Long-Run Risk is the Worst-Case Scenario

Rhys Bidder and Ian Dew-Becker*

November 13, 2014

Abstract

We study an investor who is unsure of the dynamics of consumption growth. She estimates her consumption process non-parametrically to place minimal restrictions on dynamics. Treating the agent as ambiguity-averse yields a simple and tractable framework for analyzing her model uncertainty. We analytically show that the worst-case model that she uses for pricing, given a penalty on deviations from the point estimate, is a model with long-run risks, even if consumption growth is actually white noise. With a single parameter determining risk preferences, the model generates high and volatile risk premia and matches R^2s from return forecasting regressions, even though risk aversion is equal to 4.8 and the worst-case dynamics are statistically nearly indistinguishable from the true model.

*Bidder: Federal Reserve Bank of San Francisco. Dew-Becker: Northwestern University. We appreciate helpful comments from Harjoat Bhamra, Stefano Giglio, Costis Skiadas, and seminar participants at the Federal Reserve Bank of San Francisco, Duke Fuqua, the Asset Pricing Retreat in Tilburg, Kellogg, and UC Santa Cruz.

1 Introduction

Economists do not agree on the dynamic properties of the economy. There has been a long-running debate in the finance literature over how risky consumption growth is in the long-run (e.g. Bansal et al. (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 consumer is also unsure of the true model driving the world. This paper studies the behavior of such an agent.

We analyze an investor who considers a set of models of the economy that is only weakly constrained. Rather than just being uncertain about the parameters in a specific model, or putting positive probability on a handful of models, our agent considers an infinite-dimensional space of stationary autoregressive moving average (ARMA) models. Our goal is to understand the behavior of this agent, in particular how she prices financial assets.

Technically, the question we ask is, for an agent who has Epstein–Zin (1991) preferences, what model of consumption growth yields the lowest lifetime utility, subject to a constraint on the plausibility of the model? Ambiguity aversion in our setting is expressed by the investor pricing assets using such a pessimistic model that yields low lifetime utility. The headline 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 simply adds a long-run risk component to it. 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. 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.

The long-run risk model – a combination of Epstein–Zin (1991) preferences with a consumption process featuring persistent trend shocks – is a leading explanation for a wide range of asset pricing puzzles. We claim it is perfectly defensible for investors to price assets as though the long-run risk model holds, even if their point estimate for consumption dynamics is the random walk that has also widely been used as a benchmark (e.g. Campbell and Cochrane, 1999). 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 endogenous worst-case model that we derive features trend shocks with a half life of *70* years.

But we obtain much more general results that should be of interest to researchers outside finance who have never even heard of the long-run risk model. All of our results are in closed form, and the analytic tools used to derive them likely have wider applicability. Our first result is to derive formulas for exactly what the worst-case model is that a person might face, given any point estimate for consumption dynamics, not just in the white-noise case. Regardless of the point estimate for consumption growth dynamics, the worst-case model adds a strong low-frequency component. The statement in the title of this paper is thus not an approximation; we derive it directly. Our result is similar to that of Bidder and Smith (2013), who also develop a model where the worst-case process for consumption growth features persistence not present in the true process.

Second, our results give a clear separation between an investor's aversion to cycles in consumption growth at each frequency and her uncertainty about the power of the consumption process at that frequency. Intuitively, the worst-case model places highest power relative to the point estimate on those frequencies with the highest uncertainty and those that people are most averse to. Similar to Dew-Becker and Giglio (2013), who study how the pricing kernel depends on cycles in consumption growth of different frequencies, we find that the agent's utility depends primarily on the power in consumption growth at the very lowest frequencies. In calibrations, differences across frequencies in estimation uncertainty are quantitatively minor in comparison to differences across frequencies in the fears of the agent. So the reason that the long-run risk model appears endogenously is that it involves fluctuations at exactly the frequencies that people are most averse to. At the same time, the deviations of the long-run risk model from white noise affect consumption growth at only a narrow range of frequencies, which is what makes it difficult to reject statistically.

