Long-Run Risk is the Worst-Case Scenario

Rhys Bidder
Federal Reserve Bank of San Francisco

Ian Dew-Becker
Northwestern University

May 2016

Long-Run Risk is the Worst-Case Scenario

Rhys Bidder and Ian Dew-Becker[*]

May 4, 2016

Abstract

We study an investor who is unsure of the dynamics of the economy. Not only are parameters unknown, but the investor does not even know what order model to estimate. She estimates her consumption process nonparametrically – allowing potentially infinite-order dynamics – and prices assets using a pessimistic model that minimizes lifetime utility subject to a constraint on statistical plausibility. The equilibrium is exactly solvable and we show that the pricing model always includes long-run risks. With risk aversion of 4.7, the model matches major facts about asset prices, consumption, and dividends. The paper provides a novel link between ambiguity aversion and non-parametric estimation.

1 Introduction

Economists do not agree on the dynamic properties of the economy. There has been a long debate in the finance literature over how risky consumption growth is in the long-run (e.g. Bansal, Kiku, and Yaron (2012) and Beeler and Campbell (2012)), and it is well known that long-run forecasting is econometrically difficult (Müller and Watson (2013)). It is likely that the average investor is also unsure of the true model driving the world. This paper studies the behavior of such an investor.

With exactly solved results, we show that a model in which investors have Epstein–Zin preferences and uncertainty about consumption dynamics generates high and volatile risk premia, excess volatility in stock returns, a large degree of predictability in stock returns, low and stable interest rates, and an estimated elasticity of intertemporal substitution from interest rate regressions of zero as measured in Hall (1988) and Campbell and Mankiw (1989). Moreover, variation over time in risk or model uncertainty is not required to generate predictability in returns.

We argue that investors consider a set of models of the economy that is only weakly constrained. People face pervasive ambiguity: no one can say they know the exact specification to estimate when forecasting economic activity. So rather than just allowing uncertainty about the parameters in a specific model, or putting positive probability on a handful of models, we treat investors as considering an infinite-dimensional space of autoregressive moving average (ARMA) specifications of unrestricted order. They therefore face a nonparametric problem in the sense that the number of parameters to be estimated is potentially infinite.[1]

Infinite-dimensional estimation problems are well known to be difficult or impossible to approach with standard likelihood-based methods.[2] But people must estimate *some* model. So, following the literature on nonparametric time series estimation, we assume that in a finite sample, they estimate a relatively small-scale model, which they view as an approximation to the truth. But the true model may be infinite-dimensional. So investors face a highly non-standard estimation problem, and, as Sims (1972) shows, the consequences for welfare of using a misspecified model can be severe.

The uncertainty due to estimating consumption dynamics when the true model is unknown and potentially of infinite order is clearly very different from that due to standard sources of risk, such as future innovations to the consumption process; in many cases a valid posterior distribution *cannot* be placed on the space of models. It is simply not always possible to be a Bayesian (in particular, when the model order is unknown). It is reasonable to think that people view model

---

[1]This is a typical definition of nonparametric estimation – see, e.g., Chen (2007)

[2]Diaconis and Freedman (1986) note that Doob's (1948) theorem on the consistency of Bayesian estimates only applies to finite-dimensional parameter spaces. In infinite-dimensional settings, Bayesian estimators need not be consistent for even apparently reasonable priors. Sims (1971, 1972) shows that it is generally impossible to create accurate confidence intervals in such settings. See also Chow and Grenander (1985), Faust (1999), and Hansen and Sargent (2007) (who note the links between Sims (1971), Diaconis and Freedman (1986), and robust control theory).

uncertainty fundamentally differently from other sources of risk, in the sense of Knight (1921) and Ellsberg (1961).

There are multiple potential ways of modeling attitudes towards such ambiguity. We draw upon the insights of Gilboa and Schmeidler's (1989) work on choice under ambiguity in modeling people as making choices under a "worst-case" process for consumption that is chosen to minimize lifetime utility.[3] Such analysis is highly tractable, it is a plausible and parsimonious description of how people might approach ambiguity, and it yields results that have a natural economic interpretation because they point directly to the specific model that is most painful to agents.

Our headline theoretical result is that for an ambiguity-averse agent whose point estimate is that consumption growth is white noise, the worst-case model used for decision-making, chosen from the entire space of ARMA models, is an ARMA(1,1) with a highly persistent trend – literally the homoskedastic version of Bansal and Yaron's (2004) long-run risk model. More generally, whatever the investor's point estimate, the worst-case model always adds a long-run risk component to it.

The low-frequency fluctuations that people in our model fear die out at precisely the rate of time preference. In a sense, then, they are shocks that effectively last the rest of the investor's life. So a way of interpreting our results is that they say that what people fear most, and what makes them averse to investing in equities, is that growth rates or asset returns are going to be persistently lower over the rest of their lives than they have been on average in the past. Our specific formulation of model uncertainty allows us to formalize that intuition.

Our results are derived in the frequency domain, which allows strikingly clear conceptual and analytical insights. Two factors determine the behavior of the worst-case model at each frequency: estimation uncertainty and how utility is affected by fluctuations at that frequency. Growth under the worst-case model has larger fluctuations at frequencies about which there is more uncertainty or that are more painful. Quantitatively, we find that differences in estimation uncertainty across frequencies are relatively small. Instead, since people with Epstein–Zin preferences are highly averse to low-frequency fluctuations (for parameterizations such as ours where they prefer an early resolution of uncertainty), persistent shocks play the largest role in robust decision-making.

A criticism of the long-run risk model has always been that it depends on a process for consumption growth that is difficult to test for.[4] We turn that idea on its head and argue that it is the difficulty of testing for and rejecting long-run risk that actually makes it a sensible model for investors to focus on. If anything, our result is more extreme than that of Bansal and Yaron (2004): whereas they posit a consumption growth trend with shocks that have a half-life of 3 years, the

---

[3]Alternative models of ambiguity include Hansen and Sargent (2001), Epstein and Schneider (2003), Klibanoff, Marinacci, and Mukerji (2005) and Maccheroni, Marinacci, and Rustichini (2006).

[4]See Beeler and Campbell (2012) and Marakani (2009).

endogenous worst-case model that we derive features trend shocks with a half-life of *70* years.

In a calibration of the model, we show that it explains a wide range of features of financial markets that have been previously viewed as puzzling. Similar to the intuition from Bansal and Yaron (2004), equities earn high average returns in our model because low-frequency fluctuations in consumption growth induce large movements in both marginal utility and equity prices. In our setup, though, long-run risk need not actually exist – it only needs to be plausible.

The results that we obtain on excess volatility, forecasting, and interest rate regressions all follow from the fact that the pricing model that our investor uses always involves more persistence than her point estimate (i.e. the model used by an econometrician with access to the same data sample as the investor). Since the pricing model has excess persistence, investors overextrapolate recent news relative to what the point estimate would imply. Following positive shocks, then, stock prices are relatively high and econometric forecasts of returns are low. We are thus able to generate predictability without any changes in risk or risk aversion over time. Significantly, we obtain this result in a model in which agents are rational, if uncertain. The lower case of the 'r' in rational here is important. Agents behave as if optimizing under a "worst case" distribution that is different from the true data generating process, putting us outside the standard Rational Expectations paradigm. As discussed in Hansen and Sargent (2007), this is characteristic of models of ambiguity.

In generating all these results we have no more free parameters than other standard models of consumption and asset prices. We link the parameter that determines how the agent penalizes deviations from her point estimate for consumption dynamics directly to the coefficient of relative risk aversion. There is thus a single free parameter that determines risk preferences, and it corresponds to a coefficient of relative risk aversion of only 4.7. We also take no extreme liberties with beliefs – the investor's pricing model is essentially impossible to distinguish from the true model in a 100-year sample. Using a correctly specified likelihood ratio test, the null hypothesis that the pricing model is true is rejected at the five-percent level in, at most, only 6.8 percent of samples.

Our analysis directly builds on a number of important areas of research. First, the focus on a single worst-case outcome is closely related to Gilboa and Schmeidler's (1989) work on ambiguity aversion that provides an axiomatic basis for decision making under a worst-case distribution (though here we do not start from an axiomatic foundation). Second, we build on the analysis of generalized recursive preferences to allow for the consideration of multiple models, especially Hansen and Sargent (2010) and Ju and Miao (2012).[5] The work of Hansen and Sargent (2010) is

---

[5]See, e.g., Kreps and Porteus (1978); Weil (1989); Epstein and Zin (1991); Maccheroni, Marinacci, and Rustichini (2006); and Hansen and Sargent (2005), among many others. There is also a large recent literature in finance that specializes models of ambiguity aversion to answer particularly interesting economic questions, such as Liu Pan and Wang (2004) and Drechsler's (2013) work with tail risk and the work of Uppal and Wang (2003), Maenhout (2004), Sbuelz and Trojani (2008), and Routledge and Zin (2009) on portfolio choice. Recent papers on asset pricing with learning and ambiguity aversion include Veronesi (2000), Brennan and Xia (2001), Epstein and Schneider (2007), Cogley and Sargent (2008), Leippold, Trojani, and Vanini (2008), Ju and Miao (2012), and Collin-Dufresne, and

perhaps most comparable to ours, in that they study an investor who puts positive probability on both a white-noise model and a parameterized long-run risk model for consumption growth. The key difference here is that we consider a nonparametric setting in which the agent considers *all* ARMA models, instead of only two. The emergence of the long-run risk model as the one that she focuses on is entirely endogenous.[6] We also obtain analytic results that provide economic insights into precisely what model is most painful to agents, whereas Hansen and Sargent's (2010) analysis is numerically solved.

Since the worst-case model is more persistent than the point estimate, pricing behavior is similar to the extrapolation implied by the "natural expectations" studied by Fuster, Hebert, and Laibson (2011). Our results differ from theirs, though, in that we *always* obtain excess extrapolation, whereas in their setting it results from the interaction of suboptimal estimation on the part of investors with a specific data-generating process. Cecchetti, Lam, and Mark (2000) examine a setting in which investors use a rule-of-thumb estimation technique and are subject to random belief shocks to obtain excess volatility and predictability. In our analysis, there is no assumption that the estimation method is suboptimal, and excess volatility and predictability are natural consequences of the pessimistic model that agents choose. Our paper complements the literature on belief distortions and extrapolative expectations by deriving them as a natural response to model uncertainty.[7]

The remainder of the paper is organized as follows. Section 2 discusses the agent's estimation method. Section 3 describes the basic structure of the agent's preferences, and section 4 then derives the worst-case model. We examine asset prices in general under the preferences in section 5. Section 6 then discusses the calibration and section 7 analyzes the quantitative implications of the model. Section 8 concludes.

2 Measuring model plausibility

We begin by describing the set of possible models that investors consider and the estimation method they use to measure the relative plausibility of different models.

2.1 Economic environment

We study a pure endowment economy.

Lochstoer (2013).

[6]Bidder and Smith (2015) also develop a model in which the worst-case process of an agent with multiplier preferences also features extra peristence that arises from the interaction of ambiguity aversion and stochastic volatility.

[7]See also Barsky and Delong (1993), Abel (2002), Brandt, Zheng, and Zhang (2004), and Hirshleifer and Yu (2013).

Assumption 1. *Investors form expectations for future log consumption growth, Δc, using models of the form*

$$\Delta c_t \;=\; \mu + a\left(L\right)\left(\Delta c_{t-1} - \mu\right) + \varepsilon_t \tag{1}$$

$$\varepsilon_t \;\sim\; N\left(0, \sigma^2\right) \tag{2}$$

where μ is mean consumption growth, $a\left(L\right)$ is a power series in L, the lag operator, with roots inside the unit circle, and ε_t is an innovation.

The change in log consumption on date t, Δc_t, is a function of past consumption growth and a shock. We restrict our attention to models with purely linear feedback from past to current consumption growth. It seems reasonable to assume that people use linear models for forecasting, even if consumption dynamics are not truly linear, given that the economics literature focuses almost exclusively on linear models. For our purposes, the restriction to the class of linear processes is a description of the agent's modeling method, rather than an assumption about the true process driving consumption growth. We make the further assumption that ε_t is i.i.d. normal. While the assumption of normality is not necessary, it simplifies the exposition; the key assumption is that ε_t is serially independent.[8]

In much of what follows, it will be more convenient to work with the moving average (MA) representation of the consumption process (1),

$$\Delta c_t \;=\; \mu + b\left(L\right)\varepsilon_t \tag{3}$$

$$\text{where } b\left(L\right) \;\equiv\; \left(1 - La\left(L\right)\right)^{-1} \tag{4}$$

$$=\; 1 + \sum_{j=1}^{\infty} b_j L^j \tag{5}$$

We can thus express the dynamics of consumption equivalently as depending on a or just on the power series $b\left(L\right)$, with coefficients b_j.[9] The two different representations are each more convenient than the other in certain settings, so we will refer to both in what follows. They are directly linked to each other through equation (4), so that a particular choice of a is always associated with a distinct value of b and vice versa (as long as b is invertible, which we impose).

There are no latent state variables. When a model $a\left(L\right)$ has infinite order we assume that the agent knows all the necessary lagged values of consumption growth for forecasting (or has

---

[8]It is straightforward to solve the model when ε_t has an arbitrary distribution but remains serially independent. While time varying volatility in innovations is an important area of analysis (e.g. Drechsler and Yaron (2011)) we avoid it here for simplicity.

[9]In working with finite order regressive or moving average representations, a and b can be regarded as polynomials. When, as will be necessary below, we deal with potentially infinite order representations, we continue to use the term polynomial for ease of expression, though in that case $b\left(L\right)$ and $a\left(L\right)$ are, formally, power series.

dogmatic beliefs about them) so that no filtering is required. We discuss necessary constraints on the models below. For now simply assume that they are sufficiently constrained that any quantities we derive exist.

We set the notation $\Theta \equiv \{b, \mu, \sigma^2\}$ to represent the set of parameters that defines a model of consumption growth. Note that Θ induces a joint Gaussian density over sequences of consumption growth realizations.

2.2 Model distance

For the purpose of forecasting consumption growth, the agent in our model chooses among specifications for consumption growth, Θ, partly based on their statistical plausibility. As is common in the literature, the plausibility of a model, or its distance from the agent's point estimate, is measured by the Kullback–Leibler divergence, or relative entropy, which is the expected value of a log likelihood ratio statistic comparing a pair of models.[10]

We denote the divergence between two models by $g\left(\Theta; \bar{\Theta}\right)$, where $\bar{\Theta}$ is the benchmark or point estimate, and Θ is some alternative model. Suppose the agent has a point estimate $\bar{\Theta}$ and she compares it to an alternative model Θ based on their relative log likelihoods. Then as the number of observations grows to infinity, $g\left(\Theta; \bar{\Theta}\right)$ is the limit of the expectation of that likelihood ratio statistic if the data is generated by the model Θ. The likelihood ratio is a natural choice to measure the difference between a pair of models because the Neyman–Pearson lemma shows that such a test is the most powerful way to discriminate between a pair of models.

Any sample of observed consumption growth of length T from the model (3) has a multivariate normal distribution with mean μ and a covariance matrix determined by $\{b, \sigma\}$, denoted $\Sigma_{\Theta, T}$. The log likelihood for a sample of length T is then

$$LL_T\left(\Theta\right) = -\frac{1}{2}\log\left|\Sigma_{\Theta, T}\right| - \frac{1}{2}\left(\Delta c_{1,\ldots,T} - \mu\right)' \Sigma_{\Theta, T}^{-1}\left(\Delta c_{1,\ldots,T} - \mu\right) \tag{6}$$

where $\Delta c_{1,\ldots,T}$ denotes a column vector containing the sample of observed consumption growth between dates 1 and T.

Our analysis of the model takes place in the frequency domain because it will allow us to obtain a tractable and interpretable solution. The analysis centers on the *transfer function*,

$$B\left(\omega\right) \equiv b\left(e^{i\omega}\right) \tag{7}$$

for $i \equiv \sqrt{-1}$. The transfer function measures how the filter $b\left(L\right)$ transfers power at each frequency,

---

[10]See Hansen and Sargent (2001), Maccheroni, Marinacci, and Rusticini (2006), Sbuelz and Trojani (2008), Strzalecki (2011), Drechsler (2013), among others.

ω, from the white-noise innovations, ε, to consumption growth.

Now suppose consumption growth is generated by the model Θ. One may show that as $T \to \infty$, the expected difference between the log likelihoods for the models Θ and $\bar{\Theta}$ converges to

$$
\begin{aligned}
\lim_{T \to \infty} T^{-1} E_\Theta \left[LL_T \left(\bar{\Theta} \right) - LL_T \left(\Theta \right) \right] \;=\; & -\frac{1}{2} \frac{\sigma^2}{\bar{\sigma}^2} \int \frac{\left| B \left(\omega \right) - \bar{B} \left(\omega \right) \right|^2}{\left| \bar{B} \left(\omega \right) \right|^2} d\omega \\
& -\frac{1}{2} \frac{\left(\mu - \bar{\mu} \right)^2}{\bar{\sigma}^2 \left| \bar{B} \left(0 \right) \right|^2} + \frac{1}{2} \left(\log \left(\frac{\sigma^2}{\bar{\sigma}^2} \right) - \frac{\sigma^2 - \bar{\sigma}^2}{\bar{\sigma}^2} \right) \quad (8)
\end{aligned}
$$

(where $|\cdot|$ denotes the norm of a complex number, the notation E_Θ indicates an unconditional expectation taken over the probability measure induced by Θ, and the integral sign with no limits denotes $\frac{1}{2\pi} \int_{-\pi}^{\pi}$). This result is a simple application of Whittle's (1953) limiting formula for the log likelihood (see, e.g., Dahlhaus (1996)).

The relative entropy between Θ and $\bar{\Theta}$ depends on three terms. The first measures the difference between the dynamics (in terms of autocorrelations) implied by the two models. A simple way to interpret it is that $\left| \bar{B} \left(\omega \right) \right|^2$ measures the uncertainty about the size of fluctuations at frequency ω, so the entropy distance penalizes squared deviations between the models, $\left| B \left(\omega \right) - \bar{B} \left(\omega \right) \right|^2$, less strongly at frequencies at which there is more uncertainty. This is the major term that will drive our results – it determines how the agent quantifies deviations of dynamics from the benchmark.

The second term incorporates the differences in the means of the distributions under the two models. We obtain the typical result that the uncertainty about the mean depends on the spectrum at frequency zero.[11] In other words, deviations in the mean of consumption growth between the two models are penalized with a similar scaling to deviations in the dynamics, but they are essentially infinitely low frequency differences – i.e. infinitely long lived shocks to consumption growth – so they are scaled by $\bar{\sigma}^2 \left| \bar{B} \left(0 \right) \right|^2$.

Finally, there is a contribution from the deviation of the innovation variance, σ^2, from the benchmark, which we find to have quantitatively minimal effects in our calibration.

As a divergence measure, (8) is precisely the limit of the Kullback–Leibler (KL) divergence between the models Θ and $\bar{\Theta}$ – the relative entropy between the joint distributions for consumption growth implied by Θ and $\bar{\Theta}$ as the sample length grows to infinity. We therefore define the agent's measure of model plausibility in the following assumption:

Assumption 2. *Given a point estimate $\bar{\Theta}$, investors measure the statistical plausibility of an*

---

[11] $\bar{f} \left(\omega \right) \equiv \bar{\sigma}^2 \left| B \left(\omega \right) \right|^2$ is the *spectral density* (or spectrum) of consumption growth under the model $\bar{\Theta}$. The spectral density decomposes the total variance of a time series into components coming from fluctuations at different frequencies.

alternative model Θ through the divergence measure,

$$
\begin{aligned}
g\left(\Theta;\bar{\Theta}\right) \;\equiv\; & \frac{1}{2}\frac{\sigma^2}{\bar{\sigma}^2}\int\frac{\left|B\left(\omega\right)-\bar{B}\left(\omega\right)\right|^2}{\left|\bar{B}\left(\omega\right)\right|^2}d\omega \\
& +\frac{1}{2}\frac{\left(\mu-\bar{\mu}\right)^2}{\bar{\sigma}^2\left|\bar{B}\left(0\right)\right|^2}-\frac{1}{2}\left(\log\left(\frac{\sigma^2}{\bar{\sigma}^2}\right)-\frac{\sigma^2-\bar{\sigma}^2}{\bar{\sigma}^2}\right)
\end{aligned}
\tag{9}
$$

While the KL divergence is widely studied and has a prominent place in the ambiguity literature, one could ask how an investor, in reality, might adopt such a measure of statistical discrimination. Is it reasonable to think that an investor armed with a simple estimation toolkit would behave as if she were using this distance measure?

As noted above, it is the first term in the definition of $g\left(\Theta;\bar{\Theta}\right)$ that drives our results. That component also arises if the investor uses the nonparametric estimation methods for AR and MA models described by Berk (1984) and Brockwell and Davis (1988). Specifically, one may model agents as estimating AR or MA models whose lag orders grow with the sample size. They thus have nonparametric confidence intervals. We show in the online appendix that if they follow such an estimation process, then as the sample size grows, a Wald test of the difference in the MA coefficients between the point estimate and any alternative approaches $\frac{1}{2\pi}\int_{-2\pi}^{2\pi}\frac{\left|B(\omega)-\bar{B}(\omega)\right|^2}{\left|\bar{B}(\omega)\right|^2}d\omega$. So the key part of the KL divergence that addresses dynamics may also be derived from an explicit nonparametric estimation method.[12]

3   Preferences

Given a particular model of consumption dynamics, the agent has Epstein–Zin (1991) preferences. We augment those preferences with a desire for a robustness against alternative models. The desire for robustness induces the agent to form expectations, and hence calculate utility and asset prices, under a pessimistic but plausible model, where plausibility is quantified using the estimation approach described above.

