the volatility process the solution remains a good approximation to that obtained when explicitly taking account of the censoring. The choice of an ARG process for $h_t$, rather than a Gaussian autoregression (AR), as for $v_t$ reflects the fact that, for the calibrations we will use below, the probability of the process going negative in the AR case is uncomfortably high. Of course, one could also use an ARG for $v_t$ but we eschew this because, setting aside the low probability of $v_t$ going negative under our calibrations, the analytic insight and computational benefits we obtain in the AR case are considerable.

### 2.2 Preferences

There is much evidence suggesting that, when faced with situations that are 'ambiguous' or uncertain in the Knightian sense, agents do not behave in accordance with standard axioms of choice. One commonly cited example of this is the famous paradox of Ellsberg (1961). An important literature has emerged suggesting possible formalizations for how agents make decisions in such contexts. One of the more prominent and intuitively appealing formalizations is the Robust Control framework, which has been adapted to economic applications by Hansen and Sargent (2008).

[3] The ergodic distribution for $h_t$ is Gamma with shape parameter, $\delta_h$ and scale parameter $\frac{c_h}{1-\varphi_h}$. The conditional variance of $w_{h,t+1}$ is $\delta_h c_h^2 + 2\varphi_h c_h h_t$.

### 2.2.1 An intuitive description of robustness

Our agent is endowed with a 'benchmark' model which she takes to be a reasonable description of the world, but which she suspects is misspecified. She expresses her doubts over her model by considering alternative distributions that are distorted versions of the joint distribution over sequences implied by the benchmark. The alternative distributions considered are not expressed as explicit structural models, but only as statistical objects - changes in measure relative to the benchmark distribution. This captures the intuitively appealing idea that, faced with ambiguity, an agent may not be able to exhaustively enumerate 'known unknowns' and instead faces 'unknown unknowns'.

Reflecting an aversion to model uncertainty and her desire to identify vulnerabilities to misspecification, the agent explores adverse distortions to the benchmark distribution. She disciplines this search by considering only 'plausible' alternative models. Thus, the agent discounts more heavily distributions that are more distant from the benchmark. Implicit in the discounting of alternatives according to their distance from the benchmark is the idea that the benchmark has been obtained through a reasonable selection process and is thought a 'good', if flawed model. The agent's assessment of an alternative model's plausibility is tied to whether or not it is difficult to distinguish the benchmark and the worst case using a realistic sample size of data. If they are difficult to disttinguish then it is reasonable to think that the agent might regard the worst case and other similar models as possibly the truth.[4]

In balancing the pain of an implicit misspecification against its plausibility the agent identifies what is typically referred to as a 'worst case distribution'. This distribution is not the agent's 'belief' or even the only distribution she fears but it is especially revealing of the agent's concerns about misspecification. Properties exhibited by this distribution indicate the sort of misspecifications that particularly worry the agent and, importantly, the agent behaves as if this model is the truth. We now formalize this intuition.

### 2.2.2 A formal description of robustness

A robust agent entertains a benchmark model in which the state, control and innovation sequences are related according to the (possibly nonlinear) vector valued equation

$$s_{t+1} = g(s_t, u_t, w_{t+1}) \tag{1}$$

[4]The model selection process and any learning and estimation by the agent in deriving the benchmark are left unspecified. For work combining robustness with filtering and learning, see Hansen and Sargent (2007) and the relevant chapters of Hansen and Sargent (2008).

where $s_t$ is the state vector, $u_t$ is a vector of controls and $w_t$ is a vector-valued random innovation. Since our analysis will ultimately deal exclusively with representative agent endowment economies, our inclusion of an explicit control is in some sense redundant since consumption is exogenous and all asset holdings must be in zero net supply in equilibrium. Nevertheless, it is pedagogically convenient for now to include it in this description of a robustness problem. In addition, of course, envisaging the agent as deciding on asset positions given prices is a fundamental ingredient to the equilibrium concept, even if the representative agent will hold a net zero position in equilibrium.

Given a control law, $u_t = u(s_t)$, and a density, $p_w(w_{t+1}|s_t)$, for $w_{t+1}$, equation (1) implies a benchmark transition density $p(s_{t+1}|s_t)$. It is convenient to partition the state, $s_t$ into elements unknown on entering the period, which we identify with $w_t$, and those elements that are predetermined, denoted $\hat{s}_t$. We capture the dependence of $\hat{s}_t$ on the state prevailing in the previous period by the function $f$, such that $\hat{s}_t = f(s_{t-1})$. With this decomposition we have $p(s_{t+1}|s_t) = p_w(w_{t+1}|s_t)\delta_{f(s_t)}(\hat{s}_{t+1})$.[5] In the context of our model, $w_t \equiv (w_{d,t}, w_{g,t}, w_{x,t}, w_{v,t}, w_{z,t}, w_{h,t})'$ and $\hat{s}_t \equiv (x_{t-1}, v_{t-1}, h_{t-1})'$.

We endow the agent with multiplier preferences, which are discussed extensively in Hansen and Sargent (2008) and are axiomatized in Strzalecki (2011). Our agent is not a Bayesian - her problem progresses from a situation of multiple models to making a decision not by integrating over the models with respect to a unique prior (essentially resolving the multiple models to a single hyper-model) but by a penalized max-min approach. Formally, the decision problem of the agent takes the form of a particular two-player zero-sum game between the robust agent (the maximizer) and a metaphorical 'evil agent' or 'nature' (the minimizer)

$$\max_{\{u_t\}} \min_{\{m_{t+1}\}} \sum_{t=0}^{\infty} E\left[\beta^t M_t \left\{h(s_t, u_t) + \beta\theta E\left(m_{t+1}\log m_{t+1}|\Im_t\right)\right\}|\Im_0\right] \tag{2}$$

where $h(\cdot,\cdot)$ is the period payoff function (taken to be $(1-\beta)\log C_t$ in our case) and the problem is subject to equation (1), $M_{t+1} = m_{t+1}M_t$, $E[m_{t+1}|\Im_t] = 1$, $m_{t+1} \geq 0$ and $M_0 = 1$. We assume that the robust agent's information set, $\Im_t$ contains the entire history of states. Thus, $\{m_{t+1}, t \geq 0\}$ is a sequence of Martingale increments that recursively define a non-negative Martingale, $M_t = M_0 \prod_{j=1}^{t} m_j$.

$M_t$ defines Radon-Nikodym derivatives that twist the measures implicit in the benchmark model to yield absolutely continuous measures that represent alternative distributions considered by the agent. The agent's desire for robustness is reflected in the minimization over the sequence of martingale increments, $m_t$, chosen by the 'evil' player to twist the distributions

[5]Note that the $s_t$ may contain $w_t$ as an element of the state so that an identity mapping is implicit in $g$. $\delta_{f(s_t)}(\cdot)$ takes the value of unity at $f(s_t)$ and zero elsewhere.

towards realizations of the state sequence that are painful to the robust agent. This accords with our earlier intuition of the agent identifying her vulnerability to model misspecification by envisaging adverse alternative models or distributions.

The degree of robustness is controlled by the penalty parameter, $\theta > 0$, that enters the objective by multiplying the relative entropy associated with a given distortion. The penalty reflects our earlier intuition that the agent considers models that, although different, are somehow 'near' the benchmark. A particular alternative distribution, associated with a particular Martingale, may be especially painful in the sense of implying a very expected payoff, but may not solve the minimization problem due to the offsetting effect of the entropy penalty. Thus, the two components in equation (2) capture the way in which the robust agent balances pain and plausibility.

We seek a recursive expression of the problem and, invoking results in Hansen and Sargent (2008), obtain a value function of the following form

$$\begin{aligned} V(w_t, \hat{s}_t) &= \max_{u_t} \min_{m(w_{t+1}, \hat{s}_{t+1})} h(s_t, u_t) && (3) \\ &+ \beta \int m(w_{t+1}, \hat{s}_{t+1}) V(w_{t+1}, \hat{s}_{t+1}) p_w(w_{t+1}|s_t) && (4) \\ &+ \theta m(w_{t+1}, \hat{s}_{t+1}) \log m(w_{t+1}, \hat{s}_{t+1}) p_w(w_{t+1}|s_t) dw_{t+1} && (5) \end{aligned}$$

subject to $\int m(w_{t+1}, \hat{s}_{t+1}) p(w_{t+1}|s_t) dw_{t+1} = 1$ for all values of $\hat{s}_{t+1}$.

Under the twisted measures one can form objects interpretable as expectations taken in the context of a distorted alternative model. Thus, we define a distorted conditional expectation operator to be

$$\tilde{E}_t[b_{t+1}] \equiv E[m_{t+1} b_{t+1} | \Im_t]$$

for some $\Im_{t+1}$ measurable random variable $b_{t+1}$, given $\Im_t$. The conditional relative entropy associated with the twisted conditional distribution is given by the term $E[m_{t+1} \log m_{t+1} | \Im_t]$, which is a measure of how different the distorted measure is from the benchmark.

### 2.2.3 The worst case distribution

Solving the inner minimization problem we obtain the minimizing, or 'worst case', Martingale increment, which has the form

$$m(w_{t+1}, \hat{s}_{t+1}) = \frac{e^{-\frac{V(w_{t+1}, \hat{s}_{t+1})}{\theta}}}{E\left[e^{-\frac{V(w_{t+1}, \hat{s}_{t+1})}{\theta}} | s_t\right]} \tag{6}$$

Substituting this solution into the original problem, we obtain the following expression

$$V(s_t) = \max_{u_t} h(s_t, u_t) - \beta\theta \log E\left[\exp\left(-\frac{V(s_{t+1})}{\theta}\right)|s_t\right]$$

and, finally, we abstract from the max operator and explicit presence of a control, in anticipation of our endowment economy, representative agent analysis

$$V(s_t) = h(s_t) - \beta\theta \log E\left[\exp\left(-\frac{V(s_{t+1})}{\theta}\right)|s_t\right] \tag{7}$$

The Martingale $M_t$ implied by the solution of the agent's problem is a ratio of joint densities, $\frac{\tilde{p}(s_{1:t}|\Im_0)}{p(s_{1:t}|\Im_0)}$, where $p$ and $\tilde{p}$ denote the benchmark and worst case densities over state sequences, $s_{1:t}$, conditional on information at $t = 0$. The minimizing Martingale increment in equation (6), is the associated ratio of conditional densities, $\frac{\tilde{p}(s_{t+1}|s_t)}{p(s_{t+1}|s_t)}$ that is more natural to work with in the recursive formulation of our problem.

$\tilde{p}(s_{t+1}|s_t) = m(s_{t+1})p(s_{t+1}|s_t)$ is the worst case conditional distribution of the state and implicit in this is the conditional distribution over innovations, $\tilde{p}_w(w_{t+1}|s_t) = m(w_{t+1}, \hat{s}_{t+1})p_w(w_{t+1}|s_t)$. By iteratively drawing from $\tilde{p}_w(w_{t+1}|s_t)$ and evolving the state according to the law of motion (1) we obtain draws of sequences from $\tilde{p}(s_{1:t})$.

While $\tilde{p}$ is not directly interpretable as the 'beliefs' of the agent, the fact that it differs from $p$ emphasizes that, unlike under Rational Expectations, more than one distribution plays a role in the equilibrium. In what follows, we assert equality between the agent's benchmark model and the true model so that the agent's fears of mis-specification are 'all in her head' and our only (highly restricted) deviation from Rational Expectations is to relax the requirement that the agent fully trusts the benchmark.[6]

We also note that equation (7) is algebraically equivalent to that of an agent with risk sensitive preferences (see Tallarini (2000) and Barillas, Hansen, and Sargent (2009)) and that of an Epstein-Zin agent with unity elasticity of intertemporal substitution (EIS). In the latter case, we have the relation $\alpha \equiv -\frac{1}{\theta}$ where $1-\alpha$ is the coefficient of relative risk aversion. Under the risk sensitivity and Epstein-Zin interpretations, however, $\theta$ reflects sensitivity to well defined, Knightian 'risk' whereas here it reflects the degree to which the agent fears Knightian 'uncertainty'. Further details of this equivalence are discussed in A.2.

[6] We refer the reader to Hansen and Sargent (2010a) for an extensive discussion of the relationship between Robust Control, Rational Expectations and other decision-making and modeling paradigms.