Finally, we derive formulas for interest rates and the prices and expected returns of levered consumption claims. In our main calibration, the model generates realistically large risk premia, a highly variable and persistent price/dividend ratio, R^2s in return forecasting regressions as large as we observe empirically, and an estimated elasticity of intertemporal substitution (EIS) from interest rate regressions of zero as measured in Campbell and Mankiw (1989), even though agents actually have a unit EIS. We also show that the price/dividend ratio implied by our model given historical data on consumption and dividends matches the observed price/dividend ratio for the market well from 1882 to the early 1990's, though it fails to capture the dynamics of the last 20 years. So by allowing for realistic uncertainty over simple consumption dynamics, we are able to match six major features of the data that the model would otherwise not replicate.

Critically, 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, just as in standard consumption-based models, and it corresponds to a coefficient of relative risk aversion of only 4.8. Furthermore, the worst-case model is essentially impossible to distinguish from the true model in a 100-year sample. A person who believes that the data is driven by the worst-case model will reject that hypothesis less than 7 percent of the time at the 5 percent level.

Our analysis of the worst-case model directly builds on a number of important areas of research. First, our focus on a single worst-case outcome is closely related to Gilboa and Schmeidler's (1989) work on ambiguity aversion in that they also study agents who focus on a single worst-case model, though in a static setting. Second, we build on the analysis of generalized recursive preferences to allow for the consideration of multiple models.[1] In that regard, we relate most closely to the recent work of Hansen and Sargent (2010) and Ju and Miao (2012), who analyze dynamic models of ambiguity aversion in the case of uncertainty about model parameters and hidden states.[2] The work of Hansen and Sargent (2010) is perhaps most comparable to ours, in that they study an agent who puts positive probability on both a white-noise and a long-run risk model for consumption growth. The key difference here, though, is that rather than imposing only two possible choices for dynamics, we explicitly consider the agent's estimation problem and allow her to put weight on any plausible model. The emergence of the long-run risk model as the one that she focuses on is entirely endogenous.

The restriction in past work to specific parametric models is important.[3] Limiting the agent to

[1] 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.

[2] There is also a large recent literature in finance that specializes models of ambiguity aversion to answer particularly interesting economic questions, such as Liu et al. (2004) and Drechsler's (2010) 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. There are also recent papers that ambiguity aversion asset pricing with learning, including Veronesi (2000), Brennan and Xia (2001), Epstein and Schneider (2007), Cogley and Sargent (2008), Leippold et al. (2008), Ju and Miao (2012) , and Collin-Dufresne et al. (2013).

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

only consider specific ARMA specifications can easily eliminate the true worst-case model from the space of models the agent can consider.[4] As an example, we show that if the agent is convinced that consumption growth is an AR(2) but she is unsure of the lag coefficients, the worst-case AR(2) model is substantially different from the worst-case unconstrained ARMA model. Intuitively, an AR(2) is not sufficiently flexible to generate the type of models with power at only very low frequency that our non-parametric agent fears.

Finally, we observe that because the worst-case model is more persistent than the agent's point estimate, pricing behavior is similar to the extrapolation implied by the "natural expectations" studied by Fuster et al. (2011): investors behave as though the endowment has a small highly persistent component, which leads them, from the perspective of an econometrician, to excessively extrapolate based on recent observations. In that regard, our paper is related to the literature on belief distortions and extrapolative expectations.[5]

To summarize, then, our contribution is to show that when investors have a very general form of model uncertainty and are ambiguity averse, they will tend to focus on models featuring long-run risk. We confirm that that result holds in both an endowment economy and when consumption is chosen endogenously. Our results show that in work on model uncertainty, it is important to allow people to consider models with highly persistent trends. More generally, we provide a framework for linking ambiguity aversion with non-parametric estimation, which we view as a realistic description of how people might think about the models they estimate. While investors may estimate finite-order ARMA models of the economy, they likely understand that those models almost certainly are misspecified. So if they want to be prepared for a worst-case scenario, they need to consider very general deviations from their point estimate. We provide a way to analyze those types of deviations.