3.1   Utility given a model

Assumption 3. *Given a forecasting model $\Theta \equiv \{b, \mu, \sigma^2\}$, the investor's utility is described by Epstein–Zin (1991) preferences. The coefficient of relative risk aversion is α, the time discount parameter is β, and the elasticity of intertemporal substitution (EIS) is equal to 1. Lifetime utility,*

---

[12]Intuitively, this result is related to the fact that Wald and likelihood ratio tests are asymptotically equivalent. We thank Lars Hansen for pointing out the connection between the method based on the Wald test – which is how we originally derived our results – and the KL divergence.

v, for a fixed model Θ, is

$$v\left(\Delta c^t;\Theta\right) = (1-\beta)c_t + \frac{\beta}{1-\alpha}\log E_t\left[\exp\left(v\left(\Delta c^{t+1};\Theta\right)(1-\alpha)\right)|\Theta\right] \qquad (10)$$

$$= c_t + \sum_{k=1}^{\infty}\beta^k E_t\left[\Delta c_{t+k}|\Theta\right] + \frac{\beta}{1-\beta}\frac{1-\alpha}{2}\sigma^2 b\left(\beta\right)^2 \qquad (11)$$

where $E_t\left[\cdot|\Theta\right]$ denotes the expectation operator conditional on the history of consumption growth up to date t, Δc^t, assuming that consumption is driven by the model Θ.

$\frac{\beta}{1-\beta}\frac{1-\alpha}{2}\sigma^2 b\left(\beta\right)^2$ is an adjustment to utility for risk. The investor's utility is lower when risk aversion or the riskiness of the endowment is higher. The relevant measure of the risk of the endowment is $\sigma^2 b\left(\beta\right)^2$, which measures the variance of the shocks to lifetime utility in each period. $b\left(\beta\right)$ measures the total discounted effect of a unit innovation to ε_{t+1} on consumption growth, and hence utility, in the future. It is the term involving $b\left(\beta\right)$ that causes people with Epstein–Zin preferences to be averse to long-run risk.[13] $b\left(\beta\right)$ can be written in terms of the transfer function as

$$b\left(\beta\right) = \int Z\left(\omega\right)^* B\left(\omega\right)d\omega \qquad (12)$$

$$\text{where } Z\left(\omega\right) \equiv \sum_{j=0}^{\infty}\beta^j e^{i\omega j} \qquad (13)$$

and * denotes a complex conjugate. The parameter μ enters through the discounted expectation of future consumption growth, and hence appears as $\frac{\beta}{1-\beta}\mu$.

3.2 Robustness over dynamics

Equation (11) gives lifetime utility conditional on consumption dynamics. We now discuss the investor's consideration of alternative models of dynamics.

The investor entertains a set of possible values for Θ and can associate with any model a measure of its plausibility, g (from assumption 2). Seeking robustness, the investor makes decisions that are optimal in an unfavorable world – specifically, as though consumption growth is driven by worst-case dynamics, denoted $\Theta^w \equiv \{b^w, \sigma_w^2, \mu^w\}$. These dynamics are not the worst in an unrestricted sense but, rather, are the worst among statistically plausible models. So the investor does not fear completely arbitrary models – she focuses on models that are not too far from her point estimate in terms of KL distance (expected log likelihood).

---

[13]We focus on the case of a unit EIS to ensure that we can derive analytic results. The precise behavior of interest rates is not our primary concern, so a unit EIS is not particularly restrictive. The unit EIS also allows us to retain the result that Epstein–Zin preferences are observationally equivalent to a robust control model, as in Barillas, Hansen, and Sargent (2009), which will be helpful in our calibration below

Assumption 4. *Investors use a worst-case model to form expectations – for both calculating utility and pricing assets – that is obtained as the solution to a penalized minimization problem:*

$$\Theta^w = \arg\min_{\Theta} \left\{ E\left[v\left(\Delta c^t; b, \sigma^2, \mu \right) |\Theta \right] + \lambda g\left(\Theta; \bar{\Theta} \right) \right\} \tag{14}$$

$\{b^w, \sigma_w^2, \mu^w\}$ is the model that gives the agent the lowest unconditional expected lifetime utility, subject to the penalty g.[14] λ is a parameter that determines how much weight the penalty receives. As usual, λ can either be interpreted directly as a parameter or as a Lagrange multiplier on a constraint on the KL divergence g. The KL divergence can be large for three reasons: deviations in the dynamics, b; deviations in the mean, μ; and deviations in the conditional variance, σ^2.

The agent's assessment of plausibility is based on our statistical measure of distance and controlled by λ. We are modeling the agent's beliefs about potential models by assuming that she compares possible models to a point estimate $\bar{\Theta}$. The role of g in our analysis is similar to that of the KL divergence used in the robust control model of Hansen and Sargent (2001), in that it imposes discipline on what models the investor considers.

There are three important differences between our analysis and the multiplier preferences of Hansen and Sargent (2001). First, in Hansen and Sargent's (2001) model, the deviation between the worst-case model and the benchmark is purely in the distribution of innovations, ε_{t+1}. Here we explicitly focus on uncertainty about dynamics. As we showed above, shifts in the mean of the distribution of ε_{t+1} (as obtained in Barillas, Hansen, and Sargent (2009)) represent deviations in the model only at frequency zero. Our nonparametric analysis allows for deviations at all frequencies. The allowance of models with alternative dynamics is central to our results – it is what generates predictability in returns and excess volatility in asset prices.

Second, Hansen and Sargent (2001) model agents as having log utility over fixed models, while we start from the assumption that agents have Epstein–Zin (1991) preferences over fixed models. Because Epstein–Zin preferences imply that the timing of resolution of uncertainty affects utility, our agents view more persistent processes as less favorable (for $\alpha > 1$).

Finally, under multiplier preferences, the state variables in the worst-case model are the same as the state variables in the benchmark model. So, for example, if consumption growth is an AR(1) under the benchmark model, then lagged consumption growth is also the only state variable under the worst-case model. In our setting, though, the state variables in the worst-case model are typically different from those under the benchmark (and in general in fact include the entire history of consumption growth). Our setting thus allows agents to consider potentially infinitely richer economic dynamics.

---

[14]Since consumption can be non-stationary, this expectation does not always exist. In that case, we simply rescale lifetime utility by the level of consumption yielding $E\left[v\left(\Delta c^t; b \right) - c_t |\Theta \right]$, which does exist. Scaling by consumption is a normalization that has no effect other than to ensure the expectation exists.

A natural question is why we analyze a worst case instead of allowing the agent to average as a Bayesian across all the possible models. Our answer is that people may not actually be Bayesians, or they may not be able to assign priors to all models. Machina and Siniscalchi (2014) discuss the extensive experimental evidence that people make choices consistent with ambiguity aversion.

Ambiguity aversion is particularly compelling in our context because, as we will see, it is ultimately dynamics at the very lowest frequencies that drive our results. And direct estimation of, say, 50- or 100-year autocorrelations is, for practical purposes with realistic data sources, impossible. So investors face a situation where they simply do not have data that directly measures all features of consumption dynamics. They do not face a standard estimation problem – rather, they must make decisions in the face of model uncertainty that cannot be resolved by the data at hand, a problem akin to that discussed by Knight (1921) and Ellsberg (1961). Moreover, as Hansen and Sargent (2007) note, almost any model is necessarily just an approximation, even a high-order one. If the true model has infinite order, then it can never be fully characterized in any finite sample.

Finally, (again, following Hansen and Sargent (2007)), it is well understood that when the parameter space is infinite, likelihood-based methods are difficult to implement at best, and often impossible. Specifically, frequentist methods with an infinite-dimensional parameter space lead to degenerate estimates (in the context of spectral estimation, for example, see Chow and Grenander (1985)), while constructing priors over such a space that lead to accurate posterior confidence intervals is difficult or impossible (Sims (1971) and Diaconis and Freedman (1986)). So when an investor does not know the true order of the model driving the endowment process, she may literally not be able to assign likelihoods to different specifications. Instead, she uses $g\left(\Theta; \bar{\Theta}\right)$ to measure the "plausibility" of potential models, even though she has no probabilities on models over which to integrate.

Ultimately, our ambiguity-averse investor's utility takes the form of that of an Epstein–Zin agent but using Θ^w to form expectations about future consumption growth,[15]

$$v^w\left(\Delta c^t\right) = v\left(\Delta c^t; \Theta^w\right) = c_t + \frac{\beta}{1-\beta}\frac{1-\alpha}{2}\sigma_w^2 b^w\left(\beta\right)^2 + \sum_{k=1}^{\infty}\beta^k E_t\left[\Delta c_{t+k}|\Theta^w\right] \qquad (15)$$

In modeling investors as choosing a single worst-case Θ^w, we obtain a setup similar to Gilboa and Schmeidler (1989), Maccheroni, Marinacci, and Rustichini (2006), and Epstein and Schneider (2007) in the limited sense that we are essentially constructing a set of models and minimizing over that set. Our worst-case model is, however, chosen once and for all and is not state- or choice-dependent. The choice of Θ^w is timeless – it is invariant to the time-series evolution of consumption

___

[15]Note that since utility is recursive, the agent's preferences are time-consistent, but under a pessimistic probability measure. Furthermore, the assumption that b^w is chosen unconditionally means that b^w is essentially unaffected by the length of a time period, so the finding in Skiadas (2013) that certain types of ambiguity aversion become irrelevant in continuous time does not apply here.

– so what it represents is an unconditional worst-case model: if an agent had to choose a worst-case model to experience prior to being born into the world, it would be Θ^w. The worst-case analysis is certainly not the only way to model behavior under ambiguity, but it is plausible and generates economically interpretable results.

Unlike in some recent related papers, the investor in this model does not change her probability weights every day or adjust the worst case according to a learning process.[16] She chooses a single pessimistic model to protect against. An added benefit of the assumption that the worst-case model is chosen timelessly is that it ensures time-consistency. The agent sets the model used for forming expectations once and for all, and then uses a recursive utility specification conditional on that single probability measure. In other words, since the model is chosen a single time, our agent's preferences inherit the time-consistency of Epstein–Zin preferences.

4   The worst-case scenario

Our analysis above leads us to a simple quadratic optimization problem. The solution is summarized in the following proposition.

Proposition 1 *Under assumptions 1–4, for an agent who chooses a model $\Theta^w \equiv \{b^w, \mu^w, \sigma_w^2\}$ to minimize the unconditional expectation of lifetime utility subject to the loss function $g\left(\Theta; \bar{\Theta}\right)$, that is,*

$$\Theta^w = \arg\min_{\Theta} \frac{\beta}{1-\beta} \frac{1-\alpha}{2} \sigma^2 b\left(\beta\right)^2 + \frac{\beta}{1-\beta} \mu + \lambda g\left(\Theta; \bar{\Theta}\right) \tag{16}$$

the worst-case model is determined by the set of equations

$$\textit{Dynamics:} \quad \sigma_w^2 \left|B^w\left(\omega\right)\right|^2 = \bar{f}\left(\omega\right) \tag{17}$$

$$+\lambda^{-1}\beta\left(1+\beta\right)\left(\alpha-1\right)\sigma_w^2 b^w\left(\beta\right)^2 \bar{f}\left(\omega\right)\left|Z\left(\omega\right)\right|^2 \tag{18}$$

$$\textit{Mean:} \quad \mu_w = \bar{\mu} - \lambda^{-1}\frac{\beta}{1-\beta}\bar{f}\left(0\right) \tag{19}$$

where $B\left(\omega\right) = b\left(e^{i\omega}\right)$, $Z\left(\omega\right) \equiv \sum_{j=0}^{\infty} \beta^j e^{i\omega j}$, $\sigma_w^2 \left|B^w\left(\omega\right)\right|^2$ is the spectrum under the worst-case model, and $\bar{f}$ is the benchmark spectrum. The time-domain model $b^w\left(L\right)$ has coefficients b_j^w that are obtained as the Wold representation associated with $\left|B^w\left(\omega\right)\right|^2$ (see Priestley (1981) section 10.1.1).

We thus have a closed form expression for the worst-case model. The spectral density under the worst case, $f^w\left(\omega\right) = \sigma_w^2 \left|B^w\left(\omega\right)\right|^2$, in (17) is equal to the point estimate plus a term that depends

---

[16]See, for example, Hansen and Sargent (2010), Ju and Miao (2012), and Collin-Dufresne, Johannes, and Lochstoer (2015).

Figure 1 about here.

on three factors. First, $\lambda^{-1}\beta(1+\beta)(\alpha-1)\sigma_w b^w(\beta)$ represents the ratio of the utility losses from a marginal increase in $\sigma_w b^w(\beta)$ to the cost of deviations from the point estimate, λ. When risk aversion, α, is higher or the cost of deviating from the point estimate, λ, is lower, the worst-case model is farther from the point estimate.

Second, $\bar{f}(\omega)$ represents the amount of uncertainty the agent has about consumption dynamics at frequency ω. Where the spectral density, $\bar{f}(\omega)$, is high, there is relatively more uncertainty and the worst-case model is farther from the point estimate.

Finally, $|Z(\omega)|^2$ determines how much weight the lifetime utility function places on frequency ω. The top panel of figure 1 plots $|Z(\omega)|^2$ for $\beta = 0.99$, a standard annual calibration. It is strikingly peaked near frequency zero; in fact, the x-axis does not even show frequencies corresponding to cycles lasting less than 10 years because they carry essentially zero weight. Since the mass of $|Z(\omega)|^2$ lies very close to frequency 0, the worst case shifts power to very low frequencies. In that sense, the worst-case model *always* includes long-run risk.

The mean of consumption growth also differs from the benchmark in a natural way: when people are willing to consider more extreme models, they are more risk averse, or they have more uncertainty about the mean (through $\bar{f}(0)$), μ_w is farther below μ.

When $\alpha = 1$, so that the investor's preferences reduce to time-separable log utility, the shift in the dynamics is set to zero: $f^w(\omega) = \bar{f}(\omega)$. That is because under time separable preferences, the dynamics of consumption do not affect average utility. But the mean growth rate obviously still matters, so even for $\alpha = 1$, $\mu^w < \bar{\mu}$.

Finally, people in principle also consider deviations of σ_w^2 from the point estimate $\bar{\sigma}^2$. We find below that this effect is extremely small in a typical calibration.

An important feature of the results is that the worst-case model depends on preferences. In particular, when β is higher – the investor is more patient – the worst-case model is more persistent. At the same time, when the investor is more risk averse – α is higher – the amount of long-run risk in the worst-case model is higher. The model thus has implications for disagreement across agents that is tightly linked to preferences or investment horizons.

Proposition 1 represents the completion of the solution to the model. To summarize, given a point estimate, $\bar{\Theta}$ (estimated from a finite-order model that we need not specify here), the agent selects a worst-case model Θ^w. She then uses the worst-case model when calculating expectations and pricing assets.

4.1 Long-run risk is the worst-case scenario

Corollary 2 *Suppose the agent's point estimate is that consumption growth is white noise, with $\bar{b}(L) = 1$. The worst-case model is then an ARMA(1,1), and consumption growth has the following representation under the worst-case dynamics,*

$$
\begin{align}
\Delta c_t &= (1-\beta)\mu_w + \beta\Delta c_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1} \tag{20} \\
\varepsilon_t &\sim N(0, \sigma_w^2) \tag{21}
\end{align}
$$

The above ARMA(1,1) process also has an equivalent state-space representation,

$$
\begin{align}
\Delta c_t &= \mu_w + x_{t-1} + \eta_t \tag{22} \\
x_t &= \beta x_{t-1} + v_t \tag{23}
\end{align}
$$

where

$$
\begin{align}
\eta_t &\sim N(0, \bar{\sigma}^2) \text{ and } \nu_t \sim N(0, \bar{\sigma}^2\varphi) \tag{24} \\
\varphi &\equiv \lambda^{-1}\beta(1+\beta)(\alpha-1)b^w(\beta)^2\sigma_w^2 \tag{25}
\end{align}
$$

The state-space form in equations (22–23) is observationally equivalent to the process (20–21) in the sense that they have identical autocovariances for consumption growth, and (22–23) is exactly case I from Bansal and Yaron (2004), the homoskedastic long-run risk model.

The worst-case process exhibits a small but highly persistent trend component, and the persistence is exactly equal to the time discount factor. Intuitively, since β^j determines how much weight in lifetime utility is placed on consumption j periods in the future, a shock that decays with β spreads its effects as evenly as possible across future dates, scaled by their weight in utility. And spreading out the effects of the shock over time minimizes its detectability. The worst-case/long-run risk model is thus the departure from pure white noise that generates the largest increase in risk prices (and decrease in lifetime utility) for a given level of statistical distinguishability.[17]

There is a difference between the worst-case model derived here and the long-run risk model, which is that in this setting, the long-run trend, x_t, is unobservable. Asset prices therefore carry no ability to forecast future consumption growth beyond what is available in the history of lagged consumption. However, one may show that the volatility of the pricing kernel, and hence the price

---

[17]This exact balance results from the quadratic nature of the minimization problem. All the lag coefficients b_j contribute to the KL divergence symmetrically, but their effects on utility decline with β^j, which yields our result. Specifically, the contribution of dynamics to lifetime utility is proportional to $b^w(\beta)^2$, and $\frac{d}{db_j}b(\beta)^2 \propto \beta^j$. Similarly the contribution to the KL divergence from the $\{b_j\}$ is $\sum_{j=1}^{\infty} b_j^2$, so $\frac{dg(\Theta;\bar{\Theta})}{db_j} \propto b_j$. So we end up with the result that $b_j \propto \beta^j$.

of risk, is actually higher under the process (20–21) than under (22–23) (due to the later arrival of information). The key intuition behind the long-run risk model is that a small persistent component in consumption growth can induce large fluctuations in the pricing kernel, and that force (or fear of it) is also present in our model.

The composite parameter φ determines the degree to which the worst-case dynamics differ from the point estimate and also the volatility of the trend shock ν. When the worst-case model differs from the benchmark by a greater degree, the worst-case model displays more volatile trend-type shocks, ν_t, and the white-noise shocks, η_t, become relatively smaller.

The bottom panel of figure 1 plots the spectrum for the white-noise benchmark and the worst-case model. The spectrum for white noise is totally flat, while the worst case has substantial power at the very lowest frequencies, exactly as we would expect from the top panel of figure 1.

Hansen and Sargent (2010) also study a setting in which investors price assets as though the long-run risk model might be driving the consumption process. The key difference between their analysis and ours, though, is that they start from the assumption that agents put a non-zero probability on possibility that the long-run risk model is actually true. We, on the other hand, start from a simple white-noise benchmark and obtain the long-run risk model entirely endogenously.

5   The behavior of asset prices

The investor's Euler equation is calculated under the worst-case dynamics. For any available return R_{t+1},

$$1 \quad = \quad E_t\left[R_{t+1}M_{t+1}|\Theta^w\right] \tag{26}$$

$$\text{where } M_{t+1} \quad \equiv \quad \beta \exp\left(-\Delta c_{t+1}\right)\frac{\exp\left(v\left(\Delta c^{t+1};\Theta^w\right)\times(1-\alpha)\right)}{E_t\left[\exp\left(v\left(\Delta c^{t+1};\Theta^w\right)\times(1-\alpha)\right)|\Theta^w\right]} \tag{27}$$

M_{t+1} is the stochastic discount factor (SDF). The SDF is identical to what is obtained under Epstein–Zin preferences, except that here expectations are calculated under Θ^w. The key implication of that change is that expected shocks to $v\left(\Delta c^{t+1};\Theta^w\right)$ have a larger standard deviation since the worst-case model features highly persistent shocks that affect lifetime utility much more strongly than the less persistent point estimate.

5.1   Consumption and dividend claims

It is straightforward, given that log consumption follows a linear Gaussian process, to derive approximate expressions for prices and returns on levered consumption claims. We consider an asset whose dividend is C_t^γ in every period, where γ represents leverage. Denote the return on that asset

on date $t + 1$ as r_{t+1} and the real risk-free rate as $r_{f,t+1}$. We will often refer to the levered consumption claim as an equity claim, and we view it as a simple way to model equity returns (Abel (1999)).[18]

From the perspective of an econometrician who has the same point estimate for consumption dynamics as the investor, $\bar{\Theta}$, the expected excess log return on the levered consumption claim is

$$
\begin{aligned}
E_t \left[r_{t+1} - r_{f,t+1} | \bar{\Theta} \right] + \frac{1}{2} var_t \left(r_{t+1} \right) &= -cov^w \left(r_{t+1}, m_{t+1} \right) + \frac{1}{2} \left(var_t \left(r_{t+1} \right) - var_t^w \left(r_{t+1} \right) \right) \\
&+ \frac{\gamma - \delta a^w \left(\delta \right)}{1 - \delta a^w \left(\delta \right)} \left[\begin{array}{c} \left(1 - a^w \left(1 \right) \right) \left(\mu - \mu^w \right) + \\ \left(\bar{a} \left(L \right) - a^w \left(L \right) \right) \left(\Delta c_t - \mu \right) \end{array} \right]
\end{aligned}
\tag{28}
$$

where δ is a linearization parameter from the Campbell–Shiller approximation that depends on the steady-state price/dividend ratio. The addition of the variance is a correction to make the left-hand side approximately the arithmetic mean return.

The first term, $-\frac{1}{2} cov_t^w \left(r_{t+1}, \log M_{t+1} \right)$ (i.e. the conditional covariance of log returns with the log SDF measured under the worst-case dynamics), is the standard risk premium, and it is calculated under the worst-case model. The primary way that the model increases risk premia compared to standard Epstein–Zin preferences is that the covariance of the return with the SDF is more negative. That covariance, in turn, is more negative for two reasons. First, since the agent behaves as though shocks to consumption growth are highly persistent, they have large effects on lifetime utility, thus making the SDF very volatile. Second, again because of the persistence of consumption growth under the worst case, shocks to consumption have large effects on expected long-run dividend growth, so the return on the levered consumption claim is also very sensitive to shocks.[19] These two effects cause the consumption claim to strongly negatively covary with the SDF and generate a high risk premium.

The second term is a quantitatively trivial (under our benchmark calibration) adjustment to the arithmetic mean due to the difference between the variance of returns under the benchmark and worst-case models.