### 2.2.4 The stochastic discount factor

Our agent prices assets using a stochastic discount factor of the following form

$$\begin{aligned}\Lambda_{t,t+1} &= \Lambda^R_{t,t+1}\Lambda^U_{t,t+1} \\ \Lambda^R_{t,t+1} &= \beta\left(\frac{C_{t+1}}{C_t}\right)^{-1} \\ \Lambda^U_{t,t+1} &= \frac{\exp\left(\frac{-V_{t+1}}{\theta}\right)}{E_t\left[\exp\left(\frac{-V_{t+1}}{\theta}\right)\right]}\end{aligned}$$

Thus, provided that $\theta < \infty$, the stochastic discount factor comprises two components. The first component, $\Lambda^R_{t,t+1}$, is the stochastic discount factor derived from time separable logarithmic preferences. The second component, $\Lambda^U_{t,t+1}$, is the minimizing Martingale increment discussed above. $\Lambda^U_{t,t+1}$ induces a wedge in the fundamental asset pricing equation relative to the expected utility case. Clearly, as $\theta \to \infty$ we recover expected utility as the concern for robustness vanishes. In this case, the entropy penalty term in the robustness problem (2) is given infinite weight so the solution to the minimization features no deviation from the benchmark distribution and the Martingale increment is identically unity.

Consequently, an agent who fears her model is misspecified will price assets as if she has logarithmic period utility, but under a distorted conditional expectations operator $\tilde{E}_t$ and an associated distorted density $\tilde{p}(x_t|x_{t-1})$ as follows

$$\begin{aligned}1 &= E_t\left[\Lambda^R_{t,t+1}\Lambda^U_{t,t+1}R_{t+1}\right] \\ &= \tilde{E}_t\left[\Lambda^R_{t,t+1}R_{t+1}\right]\end{aligned}$$

For pricing assets beyond the next period, a similarly distorted expectations operator will be used. This latter operator will encode how the distorted conditional distribution of the state (captured by $\Lambda^U_{t,t+1}$) varies over time and across contingencies

$$\begin{aligned}1 &= E_t\left[\Lambda^R_{t,t+k}\Lambda^U_{t,t+k}R_{t+k}\right] \\ &= \tilde{E}_t\left[\Lambda^R_{t,t+k}R_{t+k}\right] \\ \Lambda^R_{t,t+k} &\equiv \prod_{j=0}^{k-1}\Lambda^R_{t+j,t+j+1} \\ \Lambda^U_{t,t+k} &\equiv \prod_{j=0}^{k-1}\Lambda^U_{t+j,t+j+1}\end{aligned}$$

In general, the worst case conditional distribution will be state dependent. Thus, while it is true that the agent will, in $t$, price payoffs in $t+1$ as an expected utility agent with log preferences under the worst case and thus only price $t+1$ shocks to consumption growth, distorted dynamics under the worst case can introduce interesting differences in pricing, relative to the expected utility case.

Our stochastic discount factor does not have the standard interpretation of an intertemporal marginal rate of substitution under a fully trusted model. Instead it involves a component that reflects the agent's concern that her model of the economy might be misspecified in a way that is damaging for lifetime utility and does not simply relate to marginal utility in future periods.

#### 2.2.5 Interpreting $\theta$

To be able to interpret how the agent trades off her concerns for misspecifications against their plausibility, we will follow the standard approach of the literature in connecting $\theta$ to a measure of statistical closeness, referred to as 'detection error probability' (Hansen and Sargent (2008) and Barillas, Hansen, and Sargent (2009)).[7]

Detection error probabilities (DEPs) characterize a set of distorted models in terms of whether or not, with a limited amount of data, an agent could accurately distinguish between the worst case and benchmark models using likelihood ratio tests. That is, we ask what the probability of mis-identifying the data generating process would be if one were running likelihood ratio tests on data generated under the benchmark and the worst case.

Formally, one picks a sample length, $T$ and simulates many samples of length $T$ from both the benchmark and worst case distributions. Each time, one calculates a likelihood ratio test of the benchmark versus the worst case and tallies the number of times that the test favors the model that did not generate the data - a 'detection error'. One then calculates the equally weighted average of the fraction of detection errors obtained from simulations under the benchmark and under the worst case. We denote this average the 'detection error probability'.[8]

In order to calculate the likelihood ratios we must make a decision on how to initialize the simulations and how to treat the $t = 0$ period in the likelihood evaluations. We will draw the initial state for each $T$-long simulation from the relevant unconditional distribution (or an approximation thereof, if necessary). We choose not to incorporate the unconditional likelihood of the time zero observation into the likelihoods used for the DEP calculations, as

[7] If we were interpreting preferences from the Epstein-Zin perspective we could appeal to Arrow-Pratt intuition for the calibration of $\theta$, but that is not applicable here.

[8] Clearly this number is subject to sampling variability so one must use many draws of sequences of length $T$ to obtain a reliable measure of detectability.

is consistent with Hansen and Sargent (2008), Chapter 9. If we were to allow the agent to use the time zero observation then it would implicitly be as if we were allowing the agent far longer hypothetical samples than length $T$ to distinguish the two models. Further details of our DEP calculations are included in A.11 and B.6.

If two models have similar properties, it will be be difficult to determine which model generated a sample of data using sample sizes that are typically available for analysis and which implicitly might have been used to obtain the benchmark model. In this case the DEP will be close to 0.5, indicating that the models are almost indistinguishable. Thus, it is plausible that the worst case model, or one like it, might be the true generating process and it is therefore reasonable to seek robustness against such models. In contrast, if the models are easily distinguished (a DEP of close to 0) then this suggests an implausible timidity on the part of the agent in that she is seeking robustness against unreasonable models with dramatically counterfactual implications - models that only a very limited amount of data (such as the sample that implicitly led to the selection of the benchmark model) would suggest could not plausibly be the truth.

The larger is $\theta$, the closer the benchmark will be to the worst case, due to the more rapid offsetting effect of the entropy penalty in the minimization problem (see (2) and (6)). Consequently, the higher will be the detection error probabilities. It is this connection between $\theta$ and the detection error probabilities, together with an intuitive sense of a reasonable DEP, that will allow us to interpret our calibration of the agent's robustness.

## 3 Model solution

We will guess and verify an exponentially affine solution for the scaled utility function, $u_t \equiv \frac{U_t}{C_t}$

$$u_t = \exp\{F_0 + F_1 x_t + F_2 v_t + F_3 h_t\}$$

Given this solution we can obtain an expression for the stochastic discount factor that makes clear the distinction between the influence of the standard, expected utility elements of the agent's preferences, and those that reflect her preference for robustness

$$\begin{aligned}
\log \Lambda_{t+1} &= \log \Lambda^r_{t+1} + \log \Lambda^u_{t+1} \\
\log \Lambda^r_{t+1} &= k_t + \lambda^r_{g,t} w_{g,t+1} + \lambda^r_{z,t} w_{z,t+1} \\
\log \Lambda^u_{t+1} &\equiv \xi_t + \lambda^u_{g,t} w_{g,t+1} + \lambda^u_{x,t} w_{x,t+1} + \lambda^u_{v,t} w_{v,t+1} \\
&\quad + \lambda^u_{z,t} w_{z,t+1} + \lambda^u_{h,t} w_{h,t+1}
\end{aligned}$$

where the 'risk' prices, $\lambda^r_{i,t}$, 'uncertainty' prices, $\lambda^u_{i,t}$ , $k_t$ and $\xi_t$ are functions of deep parameters, the solution for the value function and, potentially, the current level of volatility and jump intensity. They are described in the appendices.[9]

### 3.1 Characterizing the worst case

The worst case distribution is an important object beyond simply being an artifact of a decision problem since it is suggestive of the nature of other plausible and damaging models that might concern the agent. It can therefore play a diagnostic and revelatory role where introspection may not be sufficient to intuit a worrisome structural misspecification. By examining the properties of the worst case one can identify classes of structural models that are worth guarding against and researching further. Indeed, much of this paper deals with suggesting structural interpretations for moments implied by the worst case distribution over sequences.

Since the worst case distribution over sequences can be built up from the worst case conditional distributions of innovations, combined with the law of motion, we will spend time deriving the properties of these innovation distributions. To do this we will make use of cumulant generating functions and Monte Carlo techniques.

#### 3.1.1 Cumulant generating functions

Under the benchmark model, innovation $w_{i,t+1}$ has conditional cumulant generating function

$$k_t^i(u) = \log E_t \left[\exp\left\{uw_{i,t+1}\right\}\right]$$

The $j^{th}$ cumulant is obtained by evaluating the $j^{th}$ derivative of the CGF at 0 with the first and second cumulants being the mean and variance respectively. Moments can be derived given knowledge of the cumulants.

For $i \in \{d, g, x, v\}$ we use the standard Normal CGF

$$k_t^i(u) = \frac{u^2}{2}$$

The Poisson-Normal mixture, $w_{z,t+1}$ has CGF

$$k_t^{wz}(u) = h_t\left(e^{u\theta + \frac{(u\delta)^2}{2}} - 1\right)$$

[9]Remaining endogenous variables can typically be calculated in a similar way, utilizing exponential affine approximations to the true equilibrium objects. In the case of equity returns we utilize the approximation discussed in Campbell and Shiller (1989) and, where necessary, an approximation based on the zero-coupon term structure of equity, advocated by Lettau and Wachter (2011).

and the innovation to the ARG process, $w_{h,t+1}$ has CGF

$$k_t^{wh}(u) = u\varphi_h \left( \frac{1}{1 - sc_h} - 1 \right) h_t - \delta_h \left( sc_h + \log\left(1 - sc_h\right) \right)$$

To obtain analogous CGFs under the worst case, denoted $\tilde{k}_t^i$, we apply the worst case change in measure captured in $\Lambda_{t+1}^u$. Note that given the linearity of $\log \Lambda_{t+1}^u$ in the innovations and their independence under the benchmark, they are also conditionally independent under the worst case. As shown in the appendix (available on request),

$$\begin{aligned} \tilde{k}_{i,t}(s) &\equiv \log \tilde{E}_t \left[ \exp\left\{ s\, w_{i,t} \right\} \right] \\ &\equiv \log E_t \left[ \exp\left\{ \left( s + \lambda_{i,t}^u \right) w_{i,t} \right\} \right] + tis \\ &\equiv k_{i,t}\left( s + \lambda_{i,t}^u \right) + tis \end{aligned}$$

where $tis$ represents terms independent of $s$. Thus, the uncertainty prices $\lambda_t^i$ shift the origin of the worst case CGFs, relative to the benchmark.

### 3.1.2 Monte Carlo Methods

Making use of the worst case CGFs, we observe that the standard Normal innovations remain Normal with unit standard deviation under the worst case, but have their means pessimistically shifted by an amount equal to their uncertainty prices. We also show in B.5.1 that $w_{z,t+1}$ retains its Poisson-Normal structure under the worst case, but with a pessimistically distorted arrival rate of the Poisson jump and, conditional on a jump, a negatively distorted mean of the Normal component. Thus, we do not only have the ability to calculate any given moment of these innovations under the worst case, but we actually know the innovations belong to a familiar class of random variables and can thus be easily characterized and drawn from.

We do not retain such clarity in the case of $w_{h,t+1}$. While we can calculate any of its cumulants under the worst case, its distribution does not belong to any recognized class. More generally, had we used a different solution or approximation method other than the exponential-affine approach we have adopted here, we would have confronted similar problems for the other innovations. However, we can still characterize and draw from these distributions if we can evaluate an approximation to the worst case density $\tilde{p}_w(w_{t+1}|s_t)$. This requires simply that we can obtain an approximation to the value function because since we can then construct the minimizing Martingale increment and use it to pre-multiply the benchmark density.

If we can evaluate the worst case density, we can draw from it using a variety of Monte

Carlo methods. For example, we can employ Sampling Importance Resampling (SIR) algorithm of Rubin (1987) and Smith and Gelfand (1992) for drawing from the worst case distribution. This entails obtaining draws from $p_w(w_{t+1}|s_t)$, computing associated importance weights (given by the minimizing Martingale increment) and then resampling with replacement according to those weights. This yields approximate draws from $\tilde{p}_w(w_{t+1}|s_t)$ which we can use to construct draws from $\tilde{p}(s_{t+1}|s_t)$. Thus, by using these methods one can apply robust control analysis to a broad class of discrete time non-linear models, assuming one is able to obtain a reasonable approximation to the value function.

## 4 Results: A robust perspective on long-run risk

In this section we will address a particular case of the general model discussed in section 2 and show that concerns for model misspecification, representable by fears of additional persistence, can substitute for more extreme calibrations of this phenomenon in the actual data generating process.

We abstract from the the presence of $w_{z,t}$ and focus on the heteroscedastic Gaussian system given below

$$\begin{aligned}
\log g_{t+1} &= G_0 + x_t + v_t^{0.5} w_{g,t+1} \\
\log g_{d,t+1} &= G_{d,0} + \phi x_t + \varphi_d v_t^{0.5} w_{d,t+1} + \tau_d v_t^{0.5} w_{g,t+1} \\
x_{t+1} &= \rho x_t + \varphi_x v_t^{0.5} w_{x,t+1} \\
v_{t+1} &= (1-\varphi_v)\,\bar{v} + \varphi_v v_t + \sigma_v w_{v,t+1}
\end{aligned}$$

Due to the analytic expressions obtained for the worst case innovations *and* the fact that they remain within convenient classes of distributions, the worst case can be defined explicitly in terms of a system much like that above.[10] The mean shifts under the worst case are simply the uncertainty prices, $\lambda_{i,t}^u$. Thus, one can re-express $w_{i,t+1} = \varepsilon_{i,t+1} + \lambda_{i,t}^u$ with $\varepsilon_{i,t+1} \backsim N(0,1)$ and then incorporate the $\lambda_{i,t}^u$ term into the conditional mean dynamics of the system. As shown in equation (21) in A, these prices may depend on $v_t^{0.5}$ (in the case of $w_{x,t}$ and $w_{g,t}$) or are simply constant ($w_{v,t}$). Thus we can obtain the following convenient representation of the worst case[11]

[10]The analytic expressions for the distorted innovation means are included in A.10 in equation (22).

[11]Similar derivations are used in Drechsler and Yaron (2011) where they interpret dynamics under the risk neutral measure.

$$
\begin{aligned}
\log g_{t+1} &= G_0 + x_t + \chi_{g,v} v_t + v_t^{0.5} \varepsilon_{g,t+1} \\
\log g_{d,t+1} &= G_{d,0} + \phi x_t + \chi_{gd,v} v_t + \varphi_d v_t^{0.5} \varepsilon_{d,t+1} \\
x_{t+1} &= \rho x_t + \chi_{x,v} v_t + \varphi_x v_t^{0.5} \varepsilon_{x,t+1} \\
v_{t+1} &= (1 - \varphi_v)\, \tilde{v} + \varphi_v v_t + \sigma_v \varepsilon_{v,t+1} \\
\\
\varepsilon_{i,t+1} &\backsim N(0,1) \;\; for \; i \in \{g, x, v, d\}
\end{aligned}
$$

where $\tilde{v} \equiv \bar{v} + \frac{\sigma_v \lambda_v^u}{1-\varphi_v}$. Notable are the additional terms in $v_t$ in the cash-flow equations. Consumption growth, dividend growth and the Long-run Risk component acquire contributions to their conditional means that depend on the level of volatility. Since $\tilde{v} > \bar{v}$ the average level of volatility will also be higher.

We can also derive a useful approximate homoscedastic VAR representation of both the benchmark and worst case models. This representation is derived by substituting $\bar{v}$ or $\tilde{v}$ for all terms in $v_t^{0.5}$ but leaving terms linear in $v_t$ unchanged. Thus, in this representation, we have homoscedastic systems but with the $v_t$ term solely acting like a component that influences the conditional mean.

Under these approximations we have two first order vector autoregressive (VAR) representations for a state $\hat{s}_{t+1} \equiv [x_t, v_t, \varepsilon_{d,t+1}, \varepsilon_{g,t+1}, \varepsilon_{x,t+1}, \varepsilon_{v,t+1}]$. These VARs preserve the 'first moment' conditional dynamics of the underlying heteroscedastic systems (benchmark and worst case) and also the unconditional second moments.

### 4.1 A simple illustrative example

We begin our quantitative analysis by estimating a simple model for consumption growth using quarterly per capita nondurable plus services consumption data for the U.S. from 1948:Q2 to 2013:Q4. The estimated model abstracts from the Long-run Risk component and so asserts that consumption growth is heteroscedastic white noise. We use the posterior means from our Bayesian estimation for our parameterization of the endowment process since this example is simply meant to be illustrative. The parameterization is given in table 1 and implies substantial variation in volatility over time.

We calibrate $\beta$ and $\theta$, to attain an annual risk free rate target of 2.59% and various values of the unconditional quarterly market price of risk.[12] The values for the market price of risk

[12] The market price of risk is taken as the ratio of the unconditional standard deviation of the stochastic discount factor to its unconditional mean. The targeted value of the risk free rate is equal to the median of the short sample simulations carried out by Beeler and Campbell (2012) under the Bansal and Yaron (2004)

that we consider are 0.25, 0.375 and 0.5.

As is well known, being able to attain an elevated market price of risk without an excessive risk free rate is essentially impossible if one uses standard, time separable expected utility. Under Epstein-Zin preferences, however, breaking the equivalence of the elasticity of substitution (EIS) and the inverse of the coefficient of relative risk aversion allows these two goals to be satisfied simultaneously. As noted in Tallarini (2000) and Barillas, Hansen, and Sargent (2009), this allows greater success in approaching the Hansen-Jaggannathan bounds as one increases the coefficient of risk aversion.

Thus, in fixing $\theta$ we implicitly fix an associated coefficient of relative risk aversion under the alternative, Epstein-Zin interpretation of our preferences. In table 2 we list our preference parameterizations associated with the unconditional market prices of risk considered. For interpretability, we also list detection error probabilities and (although not the natural interpretation in this paper) the associated risk aversion coefficient.

The DEPs are listed for two lengths of short sample simulations. Recall that to calculate DEPs one repeatedly simulates short samples under the benchmark and worst case and applies likelihood ratio tests between the two. Clearly, the longer is the 'short' sample, the more detectable the models will be and the lower will be the DEP.

It is not clear how to choose the sample length or, indeed, how to interpret detection error probabilities. With regard to sample lengths, we report detection error probabilities associated with two reasonably long sample sizes of 100 and 250 quarters, meant to be illustrative of the sort of sample lengths commonly employed in macro-finance applications. One guide to selecting these lengths is to align it with the sample we used to estimate the benchmark - which would favor the longer of our two samples - since implicitly in our analysis our agent might have used such a sample in arriving at her benchmark. Nevertheless, one might wish to envisage an agent who has used post-war data to estimate a model but fears that there may have been a structural break in more recent times. One need only look at the recent Japanese experience or the active debate in the U.S. over secular stagnation to note that fears of model misspecification might be associated with a concern that historical data may be rendered unreliable. In this case, one might want to consider model comparisons over shorter periods than those used to estimate the benchmark.

Anderson, Hansen, and Sargent (2003) suggest that detection error probabilities of approximately 0.1 might be thought to convey a plausible degree of ambiguity aversion.[13] However, there is no fixed consensus in this regard and there are no easily interpretable thought exper-

calibration. As in Beeler and Campbell (2012) we calculate the annual risk free rate by rolling over the one-period risk free rate, in this case from a quarterly frequency.

[13]Anderson, Hansen, and Sargent (2003) envisage a data set of length 200 quarters in their exercises.

iments such as the Arrow-Pratt gamble experiments to guide us, as they do for interpretation of risk aversion coefficients.

As one can see from table 2 one would require an extremely implausible degree of risk aversion to attain the desired prices of risk. In comparison, the detection error probabilities seem perhaps more reasonable, especially for a MPR of 0.25. This is the essence of the story told in Barillas, Hansen, and Sargent (2009) who argue that a little robustness can substitute for a lot of risk aversion. Nevertheless, the detection error probabilities in this simple example are rather low and suggestive of an extreme degree of robustness, even if it is less extreme than the implied risk aversion. Despite this we will analyze the properties of the worst case under these calibrations as they provide useful qualitative insight.

Taking high, medium and low values of $v_t$ to be the theoretical (not allowing for censoring) $20^{th}$, $50^{th}$ and $80^{th}$ unconditional percentiles under the benchmark, we can calculate the worst case innovation distributions. Given Normality and independence, together with the fact that the standard deviations are undistorted relative to the benchmark, these distributions are characterized by their means.

The mean of the endowment innovation is distorted downwards, while the mean of the volatility innovation is distorted upwards. These patterns are to be expected given the undesirability of low consumption growth and high volatility. The pessimistic shift to the mean of $w_{g,t+1}$ is more intense, the higher is volatility, as listed in table 3. Intuitively, as the agent is more exposed to misspecifications in her model for growth (because $w_{g,t+1}$ is pre-multiplied by a larger volatility term) she envisages greater distortions to the benchmark model as an artifact of her robust decision problem.

In low volatility times, the robust agent is informed by a conditional distribution very much like that of the benchmark and, in this sense, is close to an expected utility agent. Although the agent doubts her model at all times, the *manifestation* of these doubts varies. This is perhaps a partial response to oft-expressed concerns that models of ambiguity are counterfactual in their prediction of pessimistic behavior. The argument often is that in the real world people do not typically appear pessimistic. However, we see here that robustness is quite capable of generating variations in apparent pessimism and, as in the real world, pessimism can be quite extreme in occasional periods of unusually high volatility.

Associated with fluctuations in the worst case conditional distributions are, naturally, fluctuations in the conditional market price of risk. In table 4 we list the conditional MPR and the corresponding object in the absence of robustness. Both the level and variability of the conditional MPR derive almost entirely from the 'uncertainty component', $\Lambda^u_{t+1}$, of the stochastic discount factor.[14] The remaining columns correspond to our aforementioned

[14]Thus, as noted by Barillas, Hansen, and Sargent (2009), the term 'Market Price of Risk' is something of

calibrations to hit certain *unconditional* market prices of risk.

Intuitively, in high volatility periods the worst case conditional distribution changes: in addition to the increased volatility, the relative likelihood of a low consumption realization increases and the mean of the consumption growth series falls, pushing up the conditional market price of risk, evaluated under the benchmark measure.

### 4.2 Fears of Long-run Risk and 'Disasters'

The properties of the worst case conditional distribution discussed above are interesting but also, at least qualitatively, to be expected. Less obvious, however, are the implications of our model for the worst case distribution over sequences, induced by iterating on these distorted conditional distributions and the law of motion for the economy. It is these distributions that can give us insight into the sort of data generating processes that would concern an agent fearful of model misspecification in an ambiguous environment.

In table 5 we show various unconditional moments of the log consumption growth process at yearly aggregation (with the observation intervals at yearly intervals, also). As expected, the worst case features lower growth and elevated volatility on average. Skewness in consumption growth is enhanced, implying a greater concern for 'disasters'. Note, however, that this concept of disasters is unconditional, whereas our *conditional* distributions at the raw simulation frequency are Gaussian and thus symmetric.[15] It is arguably in a conditional sense that people typically think of disasters, and we will address this in section 5 below, by restoring Poisson-mixture Normal components to our analysis.

We also calculate autocorrelations at annual frequencies, listed in table 6. These illustrate additional persistence under the worst case, which indicates the agent's aversion to models with a small but persistent long run component in consumption growth - the hallmark of the canonical Long-run Risk framework. In fact, the worst case model at quarterly frequency inherits an autoregressive root equal to the persistence of the volatility process. The worst case quarterly autocorrelations are small but due to their persistence, they result in non-trivial autocorrelations at yearly intervals and substantially increased power at low frequencies.[16]

The volatility component $v_t$ is persistent and, thus, the conditional mean of consumption

a misnomer in this context, with 'Market Price of Uncertainty' perhaps more appropriate term.

[15]The aggregated annual series will exhibit some negative skewness as will consumption growth over any horizon greater than one quarter.

[16]Note that the number of lags correspond to quarters or years, depending on the column considered. Note also that the yearly moments are based not only on aggregating up from a quarterly model, but also from implicitly switching to yearly observation, rather than quarterly observation of year on year growth rates. Hence, given the white noise nature of the benchmark at quarterly frequency, the yearly autocorrelations are also zero. If we had used year on year growth rates at quarterly frequency, we would have had non-zero autocorrelations.

growth inherits this persistence under the worst case. This is transparent in the worst case system, which makes clear that a scaled version of $v_t$ is essentially acting as a Long-run Risk component. In this sense the agent prices assets as if there is Long-run Risk, even though there is none in the true model. The manner in which the persistent component enters the model (*via* the concerns of the agent) is highly restricted however, since the persistence of the component is pinned down by the volatility process and its size by the degree of ambiguity aversion.

The worst case system also makes clear the nature of the unconditional skewness in consumption growth under the worst case. $v_{t+1}$ drives the conditional mean and conditional volatility. Thus, on average, periods of low growth will be associated with higher volatility that raises the probability of consumption growth being thrown even further into the left tail, even while *conditional* skewness in $t+1$, given information in $t$, remains zero.

We note that the magnitude of the autocorrelations and skew are small and, beyond emphasizing the qualitative insights, we also offer an extension of this basic model that enhances this autocorrelation, by introducing 'vol-in-vol'. Campbell, Giglio, Polk, and Turley (2012) have emphasized the importance of allowing for fluctuations in the volatility of volatility and we introduce it by amending our volatility law of motion under the benchmark to become

$$v_{t+1} = (1 - \varphi_v)\,\bar{v} + \varphi_v v_t + \sigma_v v_t^{0.5} w_{v,t+1}$$

where we set $\sigma_v = \frac{\sigma_v^{old}}{\bar{v}^{0.5}}$ where $\sigma_v^{old}$ is from the constant vol-of-vol parameterization.