The second line is the part of the expected return that comes from differences in forecasts of consumption under the models of the econometrician and investors. As long as $\gamma > \delta a^w \left(\delta \right)$ (which is always satisfied if $\gamma > 1$), the coefficients on $\left(\mu - \mu^w \right)$ and $\left(\bar{a} \left(L \right) - a^w \left(L \right) \right) \left(\Delta c_t - \mu \right)$

---

[18]The model of dividends here is very simplistic and does not match the dynamics of dividend growth particularly well. The online appendix examines a more realistic model of dividends that matches the standard deviation of dividend growth and the correlation between dividend and consumption growth in a setting in which dividends and consumption are cointegrated. The behavior of equity returns is almost identical in that setting to what we obtain with the simple model $D_t = C_t^\gamma$.

[19]We have $cov^w \left(r_{t+1}, m_{t+1} \right) = \left(-1 + \left(1 - \alpha \right) b^w \left(\beta \right) \right) \frac{\gamma - \delta a^w \left(\delta \right)}{1 - \delta a^w \left(\delta \right)} \sigma_w^2$. The volatility of the SDF is determined by $\left(-1 + \left(1 - \alpha \right) b^w \left(\beta \right) \right) \sigma_w$, while the volatility of the return is $\frac{\gamma - \delta a^w \left(\delta \right)}{1 - \delta a^w \left(\delta \right)} \sigma_w$.

are positive. The former term implies that average equity returns are higher when investors are more pessimistic about mean consumption (and hence dividend) growth, μ^w.

The term involving $(\Delta c_t - \mu)$ is zero on average, but it induces predictability in returns. When the worst case implies higher future consumption growth, investors pay relatively more for equity compared to riskless assets, thus lowering expected excess returns. This channel leads to procyclical asset prices and countercyclical expected returns when $a^w(L)$ implies more persistent dynamics than $\bar{a}(L)$, similarly to Fuster, Hebert, and Laibson (2011).

The procyclicality of asset prices also depends on leverage, γ. Expected returns for assets more risky than consumption ($\gamma > 1$) are countercyclical, and we model equity as having $\gamma > 1$. But for assets safer than consumption, prices will in fact be countercyclical, because variation in the risk-free rate dominates variation in the risk premium for them.

We also note that since risk aversion and conditional variances are constant, the excess return on a levered consumption claim has a constant conditional expectation from the perspective of investors. That is, while returns are predictable from the perspective of an econometrician, investors believe that they are unpredictable. So if investors in this model are surveyed about their expectations of excess returns, their expectations will not vary, even if econometric evidence implies that returns are predictable. This model therefore generates extrapolative expectations of the type discussed by Greenwood and Shleifer (2014) (and many references therein), but that extrapolation applies to consumption growth rather than equity returns.

Finally, it is worth noting that since asset prices depend purely on the history of consumption growth, there is no way to improve to the investor's estimates of consumption dynamics by including information on asset prices. The price of a consumption claim is completely redundant given data on the history of consumption.

5.2 Interest rates

The risk-free rate follows

$$r_{f,t+1} = -\log\beta + \mu^w + a^w(L)(\Delta c_t - \mu^w) - \frac{1}{2}\sigma_w^2 + (1-\alpha)b^w(\beta)\sigma_w^2 \qquad (29)$$

With a unit EIS, interest rates move one for one with expected consumption growth. In the present model, the relevant measure of expected consumption growth is $\mu^w + a^w(L)(\Delta c_t - \mu^w)$, which is the expectation under the worst-case model.

The online appendix derives analytic expressions for the prices of long-term zero-coupon bonds, which we discuss in our calibration below.

6 Calibration

We now parameterize the model to analyze its quantitative implications. Most of our analysis is under the assumption that the agent's point estimate implies that consumption growth is white noise and that the point estimate is also the true dynamic model. Despite this parsimony, we obtain striking empirical success in terms of matching important asset pricing moments.

Many of the required parameters are standard. We use a quarterly calibration of $\beta = 0.99^{1/4}$, implying a pure rate of time preference of 1 percent per year. Setting β smaller would imply that the worst-case model is less persistent, but it also implies that interest rates are higher. We will see below that the equilibrium real interest rate in the model ends up being still too large relative to the data, which implies that β should not be lower.

The steady-state dividend/price ratio used in the Campbell–Shiller approximation is 5 percent per year, as in Campbell and Vuolteenaho (2004), so $\delta = 0.95^{1/4}$.[20] Other parameters are calibrated to match moments reported in Bansal and Yaron (2004). The agent's point estimate is that consumption growth is i.i.d. with a quarterly standard deviation of 1.47 percent, which we also assume is the true data-generating process. Finally, the leverage parameter for the consumption claim, γ, is set to 4.806 to generate mean annualized equity returns of 6.33 percent.

We calibrate λ to equal 106.8 to match the observed Sharpe ratio on equities. The calibration of α is slightly more difficult. Its effects on asset prices are highly similar to those of λ, so it is not straightforward to calibrate both α and λ from aggregate data.

It is possible to draw a link between α and λ if we interpret α as in the literature on multiplier preferences. Hansen and Sargent (2005) and Barillas, Hansen, and Sargent (2009) show that the coefficient of relative risk aversion in Epstein–Zin preferences – our α – can be interpreted as a measure of the agent's willingness to consider alternative distributions of the innovations ε, where the measure of distance between distributions is again the KL divergence. That is, our combination of model uncertainty over Θ with Epstein–Zin preferences can alternatively be viewed as a combination of two layers of model uncertainty – over Θ and the distribution of ε separately. The appendix formalizes this argument and shows that if agents use the same penalty on the KL divergence over the distribution of ε and also the model Θ, then α is linked to λ through the formula

$$\alpha = 1 + \frac{1}{\lambda \left(1 - \beta\right)} \tag{30}$$

yielding $\alpha = 4.73$.

The link between α and λ in equation (30) is useful because it eliminates a degree of freedom from our calibration. In order to justify it, though, one must accept the assumptions of multiplier

---

[20]Setting this parameter to a higher value (e.g. $0.975^{1/4}$ as in Bansal and Yaron (2004)) increases the mean and standard deviation of equity returns but has no effect on the Sharpe ratio.

preferences, which include that agents have unit risk aversion over unambiguous risks. So while (30) is helpful, it is also only a very special case. Alternatively, α can simply be taken as another free preference parameter, rather than as a measure of robustness (in the multiplier preference interpretation). So it is important that our calibration of α be objectively reasonable as a measure of risk aversion. Risk aversion of 4.73 is extraordinarily small in the context of the consumption-based asset pricing literature with Gaussian innovations. It well within the upper bound of 10 proposed by Mehra and Prescott (1985), and almost precisely equal to average risk aversion measured by Barsky et al. (1997).[21] α therefore takes on a plausible value in its own right, separate from any connection it may have to λ.

To further investigate how reasonable λ is, in the next section we show that it implies a worst-case model that is rarely rejected by statistical tests on data generated by the true model. An investor with the true model as her point estimate might reasonably believe the worst case could have actually generated the data that led to that point estimate.

7   Quantitative implications

7.1   The white noise case

We report the values of the parameters in the worst-case consumption process in table 1. As noted above, the autocorrelation of the predictable part of consumption growth under the worst-case model is β, implying that trend shocks have a half-life of 70 years, as opposed to the three-year half-life in the original calibration in Bansal and Yaron (2004).[22] $b^w(\beta)\sigma_w$, the relevant measure of the total risk in the economy, is 0.036 at the quarterly frequency in our model, compared to 0.031 in theirs.[23] Our model thus has more persistence and more total risk.

Note also that σ_w is only 0.4 percent larger than $\bar{\sigma}$. So the conditional variance of consumption growth under the worst-case model is essentially identical to that under the benchmark. However, because the worst-case model is so persistent, $b^w(\beta)$ is 2.45 times higher than $\bar{b}(\beta)$, thus implying that the worst-case economy is far riskier than the point estimate.

---

[21]Average here is the haromic mean – the inverse of the arithmetic mean of risk tolerance. In an update and extension of Barsky et al.'s (1997) analysis, Kimball, Sahm, and Shapiro (2008) estimate similar values.

Our value for α is less than half that used by Bansal and Yaron (2004), who themselves are notable for using a low value, and it is similar to that of Barro (2006), who uses values of 2 to 4 in a model with rare disasters.

[22]The 70-year half-life is sensitive to the choice of β. It is equal to $\log(1/2)/\log(\beta)$. If $\beta = 0.975^{1/4}$, the half-life falls to 27 years. Nevertheless, the worst-case is robustly far more persistent than Bansal and Yaron's (2004) trend – to obtain a half-life of 3 years, we would need an annual rate of time preference of 21 percent. Online appendix table A3 reports results under a calibration with $\beta = 0.95^{1/4}$.

[23]Using the notation for the long-run risk model from above, and denoting the AR(1) coefficient for x as ρ, we calculate the equivalent of $b(\beta)$ in Bansal and Yaron's (2004) model as $b(\beta) = \sqrt{\frac{var(\nu)}{(1-\beta\rho)^2} + var(\eta)}$ which yields $b(\beta) = 0.029$. Dew-Becker (2015) discusses this calculation and its relation with risk premia in more detail.

Table 1 about here.

7.1.1   Unconditional moments

Table 1 reports key asset pricing moments. The first column shows that the model can generate a high standard deviation for the pricing kernel (and hence a high maximal Sharpe ratio), high and volatile equity returns, and low and stable real interest rates, as in the data. The equity premium and its volatility are 6.33 and 19.42 percent respectively, identical to the data. The real risk-free rate has a mean of 1.89 percent and a standard deviation of 0.31 percent.

The second column in the bottom section of table 1 shows what would happen if we set $\lambda = \infty$ but held α fixed at 4.73, so that we would be back in the standard Epstein–Zin setting where there is no uncertainty about dynamics. The equity premium then falls from 6.33 to 1.95 percent, since the agent exhibits no concern for long-run risk. Furthermore, because the agent no longer behaves as if consumption growth is persistent, a shock to consumption has far smaller effects on asset prices. The standard deviation of returns falls from 19.4 to 14.1 percent and the standard deviation of the price/dividend ratio falls from 19 percent to exactly zero. The agent's fear of a model with long-run risk thus raises the mean of returns by a factor of more than 3 and the volatility by a factor of 1.4.

Recall from equation (28) that the mean equity premium is due both to the investor's belief about the covariance of returns with the pricing kernel – which depends on the worst-case dynamics, b^w – and also the difference between the true and worst-case mean levels of consumption growth, $\bar{\mu} - \mu^w$. Of the 633 basis point equity premium, 54 basis points come from the mean effect, while 579 come from the dynamics. So uncertainty about mean consumption growth is relatively unimportant in driving the equity premium. That result is driven by the fact that the worst-case dynamics have two compounding effects on the equity premium – they make both the pricing kernel and equity returns more volatile, interacting to generate a large increase in the equity premium.[24]

Going back to the first column, we see that there are large and persistent movements in the price/dividend ratio in our model. The one-year autocorrelation of the price/dividend ratio at 0.96 is somewhat higher than the empirical autocorrelation, while the standard deviation is 0.19, similar to the empirical value of 0.29. These results are particularly notable given that there is no free parameter that allows us to directly match the behavior of prices.

Volatility in the price/dividend ratio has the same source as the predictability in equity returns discussed above: the agent prices assets under a model where consumption growth has a persistent component. So following positive shocks, she is willing to pay relatively more, projecting that dividends will continue to grow in the future. From the perspective of an econometrician, these

---

[24]Collin-Dufresne, Johannes, and Lochstoer (2015) also find that uncertainty about dynamics is relatively more important for the equity premium than uncertainty about mean consumption growth.

Figure 2 about here.

movements seem to be entirely due to discount-rate effects: dividend growth is entirely unpredictable, since dividends are a multiple of consumption, and consumption follows a random walk. On the other hand, from the perspective of the investor (or her worst-case model), there is almost no discount-rate news. Rather, she prices the equity claim differently over time due to variation in forecasts of future cash flows.

The bottom row of table 1 reports average gap between the yields on real 1- and 10-year zero-coupon bonds. The term structure is very slightly downward-sloping in the model, a feature it shares with Bansal and Yaron's (2004) results. The downward slope is consistent with the long sample of inflation-indexed bonds from the UK reported in Evans (1998). A thorough analysis of the implications of our model for the term structure of interest rates is beyond the scope of this paper, but we simply note that the implications of the model for average yields are not wildly at odds with the data and are consistent with past work.

A final feature of the data that papers often try to match is the finding that interest rates and consumption growth seem to be only weakly correlated, suggesting that the EIS is very small. Since consumption growth in this model is unpredictable by construction, standard regressions of consumption growth on lagged interest rates that are meant to estimate the EIS, such as those in Hall (1988) and Campbell and Mankiw (1989), will generate EIS estimates of zero on average.

7.1.2   Return predictability

To quantify the degree of predictability in returns, figure 2 plots percentiles of sample R^2s from regressions of returns on price/dividend ratios in 360-quarter samples (the length of the empirical sample). The gray line is the set of corresponding values from the data between 1926 and 2015. We report R^2s for horizons of 1 quarter to 10 years. At both short and long horizons the model matches the data well. The median R^2 from the predictive regressions at the 10-year horizon is 31 percent, while in the data it is 25 percent. This is in contrast with the complete lack of predictability (reflecting the i.i.d. nature of the true data generating process) of consumption and dividend growth.

Beeler and Campbell (2012) argue that cash flows are excessively predictable in Bansal and Yaron's (2004) calibration. In our setting, there is precisely zero cash-flow predictability by construction. Our results are thus consistent with evidence from various sources on the excess volatility in aggregate asset prices (Leroy and Porter (1981), Shiller (1981), Cochrane (2008)).

In fact, our results are consistent with all three of Beeler and Campbell's (2012) criticisms of the long-run risk model: consumption and dividends are not predicted with asset prices, asset return volatility is not predicted by asset prices, and consumption growth is not predicted by interest

rates. In other words, we obtain the major intuition and results of the long-run risk model – that investors fear low-frequency fluctuations in consumption growth and thus demand large risk premia on stocks – but without relying on significant predictability in consumption, dividends, or volatility.

Of course, we obtain these results by exploiting the wedge between the true data generating process and the one that informs investors' evaluation of risky payoffs. But, as we show below, this wedge is empirically reasonable, in the sense that the worst-case model is a perfectly reasonable process for a person to believe generated the data, given, say, a century of data.

7.1.3 Probability of rejecting the worst-case dynamics

For our calibration of λ to be intuitively reasonable, the worst-case model should be thought plausible by the agent. One way of interpreting this statement is that the worst-case model should fit a sample of data generated by the true model nearly as well as the true model itself.

We consider two tests of the fit of the worst-case model to the true white-noise consumption process: Ljung and Box's (1978) portmanteau test and the likelihood-based test of an ARMA(1,1) suggested by Andrews and Ploberger (1996).[25] The likelihood-based test is in fact a correctly specified likelihood-ratio test and thus should be asymptotically most powerful. To test that the worst-case model is the correct specification, we take a simulated sample of consumption growth, Δc_t, and construct artificial residuals,

$$\varepsilon_t^{\Theta^w} \equiv \left(\Delta c_t - \mu_w - a^w \left(L \right) \left(\Delta c_{t-1} - \mu_w \right) \right) \left(\sigma_w \right)^{-1} \tag{31}$$

Under the null that the worst-case model is the correct specification, $\varepsilon_t^{\Theta^w}$ should be white noise. The Ljung–Box and Andrews–Ploberger tests both ask whether that null can be rejected. Since consumption growth is generated as white noise, $\varepsilon_t^{\Theta^w}$ is in fact not i.i.d.. In a sufficiently large sample, an investor will be able to reject the hypothesis that consumption was driven by the worst-case model by observing that $\varepsilon_t^{\Theta^w}$ is serially correlated. We obtain small-sample critical values for the two test statistics by simulating their distributions under the null.

The top section of table 2 reports the probability that the agent would reject the hypothesis that consumption growth was driven by the worst-case model after observing a sample of white-noise consumption growth. We simulate the tests in both 50- and 100-year samples. In all four cases, the rejection probabilities are only marginally higher than they would be if the null hypothesis were actually true. The ARMA(1,1) test is the stronger of the two, with rejection rates of 5.5 and 6.3

---

[25]The intuition behind this approach is similar to that underpinning the detection error probability (DEP) calculations of Barillas, Hansen, and Sargent (2009), which are widely used to calibrate robustness models. Although we do not report them here, the DEPs in our case also indicate that the worst-case and benchmark models are difficult to distinguish.

Table 2 about here.

percent in the 50- and 100-year samples, respectively, while the Ljung–Box test performs slightly worse, with rates of 4.6 and 5.1 percent. The online appendix reports further results using other statistical tests and longer samples.[26]

Table 2 thus shows that the worst-case model, while having economically large differences from the point estimate in terms of its asset pricing implications, can barely be distinguished from the point estimate in long samples of consumption growth. From a statistical perspective, it is entirely plausible that an investor would be concerned that the worst-case model could be what drives the data. Thus both λ and α (which were calibrated jointly with only a single degree of freedom) take on independently reasonable values.

Furthermore, the extreme difficulty of distinguishing the benchmark and worst-case models with long samples of data also suggests that our decision not to model learning explicitly is likely relatively innocuous. Although allowing for learning is conceptually desirable, any learning about low frequency properties of the model would still apply to frequencies far shorter than those that our investor emphasizes under the worst case, as determined by $Z(\omega)$.

Since risk aversion and the conditional volatility of consumption growth are constant, stock returns should not be predictable if the worst-case model is true. So another potential way to test the worst-case model would be to test for return predictability. Since returns are a linear function of current and past consumption growth, though, they contain no information not contained in the history of consumption growth. We focus on formal statistical tests, like the likelihood ratio, because they should use that information efficiently.

Stock prices and returns depend on the low-frequency characteristics of consumption growth, though – exactly where the worst-case deviates from the benchmark. So we might expect them to provide a stronger test. As an example, consider simulations of the R^2s for forecasts of five-year excess equity returns in 100-year samples. The average R^2 when the data is generated by the white-noise model but agents price under the worst-case is 0.19. The average R^2 when the data is generated by the worst-case model – in which case the equity premium is constant – is 0.13. So our ambiguity aversion substantially increases predictability (to the point that the model matches the data well). But the difference is small relative to the dispersion in R^2s: in only 14.7 percent of samples does the R^2 allow the agent to reject the worst-case at the 95-percent level.[27] So even with a test that focuses on the frequencies where the worst-case model deviates from the benchmark most, with a century of data, the agent would still statistically reject the worst-case

---

[26]The under-rejection in small samples comes from the fact that $\varepsilon_t^{\Theta^w}$ is negatively autocorrelated when the data is driven by the white-noise benchmark. The online appendix explores rejection probabilities for a larger set of test statistics and with samples up to 1000 years long.

[27]That is, the R^2 for data generated under the benchmark is above the 95th percentile of the R^2 for data generated by the worst-case in 14.7 percent of the samples.

very infrequently.

7.1.4 Alternative calibrations of the pricing model

We derive the worst-case model endogenously, but similar models have also been assumed for investor expectations. Bansal and Yaron (2004) argue that a model with trend shocks with a quarterly persistence 0.94 fits the data well. Hansen and Sargent (2010) consider a setting where investors focus on two models for consumption growth, the more persistent of which has a trend component with an autocorrelation of 0.99. Due to ambiguity aversion in their model, asset prices are primarily driven by the more persistent model.

The bottom section of table 2 examines how rejection probabilities change if we modify the pricing model to use a less persistent trend. In all rows we hold $b^w(\beta)$ fixed and simply modify the persistence of consumption growth under the pricing (null) model. In other words, we ask how easy different models are to distinguish from white noise, holding constant the relevant measure of risk and varying the persistence of shocks.

The top row is the calibration from the main analysis, where persistence is equal to the time discount factor. As the degree of persistence falls, the investor's ability to reject the pricing model in a century-long sample rapidly improves. When the persistence is 0.99, as in Hansen and Sargent (2010), the pricing model is rejected 13.9 percent of the time – twice as often as our endogenous worst-case model. When the persistence falls to 0.94 as in Bansal and Yaron (2004), the pricing model is rejected 82.4 percent of the time. The result that the persistence of the worst-case model should be equal to β is clearly key to ensuring that the model is difficult to reject in simulated data.

7.2 Estimated consumption dynamics

We now examine the worst-case scenario associated with an estimated small-scale model of consumption growth. The above analysis assumes that the point estimate for consumption dynamics is that consumption growth is i.i.d.. We now relax that assumption and examine a more sophisticated estimation framework.

We study quarterly data on per capita non-durables and services consumption in the United States. The Bayesian information criterion leads us to the choice of an ARMA(1,1) from a range of small-scale ARMA models (up to an ARMA(5,5)). A key feature of the estimate is that the spectral density $\bar{f}(\omega)$ varies across frequencies, which allows us to ask whether variation in estimation uncertainty across frequencies is qualitatively or quantitatively relevant in determining the worst-case model.

We concentrate on the worst-case choice of dynamics and abstract from distortions to the innovation mean and variance. We first consider what worst-case model the agent would derive if she

Figure 3 about here.

were constrained to minimize utility with respect to a transfer function implied by an ARMA(1,1). That is, utility is minimized by choosing a worst-case $\{\rho, \theta\}$ in the model

$$\Delta c_t = \mu + \rho \left(\Delta c_{t-1} - \mu \right) + \varepsilon_t - \theta \varepsilon_{t-1} \tag{32}$$

in order to minimize lifetime utility subject to the penalty $\lambda g \left(b; \bar{b} \right)$. We denote this "parametric" worst case as b^p. The restriction that the worst case only allows changes in the parameters ρ and θ is typical in the literature; it assumes that investors know the model driving consumption growth and they need only estimate its parameters.[28] We compare b^p (which depends only on $\{\rho^p, \theta^p\}$) to the b^w obtained following the main analysis above (i.e. allowing an arbitrary b^w, instead of just an ARMA(1,1)).

In figure 3 we plot the real part of the transfer functions implied under the benchmark and the two worst cases. Except at low frequencies, the parametric worst case is essentially indistinguishable from the point estimate, while the nonparametric worst case differs much more dramatically. Intuitively, since there are only two free parameters in the restricted parametric problem, it is impossible to generate the deviations very close to frequency zero that have both high utility cost and low detectability. So, instead of large deviations on a few frequencies, as in the nonparametric case, the parametric worst case puts very small deviations on a wide range of frequencies.