As shown in C, this will again imply a worst case system in which $v_t$ enters as a persistent conditional mean component in consumption growth and the LRR equations. The unconditional mean of $v_t$ is, as before, elevated. However, the worst case also features a $v_t$ component that retains an AR(1) structure but with a larger autoregressive coefficient, $\tilde{\varphi}_v > \varphi_v$. Thus, the agent will not only behave as if there is a long run component in the cashflow growth variables but that the volatility component undergoes more persistent swings.

$$v_{t+1} \;=\; (1 - \varphi_v)\,\tilde{v} + \tilde{\varphi}_v v_t + \sigma_v v_t^{0.5} \varepsilon_{t+1}$$

In table 6, we include results and calibration of the vol-in-vol case. Generally the distortions are more exaggerated then before so that, not only are our results robust to the presence of vol-in-vol, but are enhanced. Indeed, it seems an interesting avenue for further exploration of how providing more structure to the benchmark allows the agent to explore her vulnerabilities in more complicated ways.

### 4.3 Introducing (some) Long-Run Risk in the benchmark

In the previous section we illustrated the essential intuition behind our results in a highly restricted framework with white noise consumption growth. We now relax the specification of the benchmark along several directions. First we introduce elements of the Bansal and Yaron (2004) and Bansal, Kiku, and Yaron (2012) frameworks. This is because we do not want to argue that there is *no* Long-run Risk in reality. Instead we suggest there may be less than under the standard LRR specifications and that one can trade off LRR in the mind of a robust agent, against a more moderate parameterization of LRR under the benchmark.[17]

Bansal and Yaron set $\{\rho, \varphi_x\} = \{0.979, 0.044\}$. This parameterization implies short lag autocorrelations reasonably close to those observed in the data, but it has been argued that the model exaggerates persistence, as captured by variance ratios at longer lags (Beeler and Campbell (2012)) and other transformations of autocorrelations. Pursuing this further, Dew-Becker (2013) has argued that an important measure of risk in the long run is what is typically termed the long run standard deviation (LRSD) of consumption growth:

$$\begin{aligned} LRSD &\equiv \sigma\left(\Delta E_{t+1}\sum_{j=0}^{\infty}\log g_{t+1+j}\right) \\ &= \sqrt{\sum_{j=-\infty}^{\infty}\gamma_j} \\ \gamma_j &\equiv Cov\left(\log g_t, \log g_{t-j}\right) \end{aligned}$$

The LRSD is equal to the square root of the sum of all autocovariances and thus is the square root of the spectral density of consumption growth at frequency 0.[18] Beyond this intuitive connection to the long run, and the fact that it encodes all autocovariances, rather than an arbitrarily selected subset, Dew-Becker also shows that this object features prominently in determining the maximal Sharpe ratio in standard calibrations of Epstein-Zin preferences. Thus, we take the LRSD as an index of Long-run Risk in our analysis. Where relevant we will scale the LRSD by the unconditional standard deviation (USD) of consumption, so as to focus on persistence and retain comparability as the USD varies, as it does when comparing benchmark and worst case, for example.

Estimates of the LRSD tend to fall substantially below those implied by the BY calibra-

[17] See Schorfheide, Song, and Yaron (2014) for recent work that adopts a more elaborate mixed-frequency approach to estimating LRR systems which find evidence in favor of a long run component when using only cashflows in estimation - although the evidence appears stronger when incorporating asset pricing data.

[18] The spectral density at angular frequency $\omega$ is defined as $f(\omega) \equiv \sum_{j=-\infty}^{\infty}\gamma_j \cos(\omega j)$.

tion. In figure 1 we plot the LRSD for different values of $\rho$, while adjusting $\varphi_x$ to maintain the unconditional variance of $x_t$. On this plot, which is in terms of annualized percent, we also overlay horizontal lines at the BY-implied value, Dew-Becker's point estimate and his 90% and 95% confidence intervals.[19] We choose to set $\rho$ equal to the 90% boundary, which is 0.929 rather than the 0.979 chosen by Bansal and Yaron.[20] The half-lives of shocks to the LRR component in these two cases are approximately $2\frac{3}{4}$ years in the BY case and $\frac{3}{4}$ of a year in our adjusted case, so the reduction of Long-run Risk in the benchmark is economically significant.

Under the worst case we observe somewhat higher autocorrelations at short lags, relative to the benchmark. However, again, the most important aspect of their behavior is that they remain small but positive for a large number of lags - dying out at a rate approximately equal to the persistence of volatility. In contrast, the autocorrelations under the benchmark go to zero much more quickly. This, and the fact that the unconditional standard deviation is also inflated under the worst case, implies a substantially higher LRSD than the under benchmark, as shown in table 11, in annualized percent. Indeed, the LRSD is in line with the original BY benchmark, even after scaling by the unconditional standard deviation.

We adjust Bansal and Yaron's parameterization of the volatility process, inspired by their most recent work in Bansal, Kiku, and Yaron (2012), which increases the persistence of volatility, $\varphi_v$, substantially. This focuses our analysis on the source of worst case persistence in our model, which is the interaction of stochastic volatility with ambiguity. Nevertheless, to discipline our approach, we maintain the same unconditional variance of $v_t$ as in Bansal and Yaron (2004), by reducing the standard deviation of its innovations, $\sigma_v$.

Our parameterization is listed in tables 8 and 9. We set $\alpha$ and $\beta$ to calibrate according to the risk free rate and unconditional market prices of risk, as before. We use $\phi$ and $\varphi_d$ to set the equity premium and volatility of dividends. A collection of baseline moments are listed in table 10, where we compare to numbers reported in Beeler and Campbell (2012), calculated from real world data and from the Bansal and Yaron (2004) model.[21] We have chosen a parameterization yielding numbers that are broadly comparable to those of Bansal and Yaron (2004). Like them, our equity premium is somewhat low, relative to the data and the volatility of dividend growth is also counterfactually low. The implied detection error

[19]These are from looped small sample analysis of a novel estimator proposed in Dew-Becker (2013), though other estimators based on fitting low order ARMAs imply typically even lower point estimates and tighter intervals.

[20]Note that we cannot reach Dew-Becker's point estimate by varying $\rho$ while maintaining the unconditional variance of the long run component.

[21]The targeted risk free rate and MPR values were unconditional 'ergodic' moments. Note that this is in contrast to table 10 where we report medians of looped short sample moments. There is a very small gap between the median risk free rate and the unconditional ergodic mean.

probabilities using 100 and 250 period-long samples are 10.2% and 2.4% respectively for a MPR of 0.25 and 2.6% and 0.1% for a MPR of 0.375.

Inspired by Beeler and Campbell (2012) we report additional moments that they identify as important checks on the LRR model. In tables 12 and 13 we report variance ratios for consumption and dividend growth and $R^2$ from predictability regressions of excess returns, consumption growth and dividend growth on the price-dividend ratio. Reflecting the excessive persistence of the baseline Bansal and Yaron (2004) model the variance ratios of consumption growth at longer horizons are counterfactually high under their calibration and increase dramatically, relative to shorter horizons. Unsurprisingly, with our adjusted $\rho$ our model performs better along this dimension, even if the problem is not entirely solved. We are less successful in terms of dividend growth.

Turning to the predictability evidence, like the LRR model, we do poorly in terms of the predictability of excess returns. This is despite the fact that, qualitatively, the model will generate predictability arising from the additional cashflow persistence under the worst case model. We naturally do better in terms of predictability of consumption and dividend growth, relative to the LRR model - bringing the degree of predictability down towards observed moments. As to the predictability of volatility, we appear to improve somewhat on BY in some dimensions and do worse in others. Both models struggle substantially with the predictability of consumption volatility.

As shown in table 14 we have a downward sloping real term structure, as in the baseline LRR model. The slope becomes more severe and more comparable to Bansal and Yaron (2004) for the higher MPR calibration considered. This is a direct implication of the positive autocorrelation in consumption growth under the benchmark, which is exaggerated in the mind of the agent. While there is some debate regarding the slope of the *real* term structure (see Swanson (2014), for example) this appears to be counterfactual - especially to the degree seen in the high MPR calibration of our model. Nevertheless, it is no worse than the baseline LRR model.

Interestingly, our model performs better than the baseline LRR setup in terms of 'EIS regressions' of the sort discussed in Hall (1988), Campbell and Mankiw (1989) and Beeler and Campbell (2012). Table 15 reports implied estimates of the elasticity of intertemporal substitution (EIS) from regressions of the risk free rate on consumption growth (and *vice versa*). Although the estimates are still notably above those obtained from similar regressions on real-world data, we obtain substantially lower estimates than those estimated from data generated by the LRR model. The intuition for this is clearest in the case of using the risk free rate to predict consumption growth. The exaggerated persistence of the worst case model leads agents to behave as if they overextrapolate shocks to consumption growth. Given the consumption smoothing motive and log utility of the agent the risk free rate will vary in equilibrium to a greater degree than it otherwise would. However, under the maintained assumption that the true data generating process is the benchmark the additional movements in the interest rate do not reflect additional predictability, so regressions of consumption growth on the riskless rate will yield a lower coefficient.[22]

These results suggest that worst case fears of higher volatility and greater persistence in the mind of an agent can substitute for those properties in the actual data generating process. Failing to allow for this additional Long-run Risk in the mind of the agent, while relying on prices and behavior, could conceivably lead an econometrician astray as these properties are encoded in how agents' evaluate risky payoffs even if they are not present in reality. In addition, trading off LRR in the economy against worst case fears of LRR in the mind of the agent can help improve the model's ability to match certain important moments where the baseline LRR model gives counterfactual predictions.

### 4.4 Discussion: Comparison with existing literature

It is worth contrasting the nature of the time variation in our agent's 'pessimism' with that discussed in Drechsler (2013).[23] In our framework it is not the case that as volatility ($v_t$) varies over time our agent's *uncertainty* varies. Although we are working with a recursive formulation of the robust problem, the sequence problem (2) makes clear that the agent is uncertain about the worst case distribution over sequences implied by her benchmark and that this degree of uncertainty is fixed. Since we are representing the worst case distribution over sequences in terms of worst case conditional distributions (combined with the law of motion (1)), the manifestation of this uncertainty in terms of distorted conditional means varies over time with $v_t$. However, this variation is implicit in a fixed worst case joint distribution over sequences.

In Drechsler (2013), 'uncertainty' is allowed to vary over time with volatility in the sense that the penalty for distortions to benchmark conditional distributions is linked to the level of volatility - when volatility is high, the marginal penalization for entropic deviation is reduced *in addition to* the agent's exposure to misspecification being greater. Although it seems very plausible that times of high volatility are somehow associated with higher uncertainty, this connection must be modeled carefully if one is to claim a fully structural interpretation of

[22]Bidder and Dew-Becker (2014) generate low EIS estimates using a similar mechanism but a different concept of ambiguity aversion.

[23]See also Kleshchelski and Vincent (2008) and Xu, Wu, and Li (2010) who analyze the effects of stochastic volatility in consumption on asset prices in a continuous time setting with a robust agent. Our model is entirely in discrete time. The tools we present can be used in many discrete time representative agent frameworks, a standard workhorse of modern macroeconomics.

the separate influence of these two factors. Indeed, the ambiguity literature is in some sense defined by the explicitly separate treatment of 'risk' and 'uncertainty'.

Although Drechsler (2013) likely is a reduced form for a model where volatility and uncertainty are mutually dependent but distinct, we prefer to keep them explicitly separate and independent in this work. The restriction that uncertainty be fixed implies that, as will be discussed below, it is more difficult to attain empirical success, for which Drechsler (2013) is notable. Beyond that, Drechsler adds further structure to the worst case distribution in partitioning components of the state into elements whose dynamics are uncertain and those whose dynamics are fully trusted. Some additional restrictions are also placed on the nature of the jump perturbations. In contrast, we only have one parameter, $\theta$ that controls the nature of the worst case in relation to the benchmark. Once one allows the agent to envisage (absolutely continuous) distributions that deviate from the benchmark, penalizing the entropic distance according to $\theta$, our control over the the worst case is gone - $\theta$ has been 'used up'.[24]

It is also useful to compare our work with that of Barillas, Hansen, and Sargent (2009) and Hansen and Sargent (2010b). We extend Barillas, Hansen, and Sargent (2009) in terms of methodology (our algorithms for drawing from the worst case) and in terms of introducing stochastic volatility. In their model the homoscedasticity they assume for innovations under the benchmark leads to a very restricted worst case - an unconditional mean shift in the innovation distributions. It is perhaps undesirable that omitting a fairly uncontroversial phenomenon like stochastic volatility can have such an important impact on the qualitative properties of the worst case, given that robustness is largely about doubting the specification of the benchmark model. The heteroscedastic setup we use shows how a robust agent will distort conditional variances if given the opportunity.