The specific parameters in the benchmark and parametric worst-case models are $\{\bar{\rho}, \bar{\theta}\} = \{0.789, 0.587\}$ and $\{\rho^p, \theta^p\} = \{0.803, 0.600\}$ respectively, implying $\bar{b}(\beta) = 1.94$ and $b^p(\beta) = 2.02$. The parametric worst-case model is thus nearly identical to the benchmark model. The relevant measure of risk in the economy is thus essentially identical under the two models, meaning that the equity premium is almost completely unaffected by parameter uncertainty of this sort. In contrast, under the unconstrained nonparametric worst case $b^w(\beta) = 4.40$, more than two times higher than $\bar{b}(\beta)$.

Thus, when we allow the agent to choose an unrestricted worst-case model, the outcome is very similar to what we obtained for the white-noise benchmark. The worst-case deviates from the benchmark at *very* low frequencies. This is true even though in this setting the estimation uncertainty (through $\bar{f}(\omega)$) varies across frequencies.

---

[28]See Collin-Dufresne, Johannes, and Lochstoer (2013) for a deep analysis of the case of parameter uncertainty with Epstein–Zin preferences. Andrei, Carlin, and Hasler (2013) study a model in which investors disagree about the persistence of trend growth rates.

8   Conclusion

This paper studies asset pricing when agents are unsure about the endowment process. The fundamental insight is that the long-run risk model, precisely because it is difficult to test for empirically and yet has important welfare implications, represents a natural model for investors who are unsure of the true data-generating process to use for pricing assets. More technically, for an agent with Epstein–Zin preferences who estimates consumption dynamics nonparametrically, the model that leads to the lowest lifetime utility for a given level of plausibility displays large amounts of long-run risk in consumption growth. In fact, when the agent's point estimate is that consumption growth is i.i.d., the worst-case model is literally the homoskedastic long-run risk model of Bansal and Yaron (2004). Furthermore, the nonparametric worst-case model can differ substantially from a parametric worst case that only features parameter uncertainty, instead of uncertainty about the actual model driving consumption growth.

We are able to obtain solutions in a setting that previously resisted both analytic and numerical analysis. The results show exactly what types of models agents fear when they contemplate unrestricted dynamics: they fear fluctuations at the very lowest frequencies. Not only do these fears raise risk premia on average, but they also induce countercyclical risk premia, raising the volatility of asset prices and helping to match the large movements in aggregate price/dividend ratios.

In a calibration of our model where the true process driving consumption growth is white noise, we generate a realistic equity premium, a volatile and realistically persistent price/dividend ratio, returns with similar predictability to the data at both short and long horizons, and estimates of the EIS from aggregate regressions of zero. None of these results require us to posit that there is long-run risk in the economy. They are all driven by the agent's worst-case model. And we show that the worst-case model is not at all implausible: it is rejected at the 5 percent level in less than 10 percent of simulated 100-year samples.

Economists have spent years arguing over what the consumption process is. We argue that a reasonable strategy, and one that is tractable to solve, for an investor facing that type of uncertainty, would be to make plans for a worst-case scenario. The message of this paper is that worst-case planning is able to explain a host of features of the data that were heretofore viewed as puzzling and difficult to explain in a setting that was even remotely rational.

References

Abel, Andrew B. 1999. "Risk Premia and Term Premia in General Equilibrium." *Journal of Monetary Economics*, 43: 3–33.

Abel, Andrew B. 2002. "An exploration of the effects of pessimism and doubt on asset returns." *Journal of Economic Dynamics and Control*, 26(7): 1075–1092.

Andrei, Daniel, Bruce Carlin, and Michael Hasler. 2013. "Model Disagreement, Volatility, and Trading Volume." Working paper.

Andrews, Donald W.K., and Werner Ploberger. 1996. "Testing for Serial Correlation Against an ARMA(1,1) Process." *Journal of the American Statistical Association*, 91(435): 1331–1342.

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

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

Barillas, Francisco, Lars P. Hansen, and Thomas J. Sargent. 2009. "Doubts or Variability?" *Journal of Economic Theory*, 144(6): 2388–2418.

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

Barsky, Robert B., and Bradford De Long. 1993. "Why Does the Stock Market Fluctuate?" *Quarterly Journal of Economics*, 108(2): 291–311.

Barsky, Robert B., F. Thomas Juster, Miles S. Kimball, and Matthew D. Shapiro. 1997. "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study." *The Quarterly Journal of Economics*, 112(2): 537–579.

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

Berk, Kenneth N. 1974. "Consistent Autoregressive Spectral Estimates." *The Annals of Statistics*, 2(3): 489–502.

Bidder, Rhys, and Matthew E. Smith. 2015. "Doubts and Variability: A Robust Perspective on Exotic Consumption Series." Working paper.

Brandt, Michael W, Qi Zeng, and Lu Zhang. 2004. "Equilibrium stock return dynamics under alternative rules of learning about hidden states." *Journal of Economic Dynamics and Control*, 28(10): 1925–1954.

Brennan, Michael J., and Yihong Xia. 2001. "Stock Price Volatility and the Equity Premium." *Journal of Monetary Economics*, 47: 249–283.

Brockwell, P.J., and R.A. Davis. 1988*a*. "Applications of Innovation Representations in Time Series Analysis." In *Probability and Statistics: Essays in Honor of Franklin A. Graybill.* , ed. J.N. Srivastava, 61–84. Elsevier.

Brockwell, P.J., and R.A. Davis. 1988*b*. "Simple consistent estimation of the coefficients of a linear filter." *Stochastic Processes and their Applications*, 28(1): 47–59.

Campbell, John Y., and N. Gregory Mankiw. 1989. "Consumption, Income and Interest Rates: Reinterpreting the Time Series Evidence." In *NBER Macroeconomics Annual.* , ed. Olivier Jean Blanchard and Stanley Fischer.

Campbell, John Y., and Tuomo Vuolteenaho. 2004. "Bad Beta, Good Beta." *American Economic Review*, 94(5): 1249–1275.

Cecchetti, Stephen G, Pok-Sang Lam, and Nelson C Mark. 2000. "Asset Pricing with Distorted Beliefs: Are Equity Returns Too Good to Be True?" *American Economic Review*, 787–805.

Chen, Long, Zhi Da, and Richard Priestley. 2012. "Dividend Smoothing and Predictability." *Management Science*, 58(10): 1834–1853.

Chen, Xiaohong. 2007. "Large sample sieve estimation of semi-nonparametric models." *Handbook of econometrics*, 6: 5549–5632.

Chow, Yun-Shyong, and Ulf Grenander. 1985. "A Sieve Method for the Spectral Density." *The Annals of Statistics*, 13(3): 998–1010.

Cochrane, John H. 2008. "The Dog That Did Not Bark: A Defense of Return Predictability." *Review of Financial Studies*, 21(4): 1533–1575.

Cogley, Timothy, and Thomas J. Sargent. 2008. "The market price of risk and the equity premium: A legacy of the Great Depression?" *Journal of Monetary Economics*, 55(3): 454–476.

Collin-Dufrese, Pierre, Michael Johannes, and Lars A. Lochstoer. 2013. "Parameter Learning in General Equilibrium: The Asset Pricing Implications." Working paper.

Collin-Dufresne, Pierre, Michael Johannes, and Lars Lochstoer. Forthcoming. "Parameter Learning in General Equilibrium: The Asset Pricing Implications." *American Economic Review*.

Dahlhaus, R. 1996. "On the Kullback-Leibler information divergence of locally stationary processes." *Stochastic Processes and their Applications*, 62(1): 139–168.

Dahlhaus, Rainer. 2000. "A Likelihood Approximation for Locally Stationary Processes." *The Annals of Statistics*, 28(6): 1762–1794.

Dew-Becker, Ian. 2015. "How risky is consumption in the long-run? Benchmark estimates from a novel unbiased and efficient estimator." Working paper.

Diaconis, Persi, and David Freedman. 1986. "On the consistency of Bayes estimates." *The Annals of Statistics*, 14(1): 1–26.

Doob, J.L. 1948. "Application of the Theory of Martingales." Vol. 13 of *Colloques Internationaux du Centre National de la Recherche Scientifique*, 23–27.

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

Drechsler, Itamar, and Amir Yaron. 2011. "What's Vol Got to Do with it?" *The Review of Financial Studies*, 24(1): 1–45.

Ellsberg, Daniel. 1961. "Risk, ambiguity, and the Savage axioms." *The Quarterly Journal of Economics*, 75(4): 643–669.

Epstein, Larry, and Stan Zin. 1991. "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Investigation." *Journal of Political Economy*, 99: 555–576.

Epstein, Larry G., and Martin Schneider. 2003. "Recursive Multiple Priors." *Journal of Economic Theory*, 113(1): 1–31.

Epstein, Larry G., and Martin Schneider. 2007. "Learning Under Ambiguity." *Review of Economic Studies*, 74(4): 1275–1303.

Evans, Martin. 1998. "Real Rates, Expected Inflation and Inflation Risk Premia." *Journal of Finance*, 53(1): 187–218.

Faust, Jon. 1999. "Conventional Confidence Intervals for Points on Spectrum Have Confidence Level Zero." *Econometrica*, 67(3): 629–637.

Fuster, Andreas, Benjamin Hebert, and David Laibson. 2011. "Natural Expectations, Macroeconomic Dynamics, and Asset Pricing." *NBER Macroeconomics Annual*, 26(1): 1–48.

Gilboa, Itzhak, and David Schmeidler. 1989. "Maxmin expected utility with non-unique prior." *Journal of Mathematical Economics*, 18(2): 141–153.

Gray, Robert M. 2000. "Toeplitz and Circulant Matrices: A Review." *Foundations and Trends in Communications and Information Theory*, 2(3): 155–239.

Greenwood, Robin, and Andrei Shleifer. 2014. "Expectations of Returns and Expected Returns." *Review of Financial Studies*, 27(3): 714–746.

Grenader, Ulf, and Gabor Szego. 1958. *Toeplitz Forms and their Applications.* University of California Press.

Hall, Robert E. 1988. "Intertemporal Substitution in Consumption." *Journal of Political Economy*, 96(2): 339–357.

Hansen, Lars P., and Thomas J. Sargent. 2001. "Robust Control and Model Uncertainty." *American Economic Review*, 91(2): 60–66.

Hansen, Lars P., and Thomas J. Sargent. 2005. "Robust estimation and control under commitment." *Journal of Economic Theory*, 124(2): 258–301.

Hansen, Lars P., and Thomas J. Sargent. 2007. *Robustness.* Princeton University Press.

Hansen, Lars P., and Thomas J. Sargent. 2010*a*. "Fragile Beliefs and the Price of Uncertainty." *Quantitative Economics*, 1(1): 129–162.

Hansen, Lars Peter, and Thomas J. Sargent. 2010*b*. "Fragile beliefs and the price of uncertainty." *Quantitative Economics*, 1(1): 129–162.

Hirshleifer, David, and Jianfeng Yu. 2012. "Asset Pricing in Production Economies with Extrapolative Expectations." Working Paper.

Ju, Nengjiu, and Jianjun Miao. 2012. "Ambiguity, Learning, and Asset Returns." *Econometrica*, 80(2): 559–591.

Kimball, Miles S., Claudia R. Sahm, and Matthew D. Shapiro. 2008. "Imputing risk tolerance from survey responses." *Journal of the American Statistical Association*, 103(483): 1028–1038.

Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji. 2005. "A Smooth Model of Decision Making under Ambiguity." *Econometrica*, 73(6): 1849–1892.

Knight, Frank H. 1921. *Risk, Uncertainty, and Profit.* Houghton Mifflin.

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

Leippold, Markus, Fabio Trojani, and Paolo Vanini. 2008. "Learning and Asset Prices Under Ambiguous Information." *Review of Financial Studies*, 21(6): 2656–2597.

LeRoy, Stephen F., and Richard D. Porter. 1981. "The Present-Value Relation: Tests Based on Implied Variance Bounds." *Econometrica*, 49(3): 555–574.

Liu, Jun, Jun Pan, and Tan Wang. 2004. "An Equilibrium Model of Rare-Event Premia and Its Implication for Option Smirks." *Review of Financial Studies*, 18(1): 131–164.

Ljung, G. M., and G. E. P. Box. 1978. "On a Measure of Lack of Fit in Time Series Models." *Biometrika*, 65(2): pp. 297–303.

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini. 2006. "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences." *Econometrica*, 74(6): 1447–1498.

Machina, Mark J., and Marciano Siniscalchi. 2014. "Ambiguity and Ambiguity Aversion." In *Handbook of the Economics of Risk and Uncertainty*. 729–807. Elsevier.

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

Marakani, Srikant. 2009. "Long run consumption risks: Are they there?" Working paper.

Marsh, Terry A., and Robert C. Merton. 1987. "Dividend Behavior for the Aggregate Stock Market." *The Journal of Business*, 60(1): 1–40.

Mehra, Rajnish, and Edward C. Prescott. 1985. "The Equity Premium: A puzzle." *Journal of Monetary Economics*, 15(2): 145–161.

Monti, Anna Clara. 1997. "Empirical Likelihood Confidence Regions in Time Series Models." *Biometrika*, 84(2): 395–405.

Müller, Ulrich K., and Mark W. Watson. 2013. "Measuring Uncertainty about Long-Run Predictions." Working paper.

Routledge, Bryan R., and Stanley E. Zin. 2009. "Model Uncertainty and Liquidity." *Review of Economic Dynamics*, 12(4): 543–566.

Sbuelz, Alessandro, and Fabio Trojani. 2008. "Asset prices with locally constrained-entropy recursive multiple-priors utility." *Journal of Economic Dynamics & Control*, 32: 3695–3717.

Shiller, Robert J. 1981. "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" *American Economic Review*, 71(3): 421–436.

Shimotsu, Katsumi, and Peter C. B. Phillips. 2005. "Exact local Whittle estimation of fractional integration." *Annals of Statistics*, 33(4): 1890–1933.

Sims, Christopher A. 1971. "Distributed lag estimation when the parameter space is explicitly infinite-dimensional." *The Annals of Mathematical Statistics*, 42(5): 1622–1636.

Sims, Christopher A. 1972. "The Role of Approximate Prior Restrictions in Distributed Lag Estimation." *Journal of the American Statistical Association*, 67(337): 169–175.

Skiadas, Costis. 2013. "Smooth Ambiguity Aversion Toward Small Risks and Continuous-Time Recursive Utility." *Journal of Political Economy*, 121(4): 775–792.

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

Uppal, Raman, and Tan Wang. 2003. "Model Misspecification and Underdiversification." *Journal of Finance*, 58(6): 2465–2486.

Veronesi, Pietro. 2000. "How Does Information Quality Affect Stock Returns?" *Journal of Finance*, 55(2): 807–837.

Weil, Philippe. 1989. "The Equity Premium Puzzle and the Risk-Free Rate Puzzle." *Journal of Monetary Economics*, 24(3): 401–421.

Whittle, Peter. 1953. "Estimation and Information in Stationary Time Series." *Arkiv for Matematik*, 2: 423–434.

Figure 1a. Weighting function $|Z|^2$

Cycle length (years)
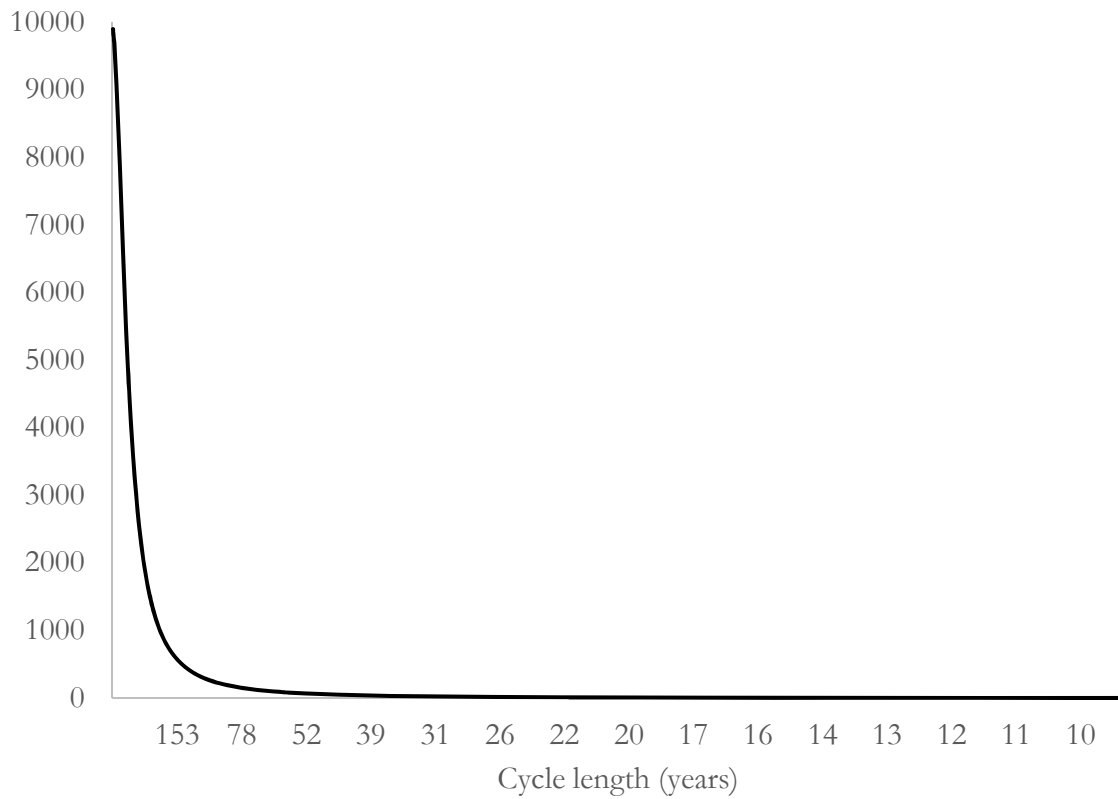
Figure 1b. Spectrum under benchmark and worst case for white-noise benchmark

Worst-case spectrum

Benchmark spectrum
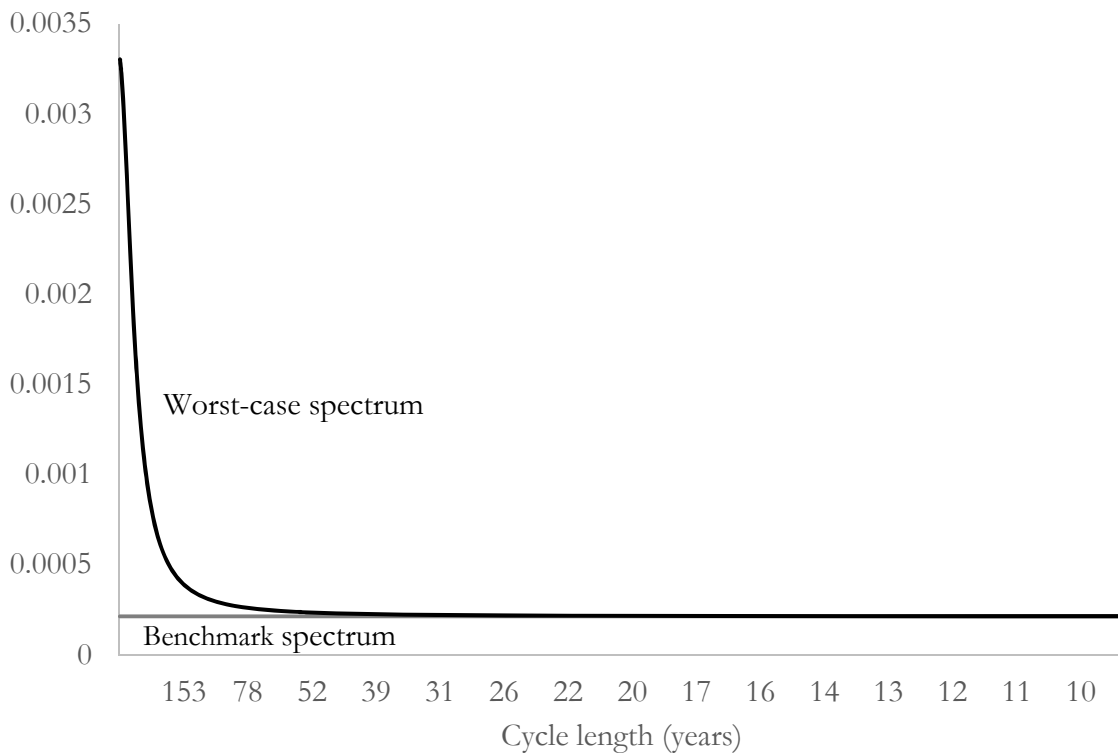
Cycle length (years)

Figure 2. Empirical and model-implied R2's from return forecasting regressions
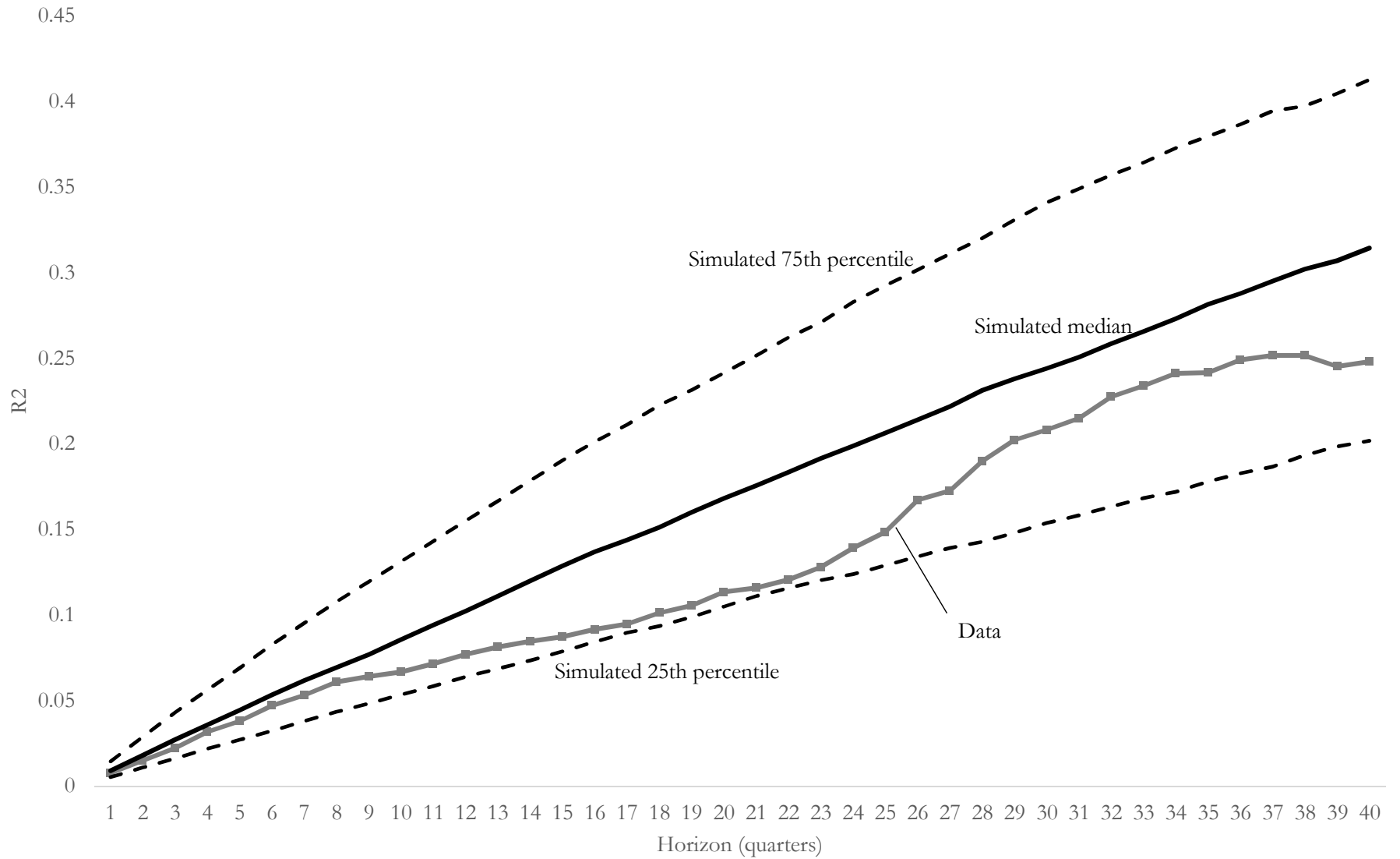
Notes: Black lines give results from simulated regressions on 90-year samples. The grey line plots R2s from regressions of aggregate equity returns on the price/dividend ratio between 1926 and 2015.

Figure 3. Benchmark and worst-case transfer function

Non-parametric worst case

Parametric worst case

Benchmark

Frequency (radians/year)
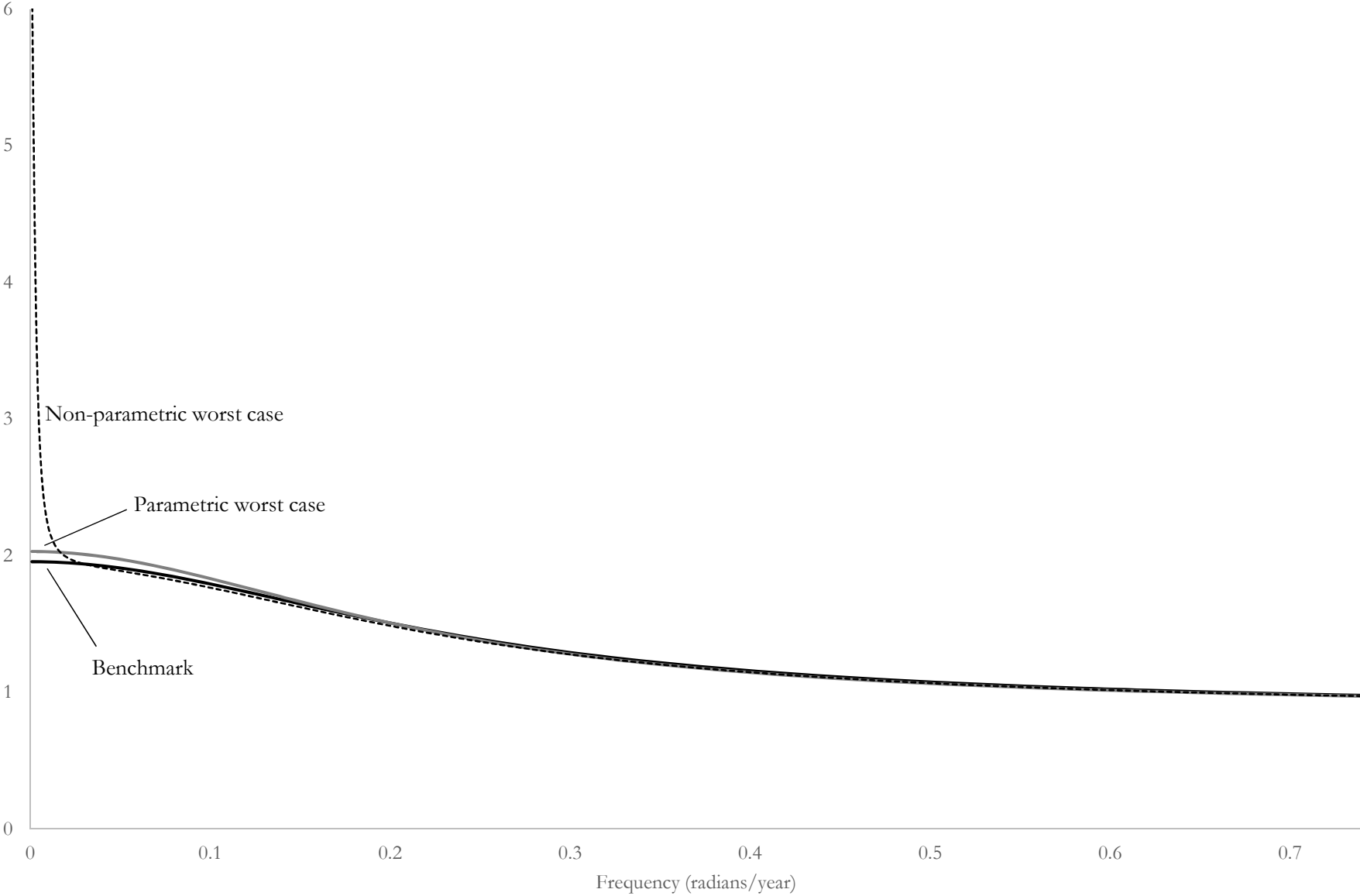
Notes: benchmark, parametric, and non-parametric transfer functions for the ARMA(1,1) specification estimated on post-war US consumption growth. The parametric worst case is the ARMA(1,1) that minimizes lifetime utility subject to the same KL divergence as is used for the nonparametric worst case.

Table 1: Asset pricing moments for the white-noise benchmark

Consumption dynamics		Benchmark	Worst-case	
$\overline{\sigma}$	Consumption conditional volatility	1.47%	1.47%	
$\overline{b}(\beta)$	Long-run volatility	1	2.449	
$\overline{\mu}$	Mean consumption growth	0.45%	0.37%	

Asset pricing moments (annualized)	Model	Standard EZ	Data
Standard deviation of pricing kernel	0.30	0.14	N/A
Mean excess equity return	6.34	1.95	6.33
Standard deviation of equity return	19.44	14.09	19.42
Mean risk-free rate	1.89	2.44	0.86
Standard deviation of risk-free rate	0.31	0	0.97
1-year autocorrelation of P/D	0.96	N/A	0.81
Standard deviation of P/D	0.19	0	0.29
Mean 10-year/1-quarter term spread	-13.3bp	0	N/A
Implied estimate of EIS	0	N/A	0.14

Notes: Moments from the model with a white-noise benchmark process for consumption growth. The "standard Epstein–Zin" results are for where the agent is sure of the consumption process. P/D is the price/dividend ratio for the levered consumption claim. The values in the data treat the aggregate equity market as analogous to the levered consumption claim. The EIS estimate is based on a regression of consumption growth on interest rates. In the second column interest rates are constant, so the regression is degenerate.

Table 2. Probability of rejecting the pricing model

Rejection probs. (5% critical value, H0=worst-case model)

	50-year sample	100-year sample
Ljung–Box	5.1%	5.1%
ARMA(1,1)	5.6%	6.0%

ARMA(1,1) rejection probabilities for alternative persistence in pricing model

Persistence	100-year sample	
0.9975	6.0%	(Our worst case)
0.995	8.7%	
0.99	13.5%	(Hansen and Sargent (2010))
0.98	28.3%	
0.94	82.7%	(Bansal and Yaron (2004))

Notes: Rejection probabilities are obtained by simulating the distributions of the test statistics in 50- and 100-year simulations of the cases where consumption growth is generated by the worst-case and white-noise models and asking how often the statistics in the latter simulation are outside the 95% range in the former simulation. In the bottom section, persistence is reduced but the price of risk in the pricing model is held constant.

Long-Run Risk is the Worst-Case Scenario
Appendix for online publication

Rhys Bidder and Ian Dew-Becker

May 4, 2016

Contents

A   Full derivation of proposition 1 and corollary 2

A.1   Deriving the Kullback–Leibler divergence

We model the agent as comparing models based on the expected value of a squared distance. In the case of a Gaussian model, the distance is exactly the expected likelihood ratio. When the time series are non-Gaussian, it becomes a quadratic distance that has been widely studied in the time series econometrics literature.

Models are indexed by the parameter set $\Theta \equiv \{b, \mu, \sigma^2\}$. The investor has a benchmark model for consumption growth dynamics, $\bar{\Theta}$. Denote the covariance matrix of consumption growth implied by a model Θ as Σ_Θ. The log likelihood for a sample of consumption growth under the model Θ is

$$-\frac{1}{2}\log|\Sigma_\Theta| - \frac{1}{2}(\Delta c_{1,...,T} - \mu)' \Sigma_\Theta^{-1} (\Delta c_{1,...,T} - \mu) \tag{A.1}$$

where $\Delta c_{1,...,T}$ denotes a column vector containing the sample of observed consumption growth between dates 1 and T. Now suppose consumption growth is generated by the model Θ. One may show that as $T \to \infty$, the expected log likelihood for the model $\bar{\Theta} = \{\bar{b}, \bar{\mu}, \bar{\sigma}^2\}$ is equal to

$$
\begin{aligned}
&\lim_{T\to\infty} T^{-1} E_\Theta \left[-\frac{1}{2}\log|\Sigma_{\bar{\Theta}}| - \frac{1}{2}(\Delta c_{1,...,T} - \bar{\mu})' \Sigma_{\bar{\Theta}}^{-1} (\Delta c_{1,...,T} - \bar{\mu}) \right] \\
&= -\frac{1}{2}\frac{1}{2\pi}\int_{-\pi}^{\pi} \log f_{\bar{\Theta}}(\omega)\, d\omega - \frac{1}{2}\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{f_\Theta(\omega)}{f_{\bar{\Theta}}(\omega)} - 1\, d\omega - \frac{1}{2}\frac{(\mu - \bar{\mu})^2}{f_{\bar{\Theta}}(0)}
\end{aligned}
\tag{A.2}
$$

where E_Θ denotes an expectation when the data is generated by the model Θ. (A.2) is simply the expected value of Whittle's (1953) expression for the log likelihood. Formally, the limit is an application of a well known result from Grenander and Szego (1958) that Toeplitz matrices converge asymptotically to circulant matrices. See Gray (2006) for a recent textbook review of such results. Examples of recent work using and extending the Whittle likelihood include Monti (1997), Dahlhaus (2000) and Shimotsu and Phillips (2005).

Now note that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{f_\Theta(\omega)}{f_{\bar{\Theta}}(\omega)}\, d\omega = \frac{1}{2\pi}\frac{\sigma^2}{\bar{\sigma}^2}\int_{-\pi}^{\pi} \frac{|B(\omega)|^2}{|\bar{B}(\omega)|^2}\, d\omega \tag{A.3}$$

Also, as long as the roots of B and $\bar{B}$ are inside the unit circle, we have $\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{B(\omega)}{\bar{B}(\omega)}\, d\omega = 1$.[1] We can

---

[1]To confirm this, write $b(L)$ as $b(L) = \prod_j (1 - a_j L)$ for $|a_j| < 1$. Using the same form for $\bar{b}$, note that each of the factors of $1/\bar{b}(L)$ has a convergent Taylor series, $\frac{1}{1-\bar{a}_j L} = \sum_{k=0}^{\infty} \bar{a}_j^k L^k$. Then the ratio $B(\omega)/\bar{B}(\omega)$ may be written as

$$B(\omega)/\bar{B}(\omega) = \prod_j \left(1 - a_j e^{i\omega}\right) \left(\sum_{k=0}^{\infty} \bar{a}_j^k e^{i\omega k}\right) \tag{A.4}$$

This function only has Fourier coefficients on the positive side of the origin, and the coefficient on the constant is $a_j^0 = 1$. That is, all the terms multiplying $e^{i\omega k}$ for $k > 0$ integrate to zero, so $\frac{1}{2\pi}\int_{-\pi}^{\pi} B(\omega)/\bar{B}(\omega)\, d\omega = 1$.

therefore write

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{|B(\omega)|^2}{|\bar{B}(\omega)|^2}d\omega = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{B(\omega)}{\bar{B}(\omega)}\right|^2 - \frac{B(\omega)}{\bar{B}(\omega)} - \frac{B(\omega)^*}{\bar{B}(\omega)^*} + 2 d\omega \tag{A.5}$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{B(\omega)-\bar{B}(\omega)}{\bar{B}(\omega)}\right|^2 d\omega + 1 \tag{A.6}$$

Which implies that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\sigma^2}{\bar{\sigma}^2}\frac{|B(\omega)|^2}{|\bar{B}(\omega)|^2} - 1 d\omega = \frac{1}{2\pi}\frac{\sigma^2}{\bar{\sigma}^2}\int_{-\pi}^{\pi}\left|\frac{B(\omega)-\bar{B}(\omega)}{\bar{B}(\omega)}\right|^2 d\omega + \frac{\sigma^2}{\bar{\sigma}^2} - 1 \tag{A.7}$$

$$= \frac{1}{2\pi}\frac{\sigma^2}{\bar{\sigma}^2}\int_{-\pi}^{\pi}\left|\frac{B(\omega)-\bar{B}(\omega)}{\bar{B}(\omega)}\right|^2 d\omega + \frac{\sigma^2 - \bar{\sigma}^2}{\bar{\sigma}^2} \tag{A.8}$$

Note also that Kolmogorov's formula implies that $\frac{1}{2\pi}\int_{-\pi}^{\pi}\log f_\Theta(\omega)\,d\omega = \log\sigma^2$.

The investor measures the distance between the benchmark model $\bar{\Theta}$ and an alternative Θ as the difference in the asymptotic expected log likelihoods of the two models when the data is generated by Θ, which is the KL divergence,

$$\lim_{T\to\infty}T^{-1}E_\Theta\left[LL(T,\Theta) - LL(T,\bar{\Theta})\right] = \frac{1}{2}\frac{1}{2\pi}\frac{\sigma^2}{\bar{\sigma}^2}\int_{-\pi}^{\pi}\frac{|B(\omega)-\bar{B}(\omega)|^2}{|\bar{B}(\omega)|^2}d\omega$$

$$-\frac{1}{2}\left(\log\left(\frac{\sigma^2}{\bar{\sigma}^2}\right) - \frac{\sigma^2 - \bar{\sigma}^2}{\bar{\sigma}^2}\right) + \frac{1}{2}\frac{(\mu - \hat{\mu})^2}{f_{\bar{\Theta}}(0)} \tag{A.9}$$

A.2 Minimization

The investor's optimization problem to find the worst-case model is

$$\min_{b,\mu,\sigma^2}\frac{\beta}{1-\beta}\frac{1-\alpha}{2}b(\beta)^2\sigma^2 + \frac{\beta}{1-\beta}\mu + \frac{\lambda}{2}\left[\int\frac{f(\omega)}{\bar{f}(\omega)} - \log\frac{f(\omega)}{\bar{f}(\omega)}d\omega + \frac{(\mu-\bar{\mu})^2}{\bar{f}(0)}\right] \tag{A.10}$$

where the integral sign without limits denotes $\frac{1}{2\pi}\int_{-\pi}^{\pi}$.

The spectral density $f(\omega)$ can be expressed as

$$f(\omega) = \exp\left(2\sum_{j=0}^{\infty}c_j\cos(\omega j)\right) \tag{A.11}$$

for a set of real coefficients c_j (Priestley (1981)). The coefficients c_j are simply the Fourier coefficients of the log of the spectrum; we only include coefficients for non-negative j because the spectrum is a real and

even function. Furthermore, setting $\sigma = \exp(c_0)$, we have

$$\sigma B(\omega) = \exp\left(\sum_{j=0}^{\infty} c_j e^{i\omega j}\right) \tag{A.12}$$

$$b_m = \int e^{-i\omega m} \exp\left(\sum_{j=1}^{\infty} c_j e^{i\omega j}\right) d\omega \tag{A.13}$$

where the b_j are the coefficients from the Wold representation for the spectrum $|B(\omega)|^2$ (Priestley (1981)). Since $\sigma = \exp(c_0)$, $b_0 = 1$. Furthermore, $b_j = 0$ for all $j < 0$. (A.12) is known as the canonical factorization of the spectrum. We solve the optimization problem by directly choosing the c_j. Since the Fourier transform is one-to-one, choosing the c_j is equivalent to optimizing over the spectrum directly. Since $B(\omega)$ is obtained from the Wold representation, it is guaranteed to be causal, invertible, and minimum-phase. Last, the innovation variance associated with the spectrum $f(\omega)$ is $\sigma^2 = \exp(2c_0)$.

We first calculate derivatives involved in the optimization

$$\frac{d}{dc_j}[\sigma b(\beta)] = \frac{d}{dc_j}\sum_{m=0}^{\infty} \beta^m \int \exp\left(\sum_{j=0}^{\infty} c_j e^{i\omega j}\right) e^{-i\omega m} d\omega \tag{A.14}$$

$$= \sum_{m=0}^{\infty} \beta^m \int \frac{d}{dc_j} \exp\left(\sum_{j=0}^{\infty} c_j e^{i\omega j}\right) e^{-i\omega m} d\omega \tag{A.15}$$

$$= \sum_{m=0}^{\infty} \beta^m \int \exp\left(\sum_{j=0}^{\infty} c_j e^{i\omega j}\right) e^{-i\omega(m-j)} d\omega \tag{A.16}$$

$$= \sigma \sum_{m=0}^{\infty} \beta^m b_{m-j} \tag{A.17}$$

$$= \sigma \sum_{m=0}^{\infty} \beta^{m+j} b_m = \sigma b(\beta)\beta^j \tag{A.18}$$

where the derivative can be passed inside the integral because $B(\omega)$ is continuous and differentiable with respect to the c_j and the last line follows from the fact that $b_j = 0$ for $j < 0$.

Next, the derivative of the ratio of the spectra is

$$\frac{d}{dc_j}\int \frac{f(\omega)}{\bar{f}(\omega)} d\omega = \frac{d}{dc_j}\int \frac{\exp\left(2\sum_{j=0}^{\infty} c_j \cos(\omega j)\right)}{\bar{f}(\omega)} d\omega \tag{A.19}$$

$$= 2\int \frac{f(\omega)}{\bar{f}(\omega)} \cos(\omega j) d\omega \tag{A.20}$$

And last,

$$\frac{d}{dc_0}\int \log \frac{f(\omega)}{\bar{f}(\omega)} d\omega = 2 \tag{A.21}$$

4

So we have

$$\min_{b,\mu,\sigma^2} \frac{\beta}{1-\beta}\frac{1-\alpha}{2}\sigma^2 b\left(\beta\right)^2 + \frac{\beta}{1-\beta}\mu + \frac{\lambda}{2}\left[\int \frac{f\left(\omega\right)}{\bar{f}\left(\omega\right)} - \log\frac{f\left(\omega\right)}{\bar{f}\left(\omega\right)}d\omega + \frac{\left(\mu-\bar{\mu}\right)^2}{\bar{f}\left(0\right)}\right] \tag{A.22}$$

The first-order condition for each $j > 0$ is

$$0 = 2\frac{\beta}{1-\beta}\frac{1-\alpha}{2}\sigma_w^2 b^w\left(\beta\right)^2\beta^j + \frac{\lambda}{2}\int \frac{f^w\left(\omega\right)}{\bar{f}\left(\omega\right)}2\cos\left(\omega j\right)d\omega \tag{A.23}$$

For $j = 0$,

$$0 = 2\frac{\beta}{1-\beta}\frac{1-\alpha}{2}\sigma_w^2 b^w\left(\beta\right)^2 + \frac{\lambda}{2}\int \frac{f^w\left(\omega\right)}{\bar{f}\left(\omega\right)}2d\omega - \lambda \tag{A.24}$$

Now multiply each of the first-order conditions by $\cos\left(j\kappa\right)$ for some κ.

$$\begin{aligned}
0 &= 2\frac{\beta}{1-\beta}\frac{1-\alpha}{2}\sigma_w^2 b^w\left(\beta\right)^2\cos\left(j\kappa\right)\beta^j + \frac{\lambda}{2}\int \frac{f^w\left(\omega\right)}{\bar{f}\left(\omega\right)}2\cos\left(j\kappa\right)\cos\left(\omega j\right)d\omega \tag{A.25}\\
&= 2\frac{\beta}{1-\beta}\frac{1-\alpha}{2}\sigma_w^2 b^w\left(\beta\right)^2\cos\left(j\kappa\right)\beta^j + \frac{\lambda}{2}\int \frac{f^w\left(\omega\right)}{\bar{f}\left(\omega\right)}\left(\cos\left(j\left(\kappa+\omega\right)\right) + \cos\left(j\left(\kappa-\omega\right)\right)\right)d\omega \tag{A.26}\\
&= 2\frac{\beta}{1-\beta}\frac{1-\alpha}{2}\sigma_w^2 b^w\left(\beta\right)^2\cos\left(j\kappa\right)\beta^j + \frac{\lambda}{2}\int \frac{f^w\left(\omega\right)}{\bar{f}\left(\omega\right)}2\cos\left(j\left(\kappa+\omega\right)\right)d\omega \tag{A.27}
\end{aligned}$$

where the third line follows by

$$\begin{aligned}
&= 2\frac{\beta}{1-\beta}\frac{1-\alpha}{2}\sigma_w^2 b^w\left(\beta\right)^2\cos\left(j\kappa\right)\beta^j + \frac{\lambda}{2}\left[\int \frac{f^w\left(\omega\right)}{\bar{f}\left(\omega\right)}\cos\left(j\left(\kappa+\omega\right)\right)d\omega + \int \frac{f^w\left(\omega\right)}{\bar{f}\left(\omega\right)}\cos\left(j\left(\kappa-\omega\right)\right)d\omega\right]\\
&= 2\frac{\beta}{1-\beta}\frac{1-\alpha}{2}\sigma_w^2 b^w\left(\beta\right)^2\cos\left(j\kappa\right)\beta^j + \frac{\lambda}{2}\left[\int \frac{f^w\left(\omega\right)}{\bar{f}\left(\omega\right)}\cos\left(j\left(\kappa+\omega\right)\right)d\omega + \int \frac{f^w\left(-\omega\right)}{\bar{f}\left(-\omega\right)}\cos\left(j\left(\kappa+\omega\right)\right)d\omega\right]\\
&= 2\frac{\beta}{1-\beta}\frac{1-\alpha}{2}\sigma_w^2 b^w\left(\beta\right)^2\cos\left(j\kappa\right)\beta^j + \frac{\lambda}{2}\left[\int \frac{f^w\left(\omega\right)}{\bar{f}\left(\omega\right)}\cos\left(j\left(\kappa+\omega\right)\right)d\omega + \int \frac{f^w\left(\omega\right)}{\bar{f}\left(\omega\right)}\cos\left(j\left(\kappa+\omega\right)\right)d\omega\right]
\end{aligned}$$

That is, since $\frac{f^w\left(\omega\right)}{\bar{f}\left(\omega\right)}$ is even, we can always reverse the sign of ω in the integration.