Hansen and Sargent (2010) set up a model in which the agent focuses on two possible models - an iid consumption growth model and a LRR consumption growth model - and her fears of misspecification are represented with distorted filtered probabilities of which (latent) model is the truth and, assuming a given model as a benchmark, how its dynamics might be misspecified. We do not consider latency or robust filtering in our paper. It is important that in our baseline heteroscedastic white noise framework we do not actually posit LRR as part of the benchmark as Hansen and Sargent do. It emerges purely in the mind of the agent, as

[24]Of course, the additional restrictions that Drecshler imposes do not invalidate his approach - they are both conceptually reasonable and helpful empirically - but for our purposes we wish to remain entirely with the tightly parameterized unstructured uncertainty of Hansen and Sargent (2008). For alternative approaches to how additional restrictions can be added to robust control problems, see Hansen and Sargent (2010b) where two entropic distances are used to envisage separate uncertainty attitudes towards filtering and dynamics or Petersen, James, and Dupuis (2000) where one asserts moment conditions that the worst case must satisfy.

captured in the autocorrelation of consumption growth under the worst case.

### 4.5 Discussion: Pricing of risk and uncertainty

It is illustrative to examine the approximate solution for the scaled (by consumption) value function. The coefficients on the constant, $x_t$ and $v_t$ in the exponentially affine approximation are given by

$$\begin{aligned} F_0 &\equiv \frac{\beta}{1-\beta}\left(G_0 + F_2\left(1-\varphi_v\right) + \frac{\alpha}{2}F_2^2\sigma_v^2\right) \\ F_1 &\equiv \frac{\beta}{1-\beta\rho} \\ F_2 &\equiv \frac{\alpha}{2}\frac{\beta}{1-\beta\varphi_v}\left(1+F_1^2\varphi_x^2\right) \end{aligned}$$

Thus, there is no concern for time variation in $v_t$ *per se* in the absence of robustness ($\alpha = 0$). That is, an agent who fully trusts her model and therefore has time separable expected utility with a log period payoff, regards her welfare as unaffected by fluctuations in the volatility state around its unconditional mean, $\bar{v}$.

In the presence of robustness the influence of the stochastic nature of $v_t$ is manifested in $F_0$ and $F_2$. Even if $v_t$ featured no persistence ($\varphi_v = 0$) the value function would still load on $v_t$. This is because recursive representation of the worst case entails a conditional mean shift in $w_{g,t+1}$ and, thus, $v_t$ is a relevant state given the fact that it pre-multiplies $w_{g,t+1}$ and thus transmits the pessimistic mean distortion to (worst case) expected consumption growth. The relevance of $v_t$ is clearly greater if it features persistence ($\varphi_v > 0$) and if the process to which it applies itself features persistence under the benchmark (as when $\varphi_x > 0$).[25]

Intuitively, the presence of $v_t$ as a welfare-relevant state reflects how, from the perspective of the time-0 sequence problem, certain path realizations expose the agent to more painful misspecifications and, therefore, feature more twisting by the minimizing Martingale. Thus $v_t$ is vital in enabling a recursive representation. Once the agent has reached a 'node' (or history) featuring high volatility it signifies that the worst case joint distribution over sequences requires substantial distortions to the conditional distributions stemming from that node.

Now, considering the stochastic discount factor we see that, even within a fully trusted model, $v_t$ controls compensation for exposure to $w_{g,t+1}$ via $\lambda^r_{g,t}$. However, there are additional effects in our case because $v_t$ enhances exposure to misspecifications represented by mean shifts to $w_{g,t+1}$ ($\lambda^u_{g,t} \equiv \alpha v_t^{0.5}$).[26] Furthermore, recalling the presence of $v_t$ in the value function

[25]In addition to the time varying influence of $v_t$ on the agent's welfare we also observe an influence on the steady state *via* the third term in the parentheses in the definition of $F_0$.

[26]Similar intuition applies to mean shifts in $w_{x,t+1}$ if there is a non-degenerate LRR component under the

discussed above, $w_{v,t+1}$ has the appearance of a 'risk' being priced ($\lambda_v^u \equiv \alpha F_2 \sigma_v$) because the innovation to $v_{t+1}$ will determine how pessimistic a change in measure the agent will be operating under from the next period forwards.

From this perspective, the algebraic equivalence between an Epstein-Zin agent and the robust agent is natural. The Epstein-Zin agent wants to hedge against shocks to future lifetime utility. The robust agent wants to hedge exposure to realizations that make her particularly vulnerable to misspecifications that are damaging for future lifetime utility.

Assuming that the data generating process is the benchmark model a 'wedge' is introduced that lowers the prices of assets whose payoffs depend on $w_{g,t}$ and $v_t$ (positively and negatively, respectively), relative to the prices that would prevail in the absence of model uncertainty. 'Risks' that are not priced by an expected log utility agent will appear to be priced by the robust agent, as in the case of $w_{v,t+1}$, because they are serving a different purpose and are not risks in the same sense as under the Epstein-Zin interpretation, but are reflecting a fear of misspecification.

### 4.6 Discussion: Relationship with Epstein-Zin

Given our decision to assert that the benchmark is the true data generating process, the likelihood function of the robustness model is the same as that of the associated reinterpreted Epstein-Zin/unity EIS model. Thus, our approach is nested by a model in which the EIS of the Epstein-Zin agent is unrestricted, as in Bansal and Yaron (2004). Thus, a larger set of moment restrictions can be satisfied by the Epstein-Zin model with unrestricted EIS than by our model. Indeed, if one were to estimate the unrestricted model on data from our benchmark, featuring robust agents, it should theoretically recover the log utility parameterization of preferences.

Nevertheless, if one imposes parameter restrictions on the economic environment there is scope for our more restricted model to outperform along dimensions of interest. Most relevant in our case are restrictions implying the degree of Long-run Risk chosen by Bansal and Yaron (2004). Our story is that this degree of Long-run Risk seems excessively high relative to what consumption data *alone* would imply, as argued powerfully in Beeler and Campbell (2012). A model featuring robustness is preferable in asserting a smaller degree of Long-run Risk while generating elevated and varying prices of risk *via* Long-run Risk in the mind of the agent.

The unity EIS restriction is required for homotheticity of multiplier preferences, which in turn allows us to obtain a stationary scaled solution for the value function (and other objects) in an environment of non-stationary consumption. An interesting avenue would be to somehow relax this restriction but this would be well beyond the scope of this paper (see

benchmark.

Meyer-Gohde (2015) for work in this direction). We believe that our intuition of a concern for persistent components in consumption growth would be preserved in such a generalized framework as it is natural that long run aspects of models are particularly prone to ambiguity as well as having important implications for welfare (see, also, Bidder and Dew-Becker (2014) and Hansen and Sargent (2010b) on these points).

## 5 Results: A robust perspective on rare disasters

In section 4.2 we showed that our agent was informed by a worst case model that implied unconditional skewness in consumption growth rates. We interpreted this as a fear of disasters. In this section we pursue a more natural interpretation of disasters, such that the benchmark model features conditional skewness in the form of a 'jump'. In the language of section 2, we reintroduce $w_{z,t+1}$, the Poisson-mixture of Normals.

Much as we developed our robust perspective of Long-Run Risk around a benchmark that borrowed from Bansal and Yaron (2004), we will here draw upon the framework laid out in Wachter (2013).[27] We will abstract from the presence of stochastic volatility and Long-Run Risk. We omit $w_{d,t}$ and, following Wachter, take dividends to be levered consumption ($D_t = C_t^{\phi}$). We will allow for both constant and time varying intensity, $h_t$. Thus, our benchmark model is again a special case of system (1), given by

$$\begin{aligned} \log g_{t+1} &= G_0 + w_{z,t+1} + \bar{v}^{0.5} w_{g,t+1} \\ \log g_{d,t+1} &= \phi \log g_{t+1} \\ h_{t+1} &= (1-\varphi_h)\bar{h} + \varphi_h h_t + w_{h,t+1} \end{aligned}$$

We set $\bar{h}$ to imply a steady state expected number of jumps in 100 years to be 2.5, informed by the calibration used by Wachter (2013). The persistence parameter, $\varphi_h$ is also based on Wachter's parameterization.[28] We choose $\delta_h$ and $c_h$ to ensure the aforementioned steady state jump intensity while also allowing substantial volatility in equity returns, as we will discuss below.[29] The mean of $w_{z,t+1}$ given a single jump, $\theta$, is set to $-0.3$ and $\delta$ is set to 0.1. This parameterization was used in Backus, Chernov, and Zin (2014) as an approximation to the multinomial distribution for consumption declines in the case of disasters, used by Wachter

[27] Wachter's model is in fact deeply connected to ours in that she asserts an EIS of unity within an Epstein-Zin framework. Thus, we provide an interesting robustness reinterpretation of her setup.

[28] $\varphi_h = 1 - \kappa/12$ where $\kappa$ is the persistence parameters in Wachter's continuous time intensity process.

[29] Backus, Chernov, and Zin (2014) map Wachter's calibration into an approximate discrete time version using an AR process for intensity but dramatically reduce the volatility and persistence of the intensity process to avoid unacceptably high probability of the process going negative. Given our ARG specification of $h_t$ we can handle a more volatile intensity process.

(2013) and Barro and Ursua (2008).

The remaining consumption growth parameters are listed in table 16. In the absence of disasters, the trend growth and volatility of the Gaussian innovation are calibrated to yield mean and standard deviation of annual consumption growth of 1.80% and 1.99% respectively.[30] In addition, as in Wachter (2013) we allow for a 40% probability of default on government debt in the case of a disaster, in which case the proportional reduction in the promised face payoff is equal to the realized disaster size, $w_{z,t}$.[31]

Table 18 lists our calibration in the constant intensity case, where we choose $\alpha$ and $\beta$ again to hit unconditional $MPR$ targets as well as a desired expected return on government debt. We ensure an average return of 1.36% conditional on no disasters occurring and approximately 1.06% unconditionally, similarly to Wachter (2013).

Referring to table 18 we note that the detection error probabilities are dramatically higher than in our Long-run Risk frameworks. Even for an MPR of 0.5, the detection error probability of approximately 9% after 250 quarters is not wildly implausible. Intuitively, this captures the fact that rare disasters are a powerful phenomenon to interact with ambiguity. Under the benchmark the expected number of jumps in a century is 2.5 while for an MPR of 0.50 the number is approximately 10, as shown in the table. This would be highly undesirable for the agent given the likely drop in consumption in the case of a disaster, but the frequency is not so much higher that she can distinguish confidently between the benchmark and worst case if, say, she observes two jumps in 25 years (corresponding to our shorter DEP sample period).

The worst case distribution concentrates the damage of a misspecification in a way that is efficient, in the sense of trading the pain of the distortion relative to the benchmark against the offsetting entropy penalty. Rare events are ideal for this since they are sufficiently infrequent to be difficult to characterize, as well as being very painful. Since the presence of jumps allows us to generate a high MPR with far lower $\alpha$ than in the purely Gaussian cases considered above, there is a concomitant reduction in the distortion to the Gaussian shock (the mean shift to $w_{g,t+1}$ is $\alpha\bar{v}^{0.5}$). In our LRR setups above the DEPs were rather low because the worst case represented the agent's fear of misspecification by large mean shifts to shocks observed period after period, thereby substantially reducing average growth over any short span of time. This rendered the benchmark and worst case easily distinguishable even in

[30] We also use a parameterization that targets a standard deviation of 1.04 as in our earlier heteroscedastic white noise example. Hence the two values of $\bar{v}$ in table 16 and the two columns in tables 17 and 18. In addition, we consider a value of the average return on government debt equal to 2.59% as in our earlier discussion of the LRR setups, with results listed in table 17. For neither of these alternative calibrations are there qualitatively different implications and the quantitative differences are also minor. For the ARG case we only report for one baseline calibration, for comparability with Wachter (2013).

[31] In appendix B.4 we derive expressions for the face and average returns on one period government debt.

small samples. In the case of jumps, there is a much smaller trend growth reduction - the worst case instead emphasizes dramatic but occasional declines.