Now take the first-order condition (FOC) for $j = 0$, multiply it by $\frac{1}{2}$, and add to the sum of the FOCs for $j > 0$ multiplied by $\cos\left(\kappa j\right)$,

$$\begin{aligned}
0 &= \frac{\beta}{1-\beta}\frac{1-\alpha}{2}\sigma_w^2 b^w\left(\beta\right)^2 + \frac{\lambda}{2}\int \frac{f^w\left(\omega\right)}{\bar{f}\left(\omega\right)}d\omega - \frac{\lambda}{2} \tag{A.28}\\
&\quad + \frac{\beta}{1-\beta}\frac{1-\alpha}{2}\sigma_w^2 b^w\left(\beta\right)^2\sum_{j=1}^{\infty}2\cos\left(j\kappa\right)\beta^j + \frac{\lambda}{2}\int \frac{f^w\left(\omega\right)}{\bar{f}\left(\omega\right)}\sum_{j=1}^{\infty}2\cos\left(j\left(\kappa+\omega\right)\right)d\omega \tag{A.29}\\
&= \frac{\beta}{1-\beta}\frac{1-\alpha}{2}\sigma_w^2 b^w\left(\beta\right)^2\left(1+\sum_{j=1}^{\infty}2\cos\left(j\kappa\right)\beta^j\right) + \frac{\lambda}{2}\int \frac{f^w\left(\omega\right)}{\bar{f}\left(\omega\right)}\left(1+\sum_{j=1}^{\infty}2\cos\left(j\left(\kappa+\omega\right)\right)\right)d\omega - \frac{\lambda}{2}
\end{aligned}$$

5

We have

$$1 + \sum_{j=1}^{\infty} 2 \cos \left(j \left(\kappa + \omega \right) \right) = \delta \left(\kappa + \omega \right) \tag{A.30}$$

where $\delta \left(\cdot \right)$ is the Dirac delta function. Furthermore, note that $Z \left(\omega \right)$ is the transfer function of an AR(1) model with autocorrelation of β. It then follows that

$$\left(1 - \beta^2 \right) \left| Z \left(\kappa \right) \right|^2 = 1 + \sum_{j=1}^{\infty} 2 \cos \left(j \kappa \right) \beta^j \tag{A.31}$$

The FOC then becomes

$$0 = \frac{\beta}{1-\beta} \frac{1-\alpha}{2} \sigma_w^2 b^w \left(\beta \right)^2 \left(1 - \beta^2 \right) \left| Z \left(\kappa \right) \right|^2 + \frac{\lambda}{2} \left(\frac{f^w \left(\omega \right)}{\bar{f} \left(\omega \right)} - 1 \right) \tag{A.32}$$

$$\frac{f^w \left(\omega \right)}{\bar{f} \left(\omega \right)} - 1 = \lambda^{-1} \frac{\beta}{1-\beta} \left(\alpha - 1 \right) \sigma_w^2 b^w \left(\beta \right)^2 \left(1 - \beta^2 \right) \left| Z \left(\kappa \right) \right|^2 \tag{A.33}$$

$$f^w \left(\omega \right) = \bar{f} \left(\kappa \right) + \lambda^{-1} \beta \left(1 + \beta \right) \left(\alpha - 1 \right) \bar{f} \left(\kappa \right) \sigma_w^2 b^w \left(\beta \right)^2 \left| Z \left(\kappa \right) \right|^2 \tag{A.34}$$

This is the main result in the text.

A.3 The white-noise benchmark

In the white noise case, $\bar{f} \left(\kappa \right) = \bar{\sigma}^2$. The mean immediately follows,

$$\mu_w = \bar{\mu} - \lambda^{-1} \frac{\beta}{1-\beta} \bar{\sigma}^2 \tag{A.35}$$

For the dynamics,

$$f^w \left(\omega \right) = \bar{\sigma}^2 + \lambda^{-1} \beta \left(1 + \beta \right) \left(\alpha - 1 \right) \bar{\sigma}^2 \sigma_w^2 b^w \left(\beta \right)^2 \left| Z \left(\omega \right) \right|^2 \tag{A.36}$$

Denote the autocovariances under the worst-case model as γ_j^w. Then

$$\gamma_0^w + 2 \sum_{j=1}^{\infty} \gamma_j^w \cos \left(\omega j \right) = \bar{\sigma}^2 \left(1 + \varphi \left(1 + 2 \sum_{j=1}^{\infty} \cos \left(\omega j \right) \beta^j \right) \right) \tag{A.37}$$

$$\text{where } \varphi \equiv \lambda^{-1} \frac{\beta}{1-\beta} \left(\alpha - 1 \right) \sigma_w^2 b^w \left(\beta \right)^2 \tag{A.38}$$

Matching coefficients on each side yields

$$\gamma_0^w = \bar{\sigma}^2 \left(1 + \varphi \right) \tag{A.39}$$

$$\gamma_j^w = \bar{\sigma}^2 \varphi \beta^j \text{ for } |j| > 0 \tag{A.40}$$

These may be recognized as the autocovariances of an ARMA(1,1) process. Specifically, set

$$\Delta c_t = x_t + v_t \tag{A.41}$$

$$x_t = \beta x_{t-1} + \mu_t \tag{A.42}$$

$$\sigma_v^2 = \bar{\sigma}^2 \tag{A.43}$$

$$\sigma_\mu^2 = \bar{\sigma}^2 \varphi \left(1 - \beta^2\right) \tag{A.44}$$

Then one may confirm that Δc_t has autocovariances γ_j^w.

To find the equivalent univariate ARMA(1,1) representation, note that

$$\Delta c_t - \beta \Delta c_{t-1} = x_t - \beta x_{t-1} + v_t - \beta v_{t-1} \tag{A.45}$$

$$= \mu_t + v_t - \beta v_{t-1} \tag{A.46}$$

The second line is an MA(1), with

$$m_t \equiv \mu_t + v_t - \beta v_{t-1} \tag{A.47}$$

$$var\left(m_t\right) = \sigma_\mu^2 + \left(1 + \beta^2\right)\sigma_v^2 = \left(1 + \theta^2\right)\sigma_w^2 \tag{A.48}$$

$$cov\left(m_t, m_{t-1}\right) = -\beta\sigma_v^2 = -\theta\sigma_w^2 \tag{A.49}$$

We then find θ and σ_w^2 by solving that pair of equations. We have

$$\theta = \frac{\left(\frac{\sigma_\mu^2}{\sigma_v^2} + \left(1 + \beta^2\right)\right)\beta^{-1} - \sqrt{\left(\frac{\sigma_\mu^2}{\sigma_v^2} + \left(1 + \beta^2\right)\right)^2 \beta^{-2} - 4}}{2} \tag{A.50}$$

which immediately yields σ_w^2. Now θ depends on σ_μ^2, which depends on φ. But φ itself depends on $b\left(\beta\right)$. We therefore solve for θ and σ_ε^2 iteratively. Specifically, begin by guessing that $\varphi = \lambda^{-1}\frac{\beta}{1-\beta}\left(\alpha - 1\right)\bar{\sigma}^2$. We then calculate θ and σ_w^2 for that guess, and update φ, with $\varphi = \lambda^{-1}\frac{\beta}{1-\beta}\left(\alpha - 1\right)\sigma_w^2 \left(\frac{1-\theta\beta}{1-\beta^2}\right)^2$ and iterate to convergence.

B   Testing the worst-case model

This section provides details and further results for the small-sample tests of the worst-case model.

B.1   Test statistics

We examine three tests: the ARMA(1,1) likelihood-ratio test suggested by Andrews and Ploberger (AP; 1996), the Ljung–Box (LB; 1978) test, and a test based on the Newey–West (1987) estimator of the long-run variance.

For the AP and LB tests, as discussed in the text, we assume that the agent takes an observed

consumption history and creates a series of residuals,

$$\varepsilon_t^{\Theta_w} \equiv \left(\Delta c_t - \mu_w - a^w(L)(\Delta c_{t-1} - \mu_w)\right)\sigma_w^{-1} \tag{B.1}$$

Under the null hypothesis that the worst-case model is true, $\varepsilon_t^{\Theta_w}$ is white noise. To see the dynamics of $\varepsilon_t^{\Theta_w}$ under the benchmark model, note that we can write $\varepsilon_t^{\Theta_w}$ as

$$\varepsilon_t^{\Theta_w} = \frac{1 - \beta L}{1 - \theta L}\sigma_w^{-1}(\Delta c_t - \mu_w) \tag{B.2}$$

(where θ is defined for the worst-case model above). Under the benchmark, $\Delta c_t \sim N\left(\bar{\mu}, \bar{\sigma}^2\right)$, so we can write

$$\varepsilon_t^{\Theta_w} = \frac{1 - \beta L}{1 - \theta L}\frac{\bar{\sigma}}{\sigma_w}\bar{\varepsilon}_t + \frac{1 - \beta}{1 - \theta}\sigma_w^{-1}(\bar{\mu} - \mu_w) \tag{B.3}$$

where $\bar{\varepsilon}_t \sim N(0,1)$.

When we simulate the distribution of the AP and LB test statistics conditional on the benchmark model being true, we construct them on simulated samples of $\varepsilon_t^{\Theta_w}$ using (B.3).

As discussed in the text, for the AP and LB tests, we first calculate critical values under the benchmark model. That is, we simulate samples of the time series $\bar{\varepsilon}_t \sim N(0,1)$ and then construct the AP and LB test statistics for each sample. The critical values are the 95th percentiles of those simulated distributions.

The AP statistic is constructed exactly as in Andrews and Ploberger (1996). Specifically, for a sample $\varepsilon_t, t \in \{1, 2, ..., T\}$, define

$$\tilde{\omega}^2 = T^{-1}\sum_{t=1}^{T}\varepsilon_t^2 \tag{B.4}$$

$\tilde{\omega}^2$ is the log likelihood (ignoring constants) under the null hypothesis that $\varepsilon_t \sim N(0,1)$

Second, define

$$\varepsilon_t^* \equiv \varepsilon_t - T^{-1}\sum_{t=1}^{T}\varepsilon_t \tag{B.5}$$

$$\hat{\omega}^2(\theta) \equiv T^{-1}\sum_{t=1}^{T}(\varepsilon_t^*)^2 - \left[T^{-1}\frac{\left(\sum_{t=2}^{T}\varepsilon_t^*\sum_{i=0}^{t-2}\theta^i\varepsilon_{t-i-1}^*\right)^2}{\sum_{t=2}^{T}\left(\sum_{i=0}^{t-2}\theta^i\varepsilon_{t-i-1}^*\right)}\right] \tag{B.6}$$

$\hat{\omega}^2(\theta)$ is the log likelihood when the mean of ε_t is estimated freely and we also allow estimation of the parameter θ.

The likelihood ratio statistic is then

$$LR \equiv \sup_{\theta} T \log \frac{\tilde{\omega}^2}{\hat{\omega}^2(\theta)} \tag{B.7}$$

For each simulated sample, we optimize over θ numerically (first searching over a grid, then using the simplex algorithm from the best grid point).

Note that the LR statistic here compares the likelihood of the data under assumptions both that ε_t is serially uncorrelated and also that its mean is zero. $\hat{\omega}^2(\theta)$ is the maximized likelihood under an alternative model that allows both for serial correlation (of an ARMA(1,1) form) and also a non-zero mean. We also consider a version of the AP test that ignores the deviation in the mean under the null. This constraint may potentially improve the power of the test, because it means that we are only testing the dynamics of consumption growth, not the level. Specifically, the AP statistic with a fixed mean is

$$LR^* \equiv \sup_\theta T \log \frac{(\tilde{\omega}^*)^2}{\hat{\omega}^2(\theta)} \tag{B.8}$$

$$(\tilde{\omega}^*)^2 \equiv T^{-1} \sum_{t=1}^{T} (\varepsilon_t^*)^2 \tag{B.9}$$

LR^* differs from LR only in that the numerator of the likelihood ratio now uses demeaned data. In other words, the null allows for an estimated mean.

The LB statistic is calculated using the autocorrelations of the sample of ε_t, which we denote $\hat{\gamma}_j$. The statistic, for a maximum lag of j, is

$$LB_j \equiv T(T+2) \sum_{k=1}^{j} \frac{\hat{\gamma}_k^2}{T-k} \tag{B.10}$$

$$\hat{\gamma}_k \equiv \frac{\sum_{t=k+1}^{T} \varepsilon_t \varepsilon_{t-k}}{\sum_{t=k+1}^{T} \varepsilon_t^2} \tag{B.11}$$

Finally, we also examine here a test based on the Newey–West (1987) estimator for the long-run variance of a time series. We ask whether, observing a sample of data generated by the benchmark model, a person would reject the hypothesis that the long-run variance is as large as implied by the worst-case model.

Specifically, we calculate the Newey–West estimate of the long-run variance

$$LRV_j = \hat{\kappa}_0 + 2 \sum_{k=1}^{j} \left(1 - \frac{k}{j}\right) \hat{\kappa}_j \tag{B.12}$$

$$\hat{\kappa}_j \equiv T^{-1} \sum_{k=1}^{T-j} \left(\Delta c_k - T^{-1} \sum_{t=1}^{T} \Delta c_t\right) \left(\Delta c_{k+j} - T^{-1} \sum_{t=1}^{T} \Delta c_t\right) \tag{B.13}$$

We simulate the distribution of LRV_j given data generated by the worst-case model and define LRV_j^* to be the 5th percentile of that distribution. The agent then rejects the hypothesis that the data was driven by the worst-case model after observing a sample drawn from the benchmark model if LRV_j in that particular sample is less than LRV_j^*. That is, we ask how often the estimated long-run variance estimated under the benchmark model is smaller than the 5th percentile of the long-run variance estimated under the worst-case model.

B.2 Extended results

The main text discusses results for the LB and AP tests on samples of 50 and 100 years. Table A2 reports results using the Newey–West based test and using longer samples up to 1000 years.

As one would expect, as the samples grow, the rejection rates across all four tests rise. For 1000-year samples, all but the Ljung–Box test reject with probabilities greater than 85 percent, confirming that they eventually converge to the correct result asymptotically. However, one can see looking across the table that all the tests converge rather slowly. With 250 years of data, the AP tests reject the worst case still less than 10 percent of the time, while the NW test rejects approximately 25 percent of the time.

A natural question is why the rejections probabilities are so low, even for the Newey–West based test. A simple way to see the intuition is to consider the periodogram. In a finite sample, the lowest frequency at which the periodogram is observed is $2\pi/T$ radians, which corresponds to a cycle with wavelength equal to the sample. Asymptotically, the periodogram is distributed exponentially with mean equal to the spectral density. What distinguishes the worst-case model from the benchmark is that its spectrum is much larger at low frequencies.

Specifically, the spectrum under the worst case has a value at frequency zero of $f^w(0) = b^w(1)^2 \sigma_w^2 = 0.00491,$, whereas under the true model, $\bar{f}(0) = 0.000215$. So $f^w(0)$ is 23 times larger than $\bar{f}(0)$. Given that the standard deviation of the periodogram is equal to the level of the spectrum itself, $f^w(0)$ is 22 standard deviations higher than $\bar{f}(0)$ and should be easily distinguishable.

However, since we do not observe the periodogram at frequency zero, what really matters is the value of the spectrum at $\omega = 2\pi/T$. For $T = 200$, $f^w(2\pi/200) = 0.000244$, which is only higher than $\bar{f}(2\pi/200) = \bar{\sigma}^2$ by a factor of 1.13. So in a sample with 200 observations, there simply is little information in the sample that reveals the deviations between f^w and $\bar{f}$.

In a 100-year sample, rejection is obviously easier. The first periodogram ordinate has mean $f^w(2\pi/400) = 0.00312$, which is now substantially larger than $\bar{f}$. On the other hand, this is still only a single data point for the estimators to use.

C Interpretation of the distance measure as a Wald test

This section provides an alternative of the distance measure used in the main text as a Wald test on estimated MA coefficients. Specifically, the part of the distance measure $\int \frac{|B(\omega) - \bar{B}(\omega)|^2}{\bar{f}(\omega)} d\omega$ represents the asymptotic expected value of a Wald statistic for a joint test of all the MA coefficients in the lag polynomial $b(L)$.

Brockwell and Davis (1988b) show that for an MA model of order m, the coefficients are asymptotically

normal with a covariance matrix denoted Σ_m. As $m \to \infty$, Σ_m converges to a product,[2]

$$\Sigma_m \quad \to \quad J_m^{True} J_m^{True\prime} \tag{C.1}$$

$$\text{where } J_m^{True} \quad \equiv \quad \begin{bmatrix} b_0^{True} & b_1^{True} & \cdots & b_m^{True} \\ 0 & b_0^{True} & \cdots & b_{m-1}^{True} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_0^{True} \end{bmatrix} \tag{C.2}$$

A natural empirical counterpart to that variance is to replace J^{True} with $\bar{J}$, defined analogously using the point estimate $\bar{b}$. The Wald statistic for the MA coefficients (ignoring scale factors) is then

$$m^{-1} \left(\mathbf{b}_{1:m} - \bar{\mathbf{b}}_{1:m} \right) \left(\bar{J}_m \bar{J}_m' \right)^{-1} \left(\mathbf{b}_{1:m} - \bar{\mathbf{b}}_{1:m} \right)' \tag{C.3}$$

where $\mathbf{b}_{1:m}$ is the row vector of the first m elements of the vector of coefficients in the model b.

J_m is a Toeplitz matrix, and it is well known that Toeplitz matrices, their products, and their inverses, asymptotically converge to circulant matrices (Grenander and Szegő (1958) and Gray (2006)). So $\bar{\Sigma}_m^{-1}$ has an approximate orthogonal decomposition, converging as $m \to \infty$, such that[3]

$$\bar{\Sigma}_m^{-1} \approx \Lambda_m \bar{F}_m^{-1} \Lambda_m^* \tag{C.4}$$

where * here represents transposition and complex conjugation, Λ_m is the discrete Fourier transform matrix with element j, k equal to $\exp\left(-2\pi i \left(j-1\right)\left(k-1\right)/m\right)$, $\bar{F}_m$ is diagonal with elements equal to the discrete Fourier transform of the autocovariances. Now if we define the vector $\mathbf{B}$ to be the Fourier transform of $\mathbf{b}$, $\mathbf{B}_{1:m} \equiv \mathbf{b}_{1:m} \Lambda_m$, we have

$$m^{-1} \left(\mathbf{b}_{1:m} - \bar{\mathbf{b}}_{1:m} \right) \bar{\Sigma}_m^{-1} \left(\mathbf{b}_{1:m} - \bar{\mathbf{b}}_{1:m} \right)' \quad \approx \quad m^{-1} \left(\mathbf{B}_m \Lambda_m^* - \bar{\mathbf{B}}_m \Lambda_m^* \right) \Lambda_m \bar{F}_m^{-1} \Lambda_m^* \left(\mathbf{B}_m^{*\prime} \Lambda_m' - \bar{\mathbf{B}}_m^{*\prime} \Lambda_m' \right)' \tag{C.5}$$

$$= \quad m^{-1} \left(\mathbf{B}_m - \bar{\mathbf{B}}_m \right) \bar{F}_m^{-1} \left(\mathbf{B}_m - \bar{\mathbf{B}}_m \right)^* \tag{C.6}$$

which itself, by Szegő's theorem, converges as $m \to \infty$ to an integral,

$$m^{-1} \left(\mathbf{B}_m - \bar{\mathbf{B}}_m \right) \bar{F}_m^{-1} \left(\mathbf{B}_m - \bar{\mathbf{B}}_m \right)^* \to \int \frac{\left| B\left(\omega\right) - \bar{B}\left(\omega\right) \right|^2}{\bar{f}\left(\omega\right)} d\omega \tag{C.7}$$

So the integral $\int \frac{\left| B(\omega) - \bar{B}(\omega) \right|^2}{\bar{f}(\omega)} d\omega$ may be interpreted as the limiting value of a Wald statistic for the lag polynomial b taking $\bar{b}$ as the point estimate.

---

[2] The distribution result used here is explicit in Brockwell and Davis (1988). It is implicit in Berk (1974) from a simple Fourier inversion of his result on the distribution of the spectral density estimates. Note that Brockwell and Davis (1988) impose the assumption that $b_0 = 1$, which we do not include here.

[3] Specifically, $\bar{J}_m \approx \Lambda_m \bar{B}_m \Lambda_m^* = \Lambda_m^{*\prime} \bar{B}_m^* \Lambda_m'$, and thus $\bar{J}_m \bar{J}_m' \approx \Lambda_m \bar{B}_m \Lambda_m^* \Lambda_m \bar{B}_m^* \Lambda_m^* = \Lambda_m \left(\bar{B}_m \bar{B}_m^* \right) \Lambda_m^* = \Lambda_m \bar{F}_m \Lambda_m^*$, where $\bar{B}_m$ is the diagonal matrix of the discrete Fourier transform of $\left[\bar{b}_0, \bar{b}_1, ..., \bar{b}_m \right]$. Again, the aproximations become exact as $m \to \infty$.

D Lifetime utility (assumption 3)

As discussed in the text, the agent's expectation of future consumption growth, $E_t[\Delta c_{t+j}|\Theta]$ is equal to expected consumption growth at date $t + j$ given the past observed history of consumption growth and the assumption that ε_t has mean zero. Given that the agent believes that the model $\Theta = \{b, \mu, \sigma^2\}$ drives consumption growth, we can write the innovations implied by that model as

$$\varepsilon_t^\Theta = (\Delta c_t - \mu - a(L)(\Delta c_{t-1} - \mu)) \tag{D.1}$$

That is, ε_t^Θ is the innovation that the agent would believe occurred given the observed history of consumption growth and the model Θ. The agent's subjective expectations for future consumption growth are then

$$E_t[\Delta c_{t+j}|\Theta] = \mu + \sum_{k=0}^{\infty} b_{k+j}\varepsilon_{t-k}^\Theta \tag{D.2}$$

with subjective distribution

$$\frac{\Delta c_{t+1} - E_t[\Delta c_{t+1}|\Theta]}{\sigma} \sim N(0,1) \tag{D.3}$$

We guess that $v(\Delta c^t; \Theta)$ takes the form

$$v(\Delta c^t; \Theta) = c_t + \bar{k} + \sum_{j=0}^{\infty} k_j \varepsilon_{t-j}^\Theta \tag{D.4}$$

Inserting into the recursion for lifetime utility yields

$$\bar{k} + \sum_{j=0}^{\infty} k_j \varepsilon_{t-j}^\Theta = \frac{\beta}{1-\alpha} \log E_t \left[\exp\left(\left(\begin{array}{c} \bar{k} + \mu + (k_0 + 1)\varepsilon_{t+1}^\Theta \\ + \sum_{j=1}^{\infty}(k_j + b_j)\varepsilon_{t-j+1}^\Theta \end{array} \right)(1-\alpha) \right) \middle| \Theta \right] \tag{D.5}$$

$$= \beta(\bar{k} + \mu) + \beta \sum_{j=0}^{\infty}(k_{j+1} + b_{j+1})\varepsilon_{t-j}^\Theta + \beta \frac{1-\alpha}{2}(k_0 + b_0)^2 \sigma^2 \tag{D.6}$$

Matching the coefficients on each side of the equality yields

$$k_j = \beta(k_{j+1} + b_{j+1}) \tag{D.7}$$

$$v(\Delta c^t; b) = c_t + \frac{\beta}{1-\beta}\frac{1-\alpha}{2}b(\beta)^2\sigma^2 + \frac{\beta}{1-\beta}\mu + \sum_{k=1}^{\infty}\beta^k \sum_{j=0}^{\infty} b_{j+k}\varepsilon_{t-j}^\Theta \tag{D.8}$$

$$= c_t + \frac{\beta}{1-\beta}\frac{1-\alpha}{2}b(\beta)^2\sigma^2 + \frac{\beta}{1-\beta}\mu + \sum_{j=0}^{\infty}\left(\sum_{k=1}^{\infty}\beta^k b_{j+k}\right)\varepsilon_{t-j}^\Theta \tag{D.9}$$

$$= c_t + \frac{\beta}{1-\beta}\frac{1-\alpha}{2}b(\beta)^2\sigma^2 + \sum_{k=1}^{\infty}\beta^k E_t[\Delta c_{t+k}|\Theta] \tag{D.10}$$

E   Multiplier preference interpretation

In our main analysis, we model agents as having Epstein–Zin preferences. Such preferences are observationally equivalent (in the sense that they rank all consumption streams identically) to Hansen and Sargent's (2001) multiplier preferences. In that model, agents have log utility over consumption, but they form expectations using a worst-case model over *innovations* to the consumption process. Specifically, their preferences are obtained through

$$v_t = \min_{h_{t+1}} c_t + \beta \left(E_t \left[h_{t+1} v_{t+1} \right] + \psi E_t \left[h_{t+1} \log h_{t+1} \right] \right) \tag{E.1}$$

where h_{t+1} is a change of measure with $E[h_{t+1}] = 1$. h_{t+1} represents an alternative distribution of the innovations to the state variables at date $t + 1$. In this model, agents select an alternative distribution for *innovations* (instead of a full distribution over consumption growth) penalizing alternative distributions based on their KL divergence ($E_t \left[h_{t+1} \log h_{t+1} \right]$).

Inserting the value of h_{t+1} that solves the minimization problem yields

$$v_t = c_t - \beta \psi \log E_t \exp \left(-\psi^{-1} v_{t+1} \right) \tag{E.2}$$

That is, the Epstein–Zin preferences used in the main text can be interpreted as multiplier preferences with $-\psi^{-1} = (1 - \alpha)$.

We can thus interpret the model described in the paper as involving two layers of robustness, or two evil agents. First, there is an evil agent who, in a timeless manner, selects a full worst-case process for consumption growth. Next, taking the preferences (E.2) a second evil agent causes further deviations in the innovations to that process.

The second evil agent's minimization problem is (E.1), and the minimized value function is then (E.2), which is exactly the preference specification that is minimized in the main text. In other words, both the minimization problem over the full models for consumption growth that we study and also the minimization over one-step deviations – which induces Epstein–Zin preferences – depend on a KL divergence penalty.

A natural benchmark is to equalize the penalty on the KL divergence that is involved in both minimization problems. Since the entropy penalty for the second agent is applied in every period, we naturally scale it up by the discount rate. That is,

$$\lambda = \psi / (1 - \beta) \tag{E.3}$$

Which immediately yields a connection between λ and α,

$$\lambda = \frac{1}{1 - \beta} \frac{1}{\alpha - 1} \tag{E.4}$$

$$\alpha = 1 + \frac{\lambda^{-1}}{1 - \beta} \tag{E.5}$$

13

F   Asset prices and expected returns

F.1   Pricing a levered consumption claim

Using the Campbell–Shiller (1988) approximation, the return on a levered consumption claim can be approximated as (with the approximation becoming more accurate as the length of a time period shrinks)

$$r_{t+1} = \delta_0 + \delta p d_{t+1} + \gamma \Delta c_{t+1} - p d_t \tag{F.1}$$

where δ is a linearization parameter slightly less than 1.

We guess that

$$p d_t = \bar{h} + \sum_{j=0}^{\infty} h_j \Delta c_{t-j} \tag{F.2}$$

for a set of coefficients $\bar{h}$ and h_j.

The innovation to lifetime utility is

$$v_{t+1} - E_t\left[v_{t+1}|b^w\right] = \sum_{k=0}^{\infty} \beta^k \Delta E_{t+1}\left[\Delta c_{t+k+1}|\Theta^w\right] \tag{F.3}$$

$$= b^w(\beta)\,\varepsilon_{t+1}^{\Theta^w} \tag{F.4}$$

where the investor prices assets as though $\varepsilon_{t+1}^{\Theta^w}$ is a standard normal.

The pricing kernel can therefore be written as

$$M_{t+1} = \beta \exp\left(-\Delta c_{t+1} + (1-\alpha)\,b^w(\beta)\,\varepsilon_{t+1}^{\Theta^w} - \frac{(1-\alpha)^2}{2} b^w(\beta)^2\,\sigma_w^2\right) \tag{F.5}$$

The pricing equation for the levered consumption claim is

$$
0 = \log E_t\left[\beta \exp\left(\begin{array}{c}\delta_0 + (\delta-1)\bar{h} + (\delta h_0 + \gamma - 1)\Delta c_{t+1} + \sum_{j=0}^{\infty}(\delta h_{j+1} - h_j)\Delta c_{t-j} \\ + (1-\alpha)\,b^w(\beta)\,\varepsilon_{t+1}^{\Theta^w} - \frac{(1-\alpha)^2}{2} b^w(\beta)^2\,\sigma_w^2\end{array}\right)\Big|\Theta^w\right] \tag{F.6}
$$

$$
= (\delta h_0 + \gamma - 1)\left((1 - a^w(1))\mu^w + a^w(L)\Delta c_t\right) + \sum_{j=0}^{\infty}(\delta h_{j+1} - h_j)\Delta c_{t-j}
$$

$$
+ \delta_0 + \left(\frac{1}{2}(\delta h_0 + \gamma - 1)^2 + (\delta h_0 + \gamma - 1)(1-\alpha)b^w(\beta)\right)\sigma_w^2 + (\delta-1)\bar{h} + \log\beta \tag{F.7}
$$

Matching coefficients on Δc_{t-j} and on the constant yields two equations,

$$
(\delta-1)\bar{h} + \log\beta + \delta_0 = -\left(\frac{1}{2}(\delta h_0 + \gamma - 1)^2 + (\delta h_0 + \gamma - 1)(1-\alpha)b^w(\beta)\right)\sigma_w^2
$$

$$
- (\delta h_0 + \gamma - 1)(1 - a^w(1))\mu^w \tag{F.8}
$$

$$
(\delta h_{j+1} - h_j) = -(\delta h_0 + \gamma - 1)a_j^w \tag{F.9}
$$

And thus

$$h_0 = \frac{(\gamma - 1) a^w (\delta)}{1 - \delta a^w (\delta)} \tag{F.10}$$

and

$$\delta h_0 + \gamma - 1 = \frac{\gamma - 1}{1 - \delta a^w (\delta)} \tag{F.11}$$

Note then that

$$var_w (r_{m,t+1}) = (\delta h_0 + \gamma)^2 \sigma_w^2 \tag{F.12}$$

$$cov_w (r_{m,t+1}, m_{t+1}) = (\delta h_0 + \gamma) (-1 + (1 - \alpha) b^w (\beta)) \sigma_w^2 \tag{F.13}$$

F.2   The risk-free rate

For the risk-free rate, we have

$$r_{f,t+1} = -\log E_t \left[\beta \exp \left(-\Delta c_{t+1} + (1 - \alpha) b^w (\beta) \varepsilon_{t+1}^{\Theta^w} - \frac{(1 - \alpha)^2}{2} b^w (\beta)^2 \sigma_w^2 \right) | \Theta^w \right] \tag{F.14}$$

$$= -\log \beta + (1 - a^w (1)) \mu^w + a^w (L) \Delta c_t - \frac{1}{2} \sigma_w^2 + (1 - \alpha) b^w (\beta) \sigma_w^2 \tag{F.15}$$

$$= -\log \beta + \mu^w + a^w (L) (\Delta c_t - \mu^w) - \frac{1}{2} \sigma_w^2 + (1 - \alpha) b^w (\beta) \sigma_w^2 \tag{F.16}$$

F.3   Expected excess returns

The expected excess return on the levered consumption claim from the perspective of an econometrician who believes that consumption dynamics are the point estimate $\bar{\Theta}$ is

$$E_t \left[r_{t+1} | \bar{\Theta} \right] = E_t \left[\delta_0 + (\delta - 1) \bar{h} + (\delta h_0 + \gamma) \Delta c_{t+1} + \sum_{j=0}^{\infty} (\delta h_{j+1} - h_j) \Delta c_{t-j} | \bar{\Theta} \right] \tag{F.17}$$

$$= \delta_0 + (\delta - 1) \bar{h} - (\delta h_0 + \gamma - 1) a^w (L) \Delta c_t + E_t \left[(\delta h_0 + \gamma) \Delta c_{t+1} | \bar{\Theta} \right] \tag{F.18}$$

$$= \delta_0 + (\delta - 1) \bar{h} + (- (\delta h_0 + \gamma - 1) a^w (L) + (\delta h_0 + \gamma) a (L)) \Delta c_t$$
$$+ (\delta h_0 + \gamma) (1 - a (1)) \mu \tag{F.19}$$

$$E_t \left[r_{t+1} - r_{f,t+1} | \bar{\Theta} \right] = \delta_0 + (\delta - 1) \bar{h} + (- (\delta h_0 + \gamma - 1) a^w (L) + (\delta h_0 + \gamma) a (L)) \Delta c_t$$
$$+ (\delta h_0 + \gamma) (1 - a (1)) \mu$$
$$+ \log \beta - (1 - a^w (1)) \mu^w - a^w (L) \Delta c_t + \frac{1}{2} \sigma_w^2 - (1 - \alpha) b^w (\beta) \sigma_w^2 \tag{F.20}$$

Inserting the formula for $(\delta - 1)\bar{h} + \log \beta + \delta_0$ from above yields

$$
\begin{aligned}
(\delta - 1)\bar{h} + \log \beta + \delta_0 \quad = \quad & -\left(\frac{1}{2}(\delta h_0 + \gamma - 1)^2 + (\delta h_0 + \gamma - 1)(1 - \alpha)b^w(\beta)\right)\sigma_w^2 \\
& -(\delta h_0 + \gamma - 1)(1 - a^w(1))\mu^w
\end{aligned}
\tag{F.21}
$$

$$
\begin{aligned}
E_t\left[r_{t+1} - r_{f,t+1}|\bar{\Theta}\right] \quad = \quad & (\delta h_0 + \gamma)(a(L) - a^w(L))(\Delta c_t - \mu) \\
& + (\delta h_0 + \gamma)(1 - a^w(1))(\mu - \mu^w) \\
& - \frac{1}{2}var_w(r_{m,t+1}) - cov_w(r_{m,t+1}, m_{t+1})
\end{aligned}
\tag{F.22}
$$

where

$$
var_w(r_{m,t+1}) \quad = \quad (\delta h_0 + \gamma)^2 \sigma_w^2
\tag{F.23}
$$

$$
cov_w(r_{m,t+1}, m_{t+1}) \quad = \quad (\delta h_0 + \gamma)(-1 + (1 - \alpha)b^w(\beta))\sigma_w^2
\tag{F.24}
$$

Substituting in

$$
\delta h_0 + \gamma = \delta \frac{(\gamma - 1)a^w(\delta)}{1 - \delta a^w(\delta)} + \gamma = \frac{\gamma - \delta a^w(\delta)}{1 - \delta a^w(\delta)}
\tag{F.25}
$$

yields the result from the text.

$$
\begin{aligned}
E_t\left[r_{t+1} - r_{f,t+1}|\bar{\Theta}\right] \quad = \quad & \frac{\gamma - \delta a^w(\delta)}{1 - \delta a^w(\delta)}(a(L) - a^w(L))(\Delta c_t - \mu) \\
& + \frac{\gamma - \delta a^w(\delta)}{1 - \delta a^w(\delta)}(1 - a^w(1))(\mu - \mu^w) \\
& - \frac{1}{2}var_w(r_{m,t+1}) - cov_w(r_{m,t+1}, m_{t+1})
\end{aligned}
\tag{F.26}
$$

F.4 The behavior of interest rates

The mean of the risk-free rate is

$$
-\log \beta + (1 - a^w(1))\mu^w + a^w(1)\mu - \frac{1}{2}\sigma_w^2 + (1 - \alpha)b^w(\beta)\sigma_w^2
\tag{F.27}
$$

And its standard deviation is

$$
std(a^w(L)\Delta c_t)
\tag{F.28}
$$

When consumption growth is white noise, this is

$$
std(a^w(L)\Delta c_t) \quad = \quad std\left((\beta - \theta)\sum_{j=0}^{\infty}\theta^j \Delta c_{t-j}\right)
\tag{F.29}
$$

$$
= \quad (\beta - \theta)\frac{\sigma_{\Delta c}}{\sqrt{1 - \theta^2}}
\tag{F.30}
$$

We denote the log price on date t of a claim to a unit of consumption paid on date $t+j$ as $p_{j,t}$, and we guess that

$$p_{j,t} = \phi^{(j)}(L)(\Delta c_t - \mu^w) + n_j \tag{F.31}$$

for a lag polynomial $\phi^{(j)}$ and a constant n_j that differ with maturity.

The pricing condition for a bond is

$$M_{t+1} = \beta \exp\left(-\Delta c_{t+1} + (1-\alpha) b^w(\beta) \varepsilon_{t+1}^{\Theta^w} - \frac{(1-\alpha)^2}{2} b^w(\beta)^2 \sigma_w^2\right) \tag{F.32}$$

$$\phi^{(j)}(L)\Delta c_t + n_j = \log E_t\left[\exp\left(\begin{array}{c}\log\beta - \Delta c_{t+1} + (1-\alpha) b^w(\beta)\varepsilon_{t+1}^{\Theta^w} \\ -\frac{(1-\alpha)^2}{2} b^w(\beta)^2 \sigma_w^2 + \phi^{(j-1)}(L)(\Delta c_{t+1} - \mu^w) + n_{j-1}\end{array}\right)|\Theta^w\right] \tag{F.33}$$

$$= \log\beta + \left(\phi_0^{(j-1)} - 1\right)(\mu^w + a^w(L)(\Delta c_t - \mu^w)) - \phi_0^{(j-1)}\mu^w + \sum_{k=0}^{\infty}\phi_{k+1}^{(j-1)}(\Delta c_{t-k} - \mu^w)$$

$$-\frac{(1-\alpha)^2}{2} b^w(\beta)^2 \sigma_w^2 + n_{j-1} + \frac{1}{2}\left((1-\alpha) b^w(\beta) - 1 + \phi_0^{(j-1)}\right)^2 \sigma_w^2 \tag{F.34}$$

Matching coefficients yields,

$$\phi^{(j)}(L) = \left(\phi_0^{(j-1)} - 1\right) a^w(L) + \sum_{k=0}^{\infty}\phi_{k+1}^{(j-1)}L^k \tag{F.35}$$

$$n_j = \log\beta - \mu^w - \frac{(1-\alpha)^2}{2} b^w(\beta)^2 \sigma_w^2 + n_{j-1} + \frac{1}{2}\left((1-\alpha) b^w(\beta) - 1 + \phi_0^{(j-1)}\right)^2 \sigma_w^2 \tag{F.36}$$

We also have the boundary condition that the price of a unit of consumption today is 1, so that $n_0 = 0$ and $\phi^{(0)}(L) = 0$. Note that the mean price of any of these claims is

$$E[p_{j,t}] = \phi^{(j)}(1)(\mu - \mu^w) + n_j \tag{F.37}$$

F.5   Results used in table 1

Under the worst-case, consumption growth follows an ARMA(1,1). We have

$$\Delta c_t = \beta\Delta c_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1} \tag{F.38}$$

$$a^w(L) = (\beta - \theta)\sum_{j=0}^{\infty}\theta^j L^j \tag{F.39}$$

where $\theta \equiv (1 - \varphi)\beta$ and φ is obtained above. We then have

$$a^w(\delta) = \frac{\beta - \theta}{1 - \theta\delta} \tag{F.40}$$

$$a^w(1) = \frac{\beta - \theta}{1 - \theta} \tag{F.41}$$

$$b_j = \beta^{j-1}(\beta - \theta) \tag{F.42}$$

For the coefficients in the price/dividend ratio, we have

$$(\delta h_{j+1} - h_j) = -(\delta h_0 + \gamma - 1) a_j^w \tag{F.43}$$

$$h_j = (\delta h_0 + \gamma - 1) \sum_{k=0}^{\infty} \delta^k a_{j+k}^w \tag{F.44}$$

$$= (\delta h_0 + \gamma - 1) \sum_{k=0}^{\infty} \delta^k (\beta - \theta) \theta^{j+k} \tag{F.45}$$

$$= (\delta h_0 + \gamma - 1)(\beta - \theta) \frac{\theta^j}{1 - \delta\theta} \tag{F.46}$$

And thus

$$pd_t = \bar{h} + \frac{(\delta h_0 + \gamma - 1)(\beta - \theta)}{1 - \delta\theta} \sum_{j=0}^{\infty} \theta^j \Delta c_{t-j} \tag{F.47}$$

The standard deviation of the price/dividend ratio under the true white-noise process for consumption growth is then

$$std(pd_t) = \frac{(\delta h_0 + \gamma - 1)(\beta - \theta)}{1 - \delta\theta} \frac{\bar{\sigma}}{\sqrt{1 - \theta^2}} \tag{F.48}$$

F.6 Returns in the absence of model uncertainty

When there is no model uncertainty, the SDF is the same as in our main case, but everything is calculated using the benchmark model instead of the worst case. For interest rates, then

$$r_{f,t+1} = -\log E_t \left[\beta \exp\left(-\Delta c_{t+1} + (1-\alpha)\varepsilon_{t+1}^{\bar{\Theta}} - \frac{(1-\alpha)^2}{2}\bar{\sigma}^2 \right) | \bar{\Theta} \right] \tag{F.49}$$

$$= -\log\beta + \bar{\mu} - \frac{1}{2}\bar{\sigma}^2 + (1-\alpha)\bar{\sigma}^2 \tag{F.50}$$

$$E[r_{f,t+1}] = -\log\beta + \bar{\mu} - \frac{1}{2}\bar{\sigma}^2 + (1-\alpha)\bar{\sigma}^2 \tag{F.51}$$

$$std(r_f) = 0 \tag{F.52}$$

For the price/dividend ratio, we have $h_j = 0$ for all j, which implies

$$var\left(r_{m,t+1}\right) \;=\; \gamma^2\bar{\sigma}^2 \tag{F.53}$$

$$cov\left(r_{m,t+1}, m_{t+1}\right) \;=\; -\alpha\gamma\bar{\sigma}^2 \tag{F.54}$$

The standard deviation of the log pricing kernel is

$$std\left(m_{t+1}\right) = -\alpha\bar{\sigma} \tag{F.55}$$

G Dividends cointegrated with consumption

Two drawbacks of our main specification for dividends are that it implies that dividend and consumption growth are perfectly correlated and that it implies dividends are slightly more volatile than observed empirically. To generate more realistic behavior for dividends, we now consider a setting where dividends and consumption are cointegrated. We want to exactly match three major features of the joint dynamics of consumption and dividends: the standard deviations of the two series, the correlation between the two series, and the fact that dividends appear to be smoothed over time (Marsh and Merton (1987); Chen, Da, and Priestley (2012)).