Finally, in table 19 we list details of our calibration in the time varying intensity case, where we have chosen $\alpha$ and $\beta$ to hit an unconditional (allowing for disaster realizations) $MPR$ target of 0.25 and the same targeted value of the mean return on government debt (conditional on no disasters) of 1.36% as in the constant intensity case. Detection error probabilities for 100 and 250 quarters are 10.0% and 4.1%, respectively, which are reasonable. When evaluated at the worst case unconditional mean of $h_t$ the 'steady state' expected number of jumps in 100 years is 11.3 (again in comparison with 2.5 under the benchmark).

We report asset pricing moments, unconditionally and conditioning on the non-occurrence of disasters, which Wachter (2013) emphasizes for comparability with the post-war experience in which there have been no episodes of the sort of disaster described by Barro (2006). We employ a higher degree of leverage than Wachter (2013) and, like her, our model implies an equity premium somewhat on the high side, with an elevated Sharpe ratio. In addition, we obtain qualitatively similar results to Wachter in terms of predictability of excess returns, reported in table 19, although we also note the caveats associated with the calculation of $R^2$ values in this context of persistent regressors.

As in Wachter (2013), our equity returns are volatile and we can provide insight into the reasons for the volatility of returns by appealing to the worst case in the mind of the agent. In B.5.2 we show that under the worst case, $h_t$ follows a process of the form

$$\begin{aligned} h_{t+1} &= (1-\tilde{\varphi}_h)\,\tilde{h} + \tilde{\varphi}_h h_t + \tilde{w}_{h,t+1} \\ \tilde{\varphi}_h &\equiv \varphi_h + a_h \\ &= \frac{\varphi_h}{(1-\lambda_h^u c_h)^2} \\ \tilde{h} &\equiv \frac{\delta_h c_h + a_0}{1-\tilde{\varphi}_h} \end{aligned}$$

where $\tilde{w}_{h,t+1}$ is a Martingale difference sequence. Importantly, $\tilde{h} > \bar{h}$ and $1 > \tilde{\varphi}_h > \varphi_h$ so that, as before, the jump arrival rate is pessimistically distorted upwards but also, now that it is allowed to vary, is more persistent than under the benchmark. Indeed, under our calibration, the half life of a shock to $h_t$ is approximately 30 years under the worst case, in comparison with 8 years under the benchmark (see B.5.2 for the derivation).[32]

Again, we show in B.5.1 that $w_{z,t+1}$ is distributed under the worst case as a Poisson-mixture of normals with inflated arrival rate $\hat{h}_t \equiv h_t e^{\lambda_z^u \theta + \frac{(\lambda_z^u \delta)^2}{2}}$ and inflated mean on arrival,

[32]By half life we mean the $j$ required for the effect on the expectation in $t+j$ to decline to half of the effect on the expectation at the first horizon.

$\tilde{\theta} > \theta$. Consequently, when innovations to the intensity process strike, they are exaggerated and over-extrapolated (relative to the true model) when projected into the future under the worst case. Therefore, the agent acts as if facing a very volatile news-flow for future consumption and, thus, dividend growth. Figure 2 captures this by plotting the effect on expectations of future consumption growth (in annualized percentage terms) following an innovation to $w_{h,t}$, under the benchmark and worst case models.[33]

This extrapolative behavior leads to excess sensitivity and underpins volatility in returns and, apparently, much of the predictability evidence as well. In the mind of the agent, shocks to expected cashflows are substantial at long horizons. Thus, following a negative (positive) shock to $h_t$ she will be prepared to pay more (less) for the claim and, in equilibrium, the price dividend ratio should rise (fall). Nevertheless, since we assert the true model is the benchmark, an econometrician would attribute a measure of the variation in worst case *cashflow* expectations to variation in expected returns as the source of the volatility and predictability of returns as there is not as much predictability in cashflows as implied under the worst case.[34]

## 6 Conclusion

Attributing a desire for robustness to Knightian uncertainty leads a robust agent to act as if guided by a worst case distribution. We suggest novel methods of characterizing and drawing from this distribution and show that it will feature a small persistent component when the agent faces persistent heteroscedasticity and exaggerated jumps (in terms of frequency and size) if the agent perceives the true data generating process to exhibit some degree of jump risk. This allows the model to match important asset pricing facts without taking a firm stance on whether LRR exists or whether extant disaster calibrations are correct.

[33] As $h_{t-1}$ approaches zero, the ARG nature of the process implies that the standard deviation of $w_{h,t}$ declines - and higher moments also change - leading to the differences between the three panels of the figure.

[34] This is similar to the phenomena discussed in Bidder and Dew-Becker (2014) where what would appear to be extrapolative behavior and predictability in returns arises from an ambiguity averse Epstein-Zin agent using a worst case model for dynamics of cashflows. The concept of ambiguity aversion in Bidder and Dew-Becker (2014) differs from the Hansen-Sargent multiplier preferences approach adopted here, however.

## References


Anderson, E. W., L. P. Hansen, and T. J. Sargent (2003): "A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk, and Model Detection," *Journal of the European Economic Association*, 1(1), 68–123.

Backus, D., M. Chernov, and I. Martin (2011): "Disasters Implied by Equity Index Options," *The Journal of Finance*, 66(6), 1969–2012.

Backus, D., M. Chernov, and S. E. Zin (2014): "Sources of Entropy in Representative Agent Models," *The Journal of Finance*, 69(1), 51–99.

Bansal, R., D. Kiku, and A. Yaron (2012): "An Empirical Evaluation of the Long-Run Risks Model for Asset Prices," *Critical Finance Review*, 1(1), 183–221.

Bansal, R., and A. Yaron (2004): "Risks for the Long-Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59(4), 1481–1509.

Barillas, F., L. P. Hansen, and T. J. Sargent (2009): "Doubts or variability?," *Journal of Economic Theory*, 144(6), 2388–2418.

Barro, R. (2006): "Rare Disasters and Asset Markets in the Twentieth Century," *The Quarterly Journal of Economics*, 121(3), 823–866.

Barro, R. J., and J. F. Ursua (2008): "Consumption Disasters in the Twentieth Century," *American Economic Review*, 98(2), 58–63.

Beeler, J., and J. Y. Campbell (2012): "The Long-Run Risks Model and Aggregate Asset Prices: An Empirical Assessment," *Critical Finance Review*, 1(1), 141–182.

Bidder, R., and I. Dew-Becker (2014): "Long-run risk is the worst-case scenario: ambiguity aversion and non-parametric estimation of the endowment process," .

Campbell, J. Y., S. Giglio, C. Polk, and R. Turley (2012): "An Intertemporal CAPM with Stochastic Volatility," Working Paper 18411, National Bureau of Economic Research.

Campbell, J. Y., and G. Mankiw (1989): "Consumption, Income, and Interest Rates: Reinterpreting the Time Series Evidence," in *NBER Macroeconomics Annual: 1989*, ed. by O. Blanchard, and S. Fischer, pp. 185–216. MIT Press, Cambridge, MA.

Campbell, J. Y., and R. J. Shiller (1989): "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1(3), 195–228.

Collin-Dufresne, P., M. Johannes, and L. A. Lochstoer (2013): "Parameter Learning in General Equilibrium: The Asset Pricing Implications," Working paper.

Dew-Becker, I. (2013): "Estimates of the volatility of the permanent component of consumption and their implications for asset pricing," Duke University, Fuqua School of Business.

Dolmas, J. (2013): "Disastrous disappointments: asset-pricing with disaster risk and disappointment aversion," Discussion paper.

Drechsler, I. (2013): "Uncertainty, Time-Varying Fear, and Asset Prices," *The Journal of Finance*, 68(5), 1843–1889.

Drechsler, I., and A. Yaron (2011): "What's Vol Got to Do with It," *Review of Financial Studies*, 24(1), 1–45.

Ellsberg, D. (1961): "Risk, Ambiguity and the Savage Axioms," *Quarterly Journal of Economics*, 75(4), 643–669.

Gabaix, X. (2012): "Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance," *The Quarterly Journal of Economics*, 127(2), 645–700.

Gourio, F. (2012): "Disaster Risk and Business Cycles," *American Economic Review*, 102(6), 2734–2766.

Hall, R. E. (1988): "Intertemporal Substitution in Consumption," *Journal of Political Economy*, 96(2), 339–57.

Hansen, L. P., J. C. Heaton, and N. Li (2008): "Consumption Strikes Back? Measuring Long-Run Risk," *Journal of Political Economy*, 116(2), 260–302.

Hansen, L. P., and R. Jagannathan (1991): "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 99(2), pp. 225–262.

Hansen, L. P., and T. J. Sargent (2007): "Recursive robust estimation and control without commitment," *Journal of Economic Theory*, 136(1), 1–27.

Hansen, L. P., and T. J. Sargent (2008): *Robustness*. Princeton University Press, Princeton, NJ.

Hansen, L. P., and T. J. Sargent (2010a): "Chapter 20 - Wanting Robustness in Macroeconomics," vol. 3 of *Handbook of Monetary Economics*, pp. 1097 – 1157. Elsevier.

——— (2010b): "Fragile beliefs and the price of uncertainty," *Quantitative Economics*, 1(1), 129–162.

Kleshchelski, I., and N. Vincent (2008): "Robust Equilibrium Yield Curves," Working Paper, HEC Montreal.

Lettau, M., and J. A. Wachter (2011): "The term structures of equity and interest rates," *Journal of Financial Economics*, 101(1), 90 – 113.

Marakani, S. (2009): "Long run consumption risks: Are they there?," .

Mehra, R., and E. C. Prescott (1985): "The equity premium: A puzzle," *Journal of Monetary Economics*, 15(2), 145–161.

Nakamura, E., J. Steinsson, R. Barro, and J. Ursa (2013): "Crises and Recoveries in an Empirical Model of Consumption Disasters," *American Economic Journal: Macroeconomics*, 5(3), 35–74.

Petersen, I., M. James, and P. Dupuis (2000): "Minimax optimal control of stochastic uncertain systems with relative entropy constraints," *Automatic Control, IEEE Transactions on*, 45(3), 398–412.

Rietz, T. A. (1988): "The equity risk premium a solution," *Journal of Monetary Economics*, 22(1), 117–131.

Rubin, D. B. (1987): "A noniterative sampling importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing in- formation are modest: the SIR algorithm," *Journal of the American Statistical Association*, 82(398).

Schorfheide, F., D. Song, and A. Yaron (2014): "Identifying Long-Run Risks: A Bayesian Mixed-Frequency Approach," Working Paper 20303, National Bureau of Economic Research.

Smith, A. F. M., and A. E. Gelfand (1992): "Bayesian Statistics WIthout Tears: A Sampling-Resampling Perspective," *The American Statistician*, 46(2).

Strzalecki, T. (2011): "Axiomatic Foundations of Multiplier Preferences," *Econometrica*, 79(1), 47–73.

Swanson, E. (2014): "A Macroeconomic Model of Equities and Real, Nominal, and Defaultable Debt," Unpublished paper, University of California at Irvine.

Tallarini, T. D. (2000): "Risk-sensitive real business cycles," *Journal of Monetary Economics*, 45(3), 507–532.

Wachter, J. A. (2013): "Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?," *The Journal of Finance*, 68(3), 987–1035.

Xu, W., C. Wu, and H. Li (2010): "Robust general equilibrium under stochastic volatility model," *Finance Research Letters*, 7(4), 224 – 231.

**Tables**

Table 1: **Parameterization of endowment process - simple heteroscedastic white noise example**

| Parameter | Value |
|---|---|
| $G_0$ | $4.66 \times 10^{-3}$ |
| $\varphi_v$ | 0.91 |
| $\bar{v}$ | $2.69 \times 10^{-5}$ |
| $\sigma_v$ | $5.45 \times 10^{-6}$ |

Table 2: **Parameterization of preferences - simple heteroscedastic white noise example**

| | $\mathbf{MPR = 0.25}$ | $\mathbf{MPR = 0.375}$ | $\mathbf{MPR = 0.5}$ |
|---|---|---|---|
| $\alpha$ | $-44.7$ | $-63.9$ | $-80.1$ |
| $\beta$ | 0.9970 | 0.9965 | 0.9959 |
| $DEP(T = 100)$ | 10.8 | 3.1 | 0.7 |
| $DEP(T = 250)$ | 2.6 | 0.2 | 0.0 |
| $CRRA$ | 45.7 | 64.9 | 81.1 |

Table 3: **Worst case innovation means - simple heteroscedastic white noise example**

| | | $\mathbf{w_{g,t+1}}$ | | $\mathbf{w_{v,t+1}}$ |
|---|---|---|---|---|
| | $\mathbf{20^{th}}$**%ile** $\mathbf{v_t}$ | $\mathbf{50^{th}}$**%ile** $\mathbf{v_t}$ | $\mathbf{80^{th}}$**%ile** $\mathbf{v_t}$ | $\forall \mathbf{v_t}$ |
| $MPR = 0.25$ | $-0.18$ | $-0.23$ | $-0.28$ | 0.06 |
| $MPR = 0.375$ | $-0.25$ | $-0.33$ | $-0.39$ | 0.12 |
| $MPR = 0.50$ | $-0.32$ | $-0.42$ | $-0.49$ | 0.19 |

Table 4: **Conditional market price of risk by $v_t$ - simple heteroscedastic white noise example**

| $\mathbf{v_t}$ | **No Robustness** | **MPR = 0.25** | **MPR = 0.375** | **MPR = 0.5** |
|---|---|---|---|---|
| *High* | 0.01 | 0.29 | 0.44 | 0.58 |
| *Medium* | 0.01 | 0.25 | 0.37 | 0.49 |
| *Low* | 0.00 | 0.19 | 0.29 | 0.39 |

Table 5: **Unconditional moments of $\log g_t$ - simple heteroscedastic white noise example**

| | **Benchmark** | **MPR = 0.25** | **MPR = 0.375** | **MPR = 0.5** |
|---|---|---|---|---|
| *Mean(%Ann.)* | 1.86 | 1.31 | 0.99 | 0.62 |
| *St.Dev.(%Ann.)* | 1.03 | 1.13 | 1.21 | 1.30 |
| *Skew.* | 0.00 | −0.22 | −0.27 | −0.28 |
| *LRSD/St.Dev.* | 1.00 | 1.09 | 1.16 | 1.21 |

Table 6: **Unconditional autocorrelations of annual $\log g_t$ - simple heteroscedastic white noise example**

| *Lag* | **Benchmark** | **MPR = 0.25** | **MPR = 0.375** | **MPR = 0.5** |
|---|---|---|---|---|
| 2 | 0.00 | 0.03 | 0.05 | 0.07 |
| 3 | 0.00 | 0.02 | 0.04 | 0.05 |
| 4 | 0.00 | 0.01 | 0.03 | 0.04 |
| 5 | 0.00 | 0.01 | 0.02 | 0.02 |
| 6 | 0.00 | 0.01 | 0.01 | 0.02 |

Table 7: **Robustness check: vol-in-vol heteroscedastic white noise example**

| | **Benchmark** | **MPR = 0.25** | **MPR = 0.375** | **MPR = 0.5** |
|---|---|---|---|---|
| $\alpha$ | - | $-44.7$ | $-62.8$ | $-77.3$ |
| $\beta$ | - | 0.9970 | 0.9965 | 0.9961 |
| $DEP(T = 100)$ | - | 10.8 | 3.0 | 0.5 |
| $DEP(T = 250)$ | - | 2.6 | 0.1 | 0.0 |
| $Skew.$ | 0.00 | $-0.31$ | $-0.47$ | $-0.64$ |
| $LRSD/St.Dev.$ | 1.00 | 1.14 | 1.38 | 1.94 |
| $v_t$ half-life | 1.86 | 2.19 | 2.74 | 4.14 |
| Acorr. lag | | | | |
| 2 | 0.00 | 0.04 | 0.10 | 0.21 |
| 3 | 0.00 | 0.03 | 0.08 | 0.18 |
| 4 | 0.00 | 0.02 | 0.06 | 0.15 |
| 5 | 0.00 | 0.02 | 0.05 | 0.13 |
| 6 | 0.00 | 0.01 | 0.04 | 0.11 |

Table 8: **Parameterization of endowment process - adjusted LRR model**

| **Parameter** | **Value** |
|---|---|
| $G_0$ | 0.0015 |
| $G_{d,0}$ | 0.0015 |
| $\phi$ | 4.368 |
| $\varphi_d$ $(MPR = 0.25/0.375)$ | 3.053/3.579 |
| $\bar{v}$ | $0.0078^2$ |
| $\varphi_v$ | 0.999 |
| $\sigma_v$ | $6.3983 \times 10^{-07}$ |
| $\rho$ | 0.9289 |
| $\varphi_x$ | 0.0799 |

Table 9: **Parameterization of preferences - adjusted LRR model**

| | **MPR = 0.25** | **MPR = 0.375** |
|---|---|---|
| $\alpha$ | $-11.3$ | $-16.42$ |
| $\beta$ | 0.9986 | 0.9983 |
| $DEP(T = 100)$ | 10.2 | 2.6 |
| $DEP(T = 250)$ | 2.4 | 0.1 |
| $CRRA$ | 12.3 | 17.42 |

Table 10: **Baseline moments - adjusted LRR model**
Medians of moments obtained from 50000 looped samples of length 756 months.

| | BS ($MPR = 0.25$) | BS ($MPR = 0.375$) | BY | Data |
|---|---|---|---|---|
| $E[\log g_t]$ | 1.80 | 1.80 | 1.80 | 2.01 |
| $\sigma(\log g_t)$ | 2.45 | 2.45 | 2.44 | 1.02 |
| $AC1(\log g_t)$ | 0.30 | 0.30 | 0.31 | 0.26 |
| $E[\log g_{d,t}]$ | 2.00 | 2.00 | 1.82 | 2.29 |
| $\sigma(\log g_{d,t})$ | 9.06 | 9.06 | 10.53 | 27.61 |
| $AC1(\log g_{d,t})$ | 0.34 | 0.34 | 0.26 | $-0.58$ |
| $E[r_t]$ | 5.21 | 6.70 | 6.75 | 6.36 |
| $\sigma(r_t)$ | 13.02 | 13.90 | 16.54 | 16.52 |
| $AC1(r_t)$ | $-0.00$ | 0.05 | 0.01 | 0.08 |
| $E[r_{f,t}]$ | 2.61 | 2.61 | 2.59 | 0.89 |
| $\sigma(r_{f,t})$ | 1.89 | 1.90 | 1.23 | 1.82 |
| $AC1(r_{f,t})$ | 0.85 | 0.85 | 0.94 | 0.84 |
| $E[p_t - d_t]$ | 3.38 | 3.02 | 3.01 | 3.46 |
| $\sigma(p_t - d_t)$ | 0.12 | 0.12 | 0.18 | 0.43 |
| $AC1(p_t - d_t)$ | 0.85 | 0.84 | 0.89 | 0.98 |

Table 11: **Raw and scaled long run standard deviation of annualized consumption growth - adjusted LRR model**

| | **Benchmark** | **W.C.** ($MPR = 0.25$) | **W.C.** ($MPR = 0.375$) | **BY** |
|---|---|---|---|---|
| $LRSD$ | 4.06 | 7.39 | 9.74 | 6.28 |
| $LRSD/\sigma\left(\log g_t^{Ann}\right)$ | 1.26 | 1.92 | 2.26 | 2.27 |

Table 12: **Variance ratios - adjusted LRR model**
The variance ratio for consumption growth is defined $V(K) \equiv \frac{Var(\log g_{t+1} + \ldots + \log g_{t+K})}{K\, Var(\log g_{t+1})}$

| Variance ratios | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Consumption | | | | Dividends | | |
| Years | 2 | 4 | 6 | Years | 2 | 4 | 6 |
| BS | 1.37 | 1.60 | 1.63 | BS | 1.41 | 1.69 | 1.73 |
| BY | 1.46 | 1.96 | 2.22 | BY | 1.36 | 1.66 | 1.79 |
| Data | 1.27 | 1.29 | 1.29 | Data | 1.32 | 1.45 | 1.32 |

Table 13: **Predictability regressions $R^2$ - adjusted LRR model**
For $z_t \in \{r_t, \log g_t, \Delta d_t\}$ we estimate $\sum_{j=1}^{J} z_{t+j} = \alpha + \beta (p_t - d_t) + \varepsilon_{t+j}$

**Predictability**

| | Excess Returns | | | | Consumption | | |
|---|---|---|---|---|---|---|---|
| Quarters | 4 | 12 | 20 | Quarters | 4 | 12 | 20 |
| BS ($MPR = 0.25$) | 0.007 | 0.015 | 0.022 | BS ($MPR = 0.25$) | 0.157 | 0.090 | 0.049 |
| BS ($MPR = 0.375$) | 0.007 | 0.016 | 0.022 | BS ($MPR = 0.375$) | 0.143 | 0.083 | 0.046 |
| BY | 0.008 | 0.022 | 0.033 | BY | 0.237 | 0.269 | 0.213 |
| Data | 0.090 | 0.187 | 0.257 | Data | 0.000 | 0.001 | 0.002 |
| | Dividends | | | | | | |
| Quarters | 4 | 12 | 20 | | | | |
| BS ($MPR = 0.25$) | 0.275 | 0.131 | 0.065 | | | | |
| BS ($MPR = 0.375$) | 0.221 | 0.107 | 0.05 | | | | |
| BY | 0.159 | 0.180 | 0.147 | | | | |
| Data | 0.000 | $-0.002$ | $-0.003$ | | | | |

**Predictability of volatility**

| | Excess Returns | | | | Consumption | | |
|---|---|---|---|---|---|---|---|
| Quarters | 4 | 12 | 20 | Quarters | 4 | 12 | 20 |
| BS ($MPR = 0.25$) | 0.008 | 0.016 | 0.021 | BS ($MPR = 0.25$) | 0.007 | 0.016 | 0.021 |
| BS ($MPR = 0.375$) | 0.008 | 0.017 | 0.022 | BS ($MPR = 0.375$) | 0.007 | 0.017 | 0.022 |
| BY | 0.009 | 0.022 | 0.030 | BY | 0.008 | 0.021 | 0.030 |
| Data | 0.002 | 0.033 | 0.061 | Data | 0.226 | 0.437 | 0.462 |
| | Dividends | | | | | | |
| Quarters | 4 | 12 | 20 | | | | |
| BS ($MPR = 0.25$) | 0.007 | 0.016 | 0.021 | | | | |
| BS ($MPR = 0.375$) | 0.007 | 0.016 | 0.021 | | | | |
| BY | 0.009 | 0.021 | 0.030 | | | | |
| Data | 0.049 | 0.027 | 0.019 | | | | |

Table 14: **Term structure of real riskless bonds - adjusted LRR model**

| Maturity | $3m$ | $1y$ | $5y$ | $10y$ | $20y$ | $30y$ |
|---|---|---|---|---|---|---|
| BS ($MPR = 0.25$) | 2.53 | 2.28 | 1.76 | 1.61 | 1.50 | 1.44 |
| BS ($MPR = 0.375$) | 2.50 | 2.12 | 1.37 | 1.13 | 0.93 | 0.79 |
| BY | 2.55 | 2.38 | 1.74 | 1.30 | 0.91 | 0.75 |

Table 15: **Elasticity of intertemporal substitution regressions - adjusted LRR model**
We estimate the coefficients on $\log g_{t+1}$ and $r_{f,t+1}$ in $r_{f,t+1} = a + \left(\frac{1}{\psi}\right) \log g_{t+1} + \varepsilon_{t+1}$ and $\log g_{t+1} = a + \psi r_{f,t+1} + \varepsilon_{t+1}$ and report the median over many small sample loops of the implied estimated $\psi$.

| | $\log g_{t+1}$ on $r_{f,t+1}$ | $r_{f,t+1}$ on $\log g_{t+1}$ |
|---|---|---|
| BS ($MPR = 0.25$) | 1.071 | 1.186 |
| BS ($MPR = 0.375$) | 1.065 | 1.185 |
| BY | 1.495 | 1.613 |
| Data | 0.306 | 0.504 |

Table 16: **Parameterization - rare disasters model**

| **Parameter** | **Constant** $h$ | **ARG** $h$ |
|---|---|---|
| $G_0$ | $1.5 \times 10^{-3}$ | $1.5 \times 10^{-3}$ |
| $\bar{v}$ | $0.90 \times 10^{-5}$ or $3.30 \times 10^{-5}$ | $3.30 \times 10^{-5}$ |
| $\bar{h}$ | $2.08 \times 10^{-3}$ | $2.08 \times 10^{-3}$ |
| $\delta$ | 0.1 | 0.1 |
| $\theta$ | $-0.3$ | $-0.3$ |
| $q$ | 0.4 | 0.4 |
| $\phi$ | $-$ | 3.8 |
| $\varphi_h$ | $-$ | 0.993 |
| $\delta_h$ | $-$ | 1.788 |
| $c_h$ | $-$ | $7.77 \times 10^{-6}$ |