We assume the following model holds

$$d_t = \gamma g_c\left(L\right) c_t + g_\zeta\left(L\right) \zeta_t \tag{G.1}$$

where ζ_t is a normally distributed innovation with unit variance and $g_\zeta\left(L\right)$ is a lag polynomial. We assume that $g_\zeta\left(L\right)\zeta_t$ is stationary with finite variance (the case where $g_\zeta\left(L\right)$ has a unit root would correspond to a situation where dividends and consumption are no longer cointegrated, but their growth rates are correlated).

The function $g_c\left(L\right)$ is what models dividends as a smoothed form of consumption. We normalize the lag polynomial so that $g_c\left(1\right) = 1$. As a simple example, if $g_c\left(L\right) = 1 + L + L^2$, then dividends are a three-year moving average of consumption plus noise $(g_\zeta\left(L\right)\zeta_t)$. Allowing a lagged response of dividends to fundamentals (consumption) allows us to model the dividend smoothing observed in Marsh and Merton (1987) and Chen, Da, and Priestley (2012).

γ represents the cointegrating coefficient between dividends and consumption – it determines how much the long-run level of dividends responds to a unit shock to the long-run level of consumption.

In terms of growth rates we have

$$\Delta d_t \;=\; \gamma g_c\left(L\right)\Delta c_t + \tilde{g}_\zeta\left(L\right)\zeta_t \tag{G.2}$$

$$\tilde{g}_\zeta\left(L\right) \;\equiv\; g_\zeta\left(L\right)\left(1 - L\right) \tag{G.3}$$

We then recapitulate the analysis from above. Specifically, we add a superscript C to the coefficients

in the price/dividend function to yield the guess

$$pd_t^C = \bar{h}^C + \sum_{j=0}^{\infty} \left(h_{c,j}^C \Delta c_{t-j} + h_{\zeta,j}^C \zeta_{t-j} \right) \tag{G.4}$$

$$r_{t+1}^C = \delta_0 + \delta pd_{t+1}^C + \gamma g_c \left(L \right) \Delta c_{t+1} + \tilde{g}_\zeta \left(L \right) \zeta_{t+1} - pd_t^C \tag{G.5}$$

The pricing equation for the dividend claim is

$$
\begin{aligned}
0 &= \log E_t \left[\beta \exp \left(
\begin{array}{c}
\delta_0 + (\delta - 1)\bar{h} + \left(\delta h_{c,0}^C + \gamma g_{c,0} - 1 \right) \Delta c_{t+1} \\
+ \sum_{j=0}^{\infty} \left(\delta h_{c,j+1}^C + \gamma g_{c,j+1} - h_{c,j}^C \right) \Delta c_{t-j} \\
+ (1 - \alpha) b^w \left(\beta \right) \varepsilon_{t+1}^{\Theta^w} - \frac{(1-\alpha)^2}{2} b^w \left(\beta \right)^2 \sigma_w^2 \\
+ \left(\delta h_{\zeta,0}^C + \tilde{g}_{\zeta,0} \right) \zeta_{t+1} + \sum_{j=0}^{\infty} \left(\delta h_{\zeta,j+1}^C - h_{\zeta,j}^C + \tilde{g}_{\zeta,j+1} \right) \zeta_{t-j}
\end{array}
\right) | \Theta^w \right] \\
&= \left(\delta h_{c,0}^C + \gamma g_{c,0} - 1 \right) \left((1 - a^w (1)) \mu^w + a^w (L) \Delta c_t \right) + \sum_{j=0}^{\infty} \left(\delta h_{c,j+1}^C + \gamma g_{c,j+1} - h_{c,j}^C \right) \Delta c_{t-j} \\
&\quad + \delta_0 + \left(\frac{1}{2} \left(\delta h_{c,0}^C + \gamma g_{c,0} - 1 \right)^2 + \left(\delta h_{c,0}^C + \gamma g_{c,0} - 1 \right) (1 - \alpha) b^w \left(\beta \right) \right) \sigma_w^2 + (\delta - 1)\bar{h}^C + \log \beta \\
&\quad + \sum_{j=0}^{\infty} \left(\delta h_{\zeta,j+1}^C - h_{\zeta,j}^C + \tilde{g}_{\zeta,j+1} \right) \zeta_{t-j} + \frac{1}{2} \left(\delta h_{\zeta,0}^C + \tilde{g}_{\zeta,0} \right)^2 \sigma_\zeta^2
\end{aligned} \tag{G.6, G.7}
$$

Matching coefficients on Δc_{t-j}, ζ_{t-j}, and on the constant yields three equations,

$$
\begin{aligned}
(\delta - 1)\bar{h}^C + \log \beta + \delta_0 &= -\left(\frac{1}{2} \left(\delta h_{c,0}^C + \gamma g_{c,0} - 1 \right)^2 + \left(\delta h_{c,0}^C + \gamma g_{c,0} - 1 \right) (1 - \alpha) b^w \left(\beta \right) \right) \sigma_w^2 \\
&\quad - \left(\delta h_{c,0}^C + \gamma g_{c,0} - 1 \right) (1 - a^w (1)) \mu^w - \frac{1}{2} \left(\delta h_{\zeta,0}^C + \tilde{g}_{\zeta,0} \right)^2 \sigma_\zeta^2
\end{aligned} \tag{G.8}
$$

$$\delta h_{c,j+1}^C - h_{c,j}^C = -\left(\delta h_{c,0}^C + \gamma g_{c,0} - 1 \right) a_j^w - \gamma g_{c,j+1} \tag{G.9}$$

$$\delta h_{\zeta,j+1}^C - h_{\zeta,j}^C = -\tilde{g}_{\zeta,j+1} \tag{G.10}$$

And thus

$$h_{c,j}^C = \delta h_{c,j+1}^C + \left(\delta h_{c,0}^C + \gamma g_{c,0} - 1 \right) a_j^w + \gamma g_{c,j+1} \tag{G.11}$$

$$h_{c,0}^C = \sum_{j=0}^{\infty} \left(\delta h_{c,0}^C + \gamma g_{c,0} - 1 \right) a_j^w \delta^j + \delta^{-1} \sum_{j=1}^{\infty} \gamma g_{c,j} \delta^j \tag{G.12}$$

$$\delta h_{c,0}^C + \gamma g_{c,0} - 1 = \left(\delta h_0^C + \gamma g_{c,0} - 1 \right) \delta a^w \left(\delta \right) + \gamma g_c \left(\delta \right) - 1 \tag{G.13}$$

$$\delta h_{c,0}^C + \gamma g_{c,0} - 1 = \frac{\gamma g_c \left(\delta \right) - 1}{1 - \delta a^w \left(\delta \right)} \tag{G.14}$$

$$\delta h_{c,0}^{C} + \gamma g_{c,0} = \frac{\gamma g_c(\delta) - \delta a^w(\delta)}{1 - \delta a^w(\delta)} \tag{G.15}$$

Note that when $g_c(L) = 1$, $g_c(\delta) = 1$, and $g_c = 1$, so the above equation reduces to precisely what is obtained above for $\delta h_{c,0}^{C} + \gamma - 1$. Furthermore, note that for $\delta \approx 1$, $g_c(\delta) \approx g_c(1) = 1$.

For the coefficients on ζ, we have

$$\delta h_{\zeta,0}^{C} + \tilde{g}_{\zeta,0} = \tilde{g}_\zeta(\delta) \tag{G.16}$$

Note then that

$$var_w(r_{m,t+1}) = \left(\delta h_{c,0}^{C} + \gamma g_{c,0}\right)^2 \sigma_w^2 + \tilde{g}_\zeta(\delta)^2 \tag{G.17}$$

$$cov_w(r_{m,t+1}, m_{t+1}) = \left(\delta h_{c,0}^{C} + \gamma g_{c,0}\right)(-1 + (1 - \alpha) b^w(\beta)) \sigma_w^2 \tag{G.18}$$

So what we have is that the variance of the return is simply increased through the additional noise added to dividends, $\tilde{g}_\chi(\delta)^2$, while the covariance is unaffected. Furthermore, we note that $\tilde{g}_\xi(1) = 0$, so for δ close to 1, we would expect the term $\tilde{g}_\chi(\delta)^2$ to be small.

G.1   Calibration

We leave the calibration of γ the same as in the main text. We also maintain the calibration that consumption growth in the benchmark model is white noise. We then have

$$corr(\Delta c, \Delta d) = \gamma g_{c,0} \frac{std(\Delta c)}{std(\Delta d)} \tag{G.19}$$

Following Bansal and Yaron (2004) (who use real dividend growth for the CRSP value-weighted index), we set $std(\Delta d) = 0.057$ and $corr(\Delta d, \Delta c) = 0.55$, which then implies $g_{c,0} = 0.44$ (given the value of γ from table 1). For the sake of simplicity, we assume that g_c is a simple MA(1), yielding $g_{c,,1} = 0.56$ and $g_{c,j} = 0$ for $j > 1$.

Finally, we calibrate $\tilde{g}_\zeta$ to match the variance of dividend growth. We have

$$var(\Delta d) = \gamma^2 \left(g_{c,0}^2 + g_{c,1}^2\right) var(\Delta c) + var(\tilde{g}_\zeta(\delta) \zeta_t) \tag{G.20}$$

Again, for the same of simplicity, we assume that the error $g_\zeta(L) = g_{\zeta,0}$, which implies that $\tilde{g}_\xi(L) = g_{\zeta,0} - g_{\zeta,0} L$. Finally,

$$var(\Delta d) = \gamma^2 \left(g_{c,0}^2 + g_{c,1}^2\right) var(\Delta c) + 2g_{\zeta,0}^2 \tag{G.21}$$

(under the normalization that $var(\zeta_t) = 1$). Inserting the calibrated values for the other parameters, we obtain

$$g_{\zeta,0}^2 = \frac{1}{2}\left(var(\Delta d) - \gamma^2\left(g_{c,0}^2 + g_{c,1}^2\right) var(\Delta c)\right) \tag{G.22}$$

$$g_{\zeta,0} = 0.019 \tag{G.23}$$

That is, the final model of dividends is

$$d_t \quad = \quad 2.13c_t + 2.67c_{t-1} + 0.019\zeta_t \tag{G.24}$$

$$\zeta_t \quad \sim \quad N(0,1) \tag{G.25}$$

G.2 Expected excess returns

The expected excess return on the levered consumption claim from the perspective of an econometrician who believes that consumption dynamics are the point estimate $\bar{\Theta}$ is

$$
E_t\left[r_{t+1}|\bar{\Theta}\right] \quad = \quad E_t\left[
\begin{array}{c}
\delta_0 + (\delta - 1)\,\bar{h} + \left(\delta h^C_{c,0} + \gamma g_{c,0}\right)\Delta c_{t+1} \\
+ \sum_{j=0}^{\infty}\left(\delta h^C_{c,j+1} + \gamma g_{c,j+1} - h^C_{c,j}\right)\Delta c_{t-j} \\
+ \left(\delta h^C_{\zeta,0} + \tilde{g}_{\zeta,0}\right)\zeta_{t+1} + \sum_{j=0}^{\infty}\left(\delta h^C_{\zeta,j+1} - h^C_{\zeta,j} + \tilde{g}_{\zeta,j+1}\right)\zeta_{t-j}
\end{array}
\Bigg| \bar{\Theta}\right] \tag{G.26}
$$

$$
= \quad \delta_0 + (\delta - 1)\,\bar{h} + \left(\delta h^C_{c,0} + \gamma g_{c,0}\right) E_t\left[\Delta c_{t+1}|\bar{\Theta}\right] - \left(\delta h^C_{c,0} + \gamma g_{c,0} - 1\right) a^w(L)\Delta c_t \tag{G.27}
$$

$$
\begin{aligned}
= \quad & \delta_0 + (\delta - 1)\,\bar{h} - \left(\delta h^C_{c,0} + \gamma g_{c,0} - 1\right) a^w(L)\Delta c_t \\
& + \left(\delta h^C_{c,0} + \gamma g_{c,0}\right)\left(\mu + a(L)(\Delta c_t - \mu)\right)
\end{aligned} \tag{G.28}
$$

$$
\begin{aligned}
E_t\left[r_{t+1} - r_{f,t+1}|\bar{\Theta}\right] \quad = \quad & \delta_0 + (\delta - 1)\,\bar{h} + \left(\delta h^C_{c,0} + \gamma g_{c,0}\right)\left(a(L) - a^w(L)\right)(\Delta c_t - \mu) \\
& + \left(\delta h^C_{c,0} + \gamma g_{c,0}\right)\left(1 - a^w(1)\right)\mu \\
& + \log\beta - \left(1 - a^w(1)\right)\mu^w_t + \frac{1}{2}\sigma^2_w - (1-\alpha)b^w(\beta)\sigma^2_w
\end{aligned} \tag{G.29}
$$

Inserting the formula for $(\delta - 1)\,\bar{h} + \log\beta + \delta_0$ from above yields

$$
\begin{aligned}
E_t\left[r_{t+1} - r_{f,t+1}|\bar{\Theta}\right] \quad = \quad & \left(\delta h^C_{c,0} + \gamma g_{c,0}\right)\left(a(L) - a^w(L)\right)(\Delta c_t - \mu) \\
& + \left(\delta h^C_{c,0} + \gamma g_{c,0}\right)\left(1 - a^w(1)\right)(\mu - \mu^w) \\
& + \frac{1}{2}\sigma^2_w - (1-\alpha)b^w(\beta)\sigma^2_w \\
& - \left(\frac{1}{2}\left(\delta h^C_{c,0} + \gamma g_{c,0} - 1\right)^2 + \left(\delta h^C_{c,0} + \gamma g_{c,0} - 1\right)(1-\alpha)b^w(\beta)\right)\sigma^2_w \\
& - \frac{1}{2}\left(\delta h^C_{\zeta,0} + \tilde{g}_{\zeta,0}\right)^2 \sigma^2_\zeta
\end{aligned} \tag{G.30}
$$

$$
\begin{aligned}
E_t\left[r_{t+1} - r_{f,t+1}|\bar{\Theta}\right] \quad = \quad & \frac{\gamma g_c(\delta) - \delta a^w(\delta)}{1 - \delta a^w(\delta)}\left(a(L) - a^w(L)\right)(\Delta c_t - \mu) \\
& + \frac{\gamma g_c(\delta) - \delta a^w(\delta)}{1 - \delta a^w(\delta)}\left(1 - a^w(1)\right)(\mu - \mu^w) \\
& - cov_w\left(r_{m,t+1}, m_{t+1}\right) - \frac{1}{2}var_w\left(r_{m,t+1}\right)
\end{aligned} \tag{G.31}
$$

22

where, from above,

$$var_w\left(r_{m,t+1}\right) = \left(\frac{\gamma g_c\left(\delta\right) - \delta a^w\left(\delta\right)}{1 - \delta a^w\left(\delta\right)}\right)^2 \sigma_w^2 + \tilde{g}_\zeta\left(\delta\right)^2 \tag{G.32}$$

$$cov_w\left(r_{m,t+1}, m_{t+1}\right) = \frac{\gamma g_c\left(\delta\right) - \delta a^w\left(\delta\right)}{1 - \delta a^w\left(\delta\right)}\left(-1 + \left(1 - \alpha\right) b^w\left(\beta\right)\right)\sigma_w^2 \tag{G.33}$$

G.3   Price/dividend ratio

$$\delta h_{c,j+1}^C - h_{c,j}^C = -\left(\delta h_{c,0}^C + \gamma g_{c,0} - 1\right) a_j^w - \gamma g_{c,j+1} \tag{G.34}$$

$$\delta h_{\zeta,j+1}^C - h_{\zeta,j}^C = -\tilde{g}_{\zeta,j+1} \tag{G.35}$$

$$h_{\zeta,0} = g_{\zeta,1} \tag{G.36}$$

$$h_{c,0} = \frac{\gamma g_c\left(\delta\right) - 1}{1 - \delta a^w\left(\delta\right)}\left(\beta - \theta\right)\frac{1}{1 - \delta\theta} + \gamma g_{c,1} \tag{G.37}$$

$$h_{c,j} = \frac{\gamma g_c\left(\delta\right) - 1}{1 - \delta a^w\left(\delta\right)}\left(\beta - \theta\right)\frac{\theta^j}{1 - \delta\theta} \tag{G.38}$$

So the standard deviation of the pricing kernel is now

$$pd_t^C = \sum_{j=0}^\infty h_{c,j}^C \Delta c_{t-j} + g_{\zeta,1}\zeta_t \tag{G.39}$$

$$var\left(pd_t^C\right) = \left(\frac{\gamma g_c\left(\delta\right) - 1}{1 - \delta a^w\left(\delta\right)}\frac{\left(\beta - \theta\right)}{1 - \delta\theta}\right)^2 \frac{\bar{\sigma}^2}{1 - \theta^2} + g_{\zeta,1}^2 \tag{G.40}$$

$$corr\left(pd_t, pd_{t-4}\right) = \frac{\theta^4\left(\frac{\gamma g_c\left(\delta\right) - 1}{1 - \delta a^w\left(\delta\right)}\frac{\left(\beta - \theta\right)}{1 - \delta\theta}\right)^2 \frac{\bar{\sigma}^2}{1 - \theta^2}}{var\left(pd_t^C\right)} \tag{G.41}$$

G.4   Results

Table A1 reports an alternative version of table 1 in which we use the more sophisticated model of dividends that are cointegrated with consumption growth. Since the consumption process is unchanged, there is no effect on the worst-case model of consumption. The only difference between table A1 and table 1 is that they use different models of dividends and hence have different implications for equity returns.

The mean and standard deviation of returns are both slightly reduced – the mean is lower by 5 basis points and the standard deviation by 14 basis points. The small reduction is due to the fact that $g_c\left(\delta\right) = 0.993$. The difference between the returns under the two models of dividends depends purely on that term being different from 1. The fact that it is not (which is a consequence of cointegration) is why the returns are essentially unchanged. The autocorrelation and standard deviation of the price/dividend ratio are also numerically nearly identical to what is obtained in table 1. Finally, the bottom two rows of table A1 confirm that the model is calibrated here so that the standard deviation of dividend growth and the

correlation between dividend growth and consumption growth is identical to the data (the data moments are drawn from Bansal and Yaron (2004), as is the case with our other empirical targets).

Table A1: Asset pricing moments for the white-noise benchmark – extended dividend model

Fundamental parameters   *Implied worst-case*

$\overline{\sigma}$	Cons. vol. point est.	0.01465	σ_w	0.01470
$\overline{b}(\beta)$	Long-run vol. point est.	1	$b^w(\beta)$	2.449
$\overline{\mu}$	Mean cons. growth	0.0045	μ_w	0.0073
β	Time discount	0.997	θ	0.99021
λ	Ambiguity aversion	106.8	*Standard Epstein–Zin / robust-control*	
α	RRA (implied by λ)	4.73	b_0	0.01465
γ	Leverage	4.806	$b(\beta)$	1

Asset pricing moments (annualized)

	Model	Data
std(M)	0.30	N/A
E[R-Rf]	6.28	6.33
std(r)	19.27	19.42
AC1(PD)	0.95	0.81
std(P/D)	0.19	0.29
std(Δd)	11.49	11.49
corr(Δd,Δc)	0.55	0.55

Notes: see table 1. Results for the case where dividends are cointegrated with consumption but have transitory error. The standard deviation of dividends and correlation with consumption growth are taken from Bansal and Yaron (2004).

Table A2. Probability of rejecting the pricing model – extended results

Rejection probs. (5% critical value, H0=worst-case model)

	50 years	100 years	250 years	500 years	1000 years
Ljung–Box	5.1%	5.1%	5.3%	5.7%	6.6%
ARMA(1,1)	3.7%	4.4%	9.0%	27.3%	77.9%
ARMA(1,1), fixed mean	5.4%	6.8%	13.3%	35.0%	82.3%
Newey–West(5)	6.0%	8.4%	24.9%	58.7%	89.7%
Newey–West(10)	5.9%	8.6%	23.2%	48.3%	75.6%
Newey–West(20)	5.8%	8.8%	23.5%	48.6%	73.4%

Notes: Rejection probabilities are obtained by simulating the distributions of the three statistics in 50- and 100-year simulations of the cases where consumption growth is generated by the worst-case and white-noise models and asking how often the statistics in the latter simulation are outside the 95% range in the former simulation. The numbers in parentheses in the Newey–West rows are lag orders.

Table A3: Asset pricing moments for the white-noise benchmark – high discounting

Fundamental parameters			*Implied worst-case*	
$\overline{\sigma}$	Cons. vol. point est.	0.01465	σ_w	0.01472
$\overline{b}(\beta)$	Long-run vol. point est.	1	$b^w(\beta)$	1.353
$\overline{\mu}$	Mean cons. growth	0.0045	μ_w	0.0033
β	Time discount	0.987	θ	0.97821
λ	Ambiguity aversion	13.88	*Standard Epstein–Zin / robust-control*	
α	RRA (implied by λ)	6.65	b_0	0.01465
γ	Leverage	4.806	$b(\beta)$	1

Asset pricing moments (annualized)

	Model	Standard EZ	Data
std(M)	0.25	0.19	N/A
E[R-Rf]	6.33	2.75	6.33
std(r)	18.10	14.08	19.42
E[rf]	5.94	6.40	0.86
std(rf)	0.26	0	0.97
AC1(PD)	0.92	N/A	0.81
std(P/D)	0.10	0	0.29
$E[y_{10}-r_f]$	-7.8bp	0	N/A
EIS estimate	0	N/A	0.14

Notes: see table 1. This table uses a higher rate of time preference – 5 percent per year. λ is then reduced to a value low enough to match the equity premium, which also implies a higher value of risk aversion.