Table 17: **Calibration - rare disaster model (constant intensity) -** $E[r_B] = 2.59\%$

| | **MPR=0.25** | | **MPR=0.375** | | **MPR=0.5** | |
|---|---|---|---|---|---|---|
| | Low $\bar{v}$ | High $\bar{v}$ | Low $\bar{v}$ | High $\bar{v}$ | Low $\bar{v}$ | High $\bar{v}$ |
| $E\left[r^B\right]$ | 2.28 | 2.26 | 2.27 | 2.27 | 2.28 | 2.28 |
| $E\left[r^B \vert ND\right]$ | 2.59 | 2.59 | 2.59 | 2.59 | 2.59 | 2.59 |
| $\tilde{\theta}$ | $-0.33$ | $-0.33$ | $-0.34$ | $-0.34$ | $-0.34$ | $-0.34$ |
| $\tilde{E}[\#jumps/cent.]$ | 6.4 | 6.3 | 8.2 | 8.2 | 9.9 | 9.7 |
| $\alpha$ | $-2.99$ | $-2.96$ | $-3.74$ | $-3.72$ | $-4.27$ | $-4.22$ |
| $\beta$ | 0.99823 | 0.99816 | 0.99788 | 0.99779 | 0.99755 | 0.99748 |
| $DEP(T = 100)$ | 31.2 | 30.1 | 24.9 | 23.8 | 20.6 | 20.5 |
| $DEP(T = 250)$ | 21.3 | 20.9 | 13.6 | 13.1 | 9.4 | 9.4 |
| $CRRA$ | 3.99 | 3.96 | 4.74 | 4.72 | 5.27 | 5.22 |

Table 18: **Calibration - rare disaster model (constant intensity) -** $E\left[r_B\right] = 1.36\%$

| | **MPR=0.25** | | **MPR=0.375** | | **MPR=0.5** | |
|---|---|---|---|---|---|---|
| | Low $\bar{v}$ | High $\bar{v}$ | Low $\bar{v}$ | High $\bar{v}$ | Low $\bar{v}$ | High $\bar{v}$ |
| $E\left[r^B\right]$ | 1.07 | 1.07 | 1.05 | 1.04 | 1.05 | 1.05 |
| $E\left[r^B \mid ND\right]$ | 1.36 | 1.36 | 1.36 | 1.36 | 1.36 | 1.36 |
| $\tilde{\theta}$ | $-0.33$ | $-0.33$ | $-0.34$ | $-0.34$ | $-0.34$ | $-0.34$ |
| $\tilde{E}\left[\#jumps/cent.\right]$ | 6.4 | 6.3 | 8.2 | 8.2 | 9.9 | 9.7 |
| $\alpha$ | $-2.99$ | $-2.96$ | $-3.74$ | $-3.71$ | $-4.27$ | $-4.22$ |
| $\beta$ | 0.99923 | 0.99916 | 0.99888 | 0.99879 | 0.99856 | 0.99848 |
| $DEP(T=100)$ | 30.8 | 30.6 | 25.3 | 24.5 | 20.3 | 20.5 |
| $DEP(T=250)$ | 21.1 | 20.6 | 13.8 | 13.2 | 9.7 | 9.9 |
| $CRRA$ | 3.99 | 3.96 | 4.74 | 4.71 | 5.27 | 5.22 |

Table 19: **Calibration - rare disaster model (ARG intensity)**
We obtain the return series from aggregating monthly returns and then construct the excess return by comparing the annualized gross returns. The moments are based on data from approximately $100,000$ periods of simulations. We drop observations for periods including a jump, when calculating the 'conditional' regressions. We have $\alpha = -2.97$, $\beta = 0.99921$ and DEP for 100 and 250 quarters of 10.0 and 4.1, respectively.

| | **Unconditional** | **Conditional** | **Wachter (2013) Conditional** |
|---|---|---|---|
| $E[r^B]$ | 1.06 | 1.36 | 1.36 |
| $\sigma(r^b)$ | 3.31 | 0.96 | 2.00 |
| $E[r^e - r^b]$ | 8.29 | 9.61 | 8.85 |
| $\sigma(r^e)$ | 20.25 | 17.19 | 17.66 |
| $Sharpe$ | 0.43 | 0.55 | 0.49 |
| $\sigma(\log(g_t))$ | 5.16 | 1.98 | 1.99 |
| $\sigma(\log(g_{d,t}))$ | 19.59 | 7.51 | 5.16 |

Table 20: **Predictability - rare disaster model (ARG intensity)**
We estimate the long horizon regression $\sum_{j=0}^{k}\log(R^e_{t+j}) - \log(R^b_{t+j}) = \beta_0 + \beta_1(p_t - d_t) + \varepsilon_t$ where we obtain the return series from aggregating monthly returns and then construct the excess return by comparing the annualized gross returns. $p_t - d_t$ is the log price dividend ratio, where the dividend is identified with the theoretical object in the model - rather than using a trailing sum of dividends in the denominator. We estimate the regressions on data from approximately $100,000$ periods of simulations. We drop observations where the long horizon for returns includes a jump, when calculating the 'conditional' regressions.

| | **Horizon in Years** | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 4 | 6 | 8 | 10 |
| | **Panel A: Model - Unconditional Moments** | | | | | |
| $\beta_1$ | $-0.09$ | $-0.17$ | $-0.32$ | $-0.46$ | $-0.56$ | $-0.64$ |
| $R^2$ | 0.02 | 0.04 | 0.06 | 0.09 | 0.10 | 0.11 |
| | **Panel B: Model - No disasters** | | | | | |
| $\beta_1$ | $-0.13$ | $-0.26$ | $-0.48$ | $-0.68$ | $-0.83$ | $-0.95$ |
| $R^2$ | 0.09 | 0.16 | 0.28 | 0.38 | 0.43 | 0.47 |
| | **Panel C: Wachter (2013) - No disasters** | | | | | |
| $\beta_1$ | $-0.16$ | $-0.30$ | $-0.56$ | $-0.77$ | $-0.95$ | $-1.10$ |
| $R^2$ | 0.13 | 0.24 | 0.41 | 0.52 | 0.59 | 0.63 |

**Figures**

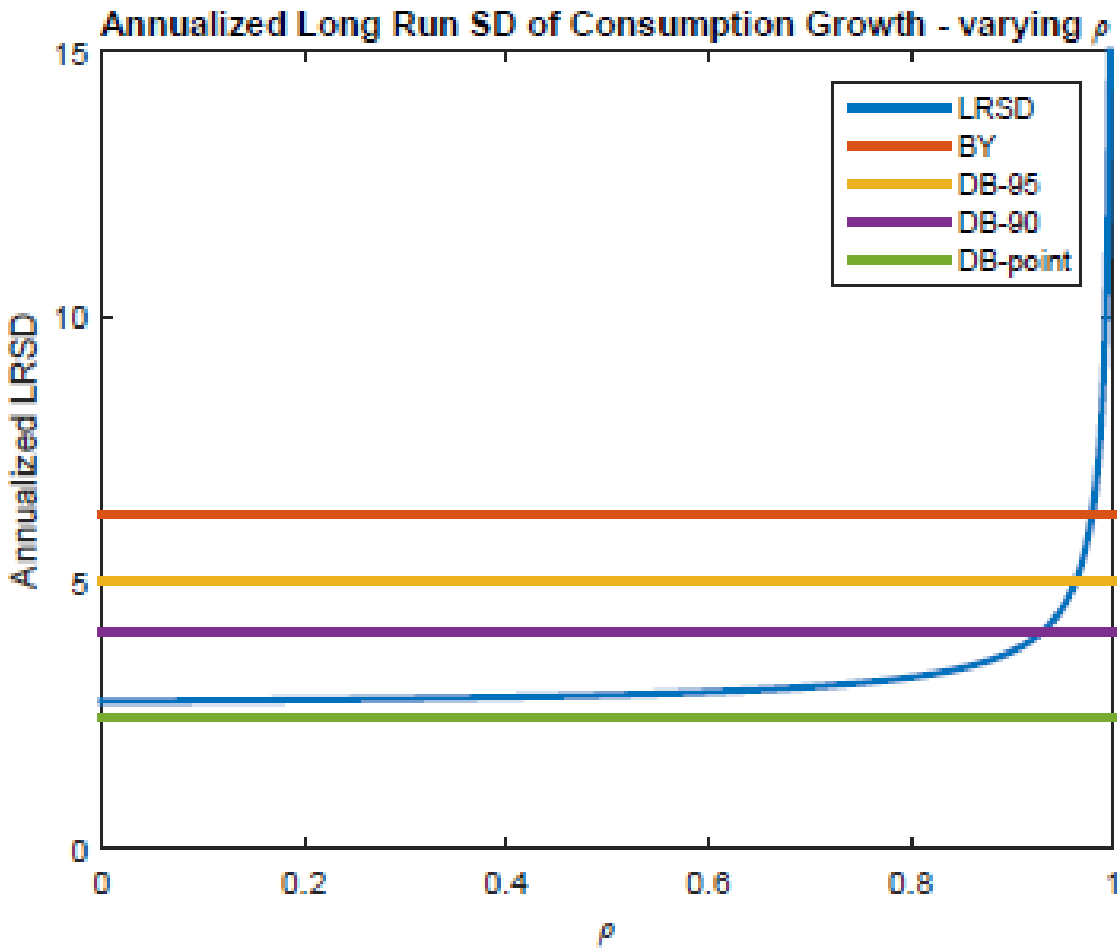


Figure 1: Long run standard deviation - varying $\rho$ maintaining $var(\Delta c_t)$

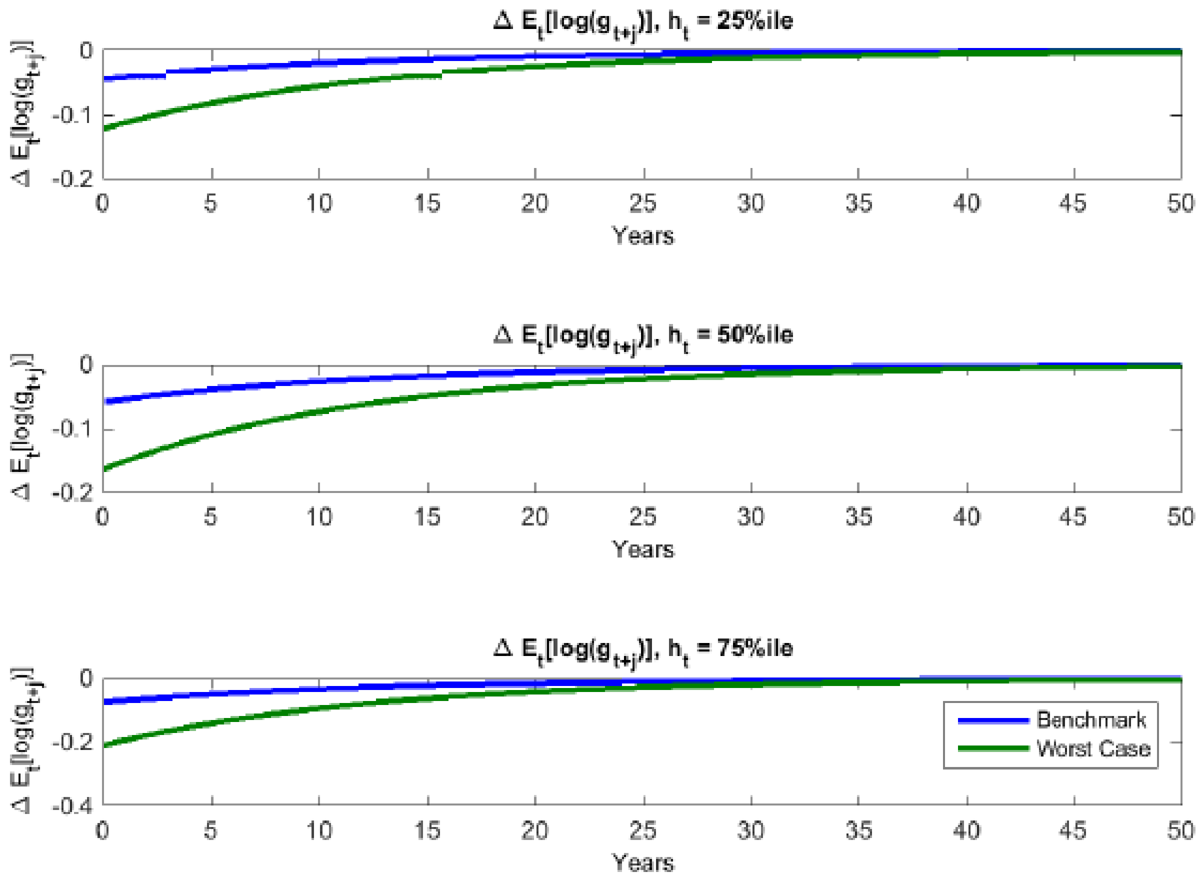


Figure 2: News to expected growth rate (annualized percent) following a 1 S.D. innovation to $w_{h,t}$ - conditional on $h_{t-1}$ percentiles.