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. Finally, section 7 analyzes the quantitative implications

(2013).

[4] Hansen and Sargent (2010) allow the agent to consider both a white-noise process and an ARMA(1,1), exactly as we would suggest (though our results would imply that the ARMA(1,1) should have been more persistent than they calibrated it to be).

[5] See Barsky and De Long (1993), Cecchetti et al. (1997), Fuster et al. (2011) and Hirshleifer and Yu (2012)

of the model and explores robustness tests. Two key extensions of the analysis that we provide are to show that the results are unchanged with endogenous consumption and that when the investor is allowed to estimate volatility dynamics the worst-case model becomes the heteroskedastic version of the long-run risk model. Section 8 concludes.

2 Non-parametric estimation

We model investors who are uncertain about the true process driving the economy. The first step, then, is to describe the set of possible models that investors consider and how they measure their plausibility.

2.1 Economic environment

We study a pure endowment economy. Investors form expectations for future consumption growth using models of the form

$$\Delta c_t = \mu + a(L)(\Delta c_{t-1} - \mu) + b_0 \varepsilon_t \tag{1}$$

$$\varepsilon_t \sim N(0,1) \tag{2}$$

where μ is mean consumption growth, $a(\cdot)$ is a polynomial function, L is the lag operator, 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 agents use linear models for forecasting, even if consumption dynamics are not truly linear, given that the economics literature focuses almost exclusively on linear models. Moreover, the Wold theorem states that any covariance stationary process can be represented in the form (1), or the associated moving average representation, with uncorrelated, though not necessarily independent Gaussian innovations ε_t. For our purposes, the restriction to the class of linear processes is important as a description of the agent's modeling method, rather than as a restriction on the true process driving consumption growth.[6] The as-

[6] Allowing the agent to model heteroskedasticity in consumption growth would be an interesting extension to our results, but we do not analyze it here.

sumption that ε_t is normally distributed is not necessary, but it simplifies the exposition somewhat. The appendix completes the derivation of the model for the case when ε_t is only restricted to have zero mean and be serially independent.

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

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

$$b(L) \equiv (1 - La(L))^{-1} b_0 \tag{4}$$

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

We can thus express the dynamics of consumption equivalently as depending on $\{a, b_0\}$ or just on the polynomial b. 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 $\{a, b_0\}$ is always associated with a distinct value of b and vice versa.

There are no latent state variables. When a model $a(L)$ has infinite order we assume that the agent knows all the necessary lagged values of consumption growth for forecasting (or has dogmatic beliefs about them) so that no filtering is required.[7] We discuss necessary constraints on the models below. For now simply assume that they are sufficiently constrained that any quantities we must derive exist.

We assume that the investor knows the value of μ with certainty but is uncertain about consumption dynamics. In addition to parsimony, there are two justifications for that assumption. First, for the estimation method we model the agent as using, estimates of the coefficients $\{a, b_0\}$ converge at an asymptotically slower rate than estimates of μ. Second, we will model the agent as having Epstein–Zin (1991) preferences, which, as shown by Barillas, Hansen, and Sargent (2009), can be viewed as appearing when an agent with power utility is unsure of the distribution of the innovations of consumption growth. So uncertainty about the mean will implicitly be accounted

[7] Croce, Lettau, and Ludvigson (2014) study in detail issues surrounding filtering and the type of information available to agents in long-run risk models. Our agent has less information than in Bansal and Yaron (2004), in some sense, because she can only observe past consumption growth and no other state variables. Conditional on that a model for consumption growth, though, our agent filters optimally. Croce, Lettau, and Ludvigson (2014) suggest optimal filtering may in fact be rather difficult.

for through the preferences.

2.2 Estimation

The agent in our model chooses among specifications for consumption growth, b, partly based on their statistical plausibility. That plausibility is measured by a function $g(b)$, where high values of $g(b)$ are associated with less plausible models. As a simple example, if a person were to estimate a parameterized model, such as an AR(1), on a sample of data, she would have a likelihood (or posterior distribution) over the autocorrelation, and she could rank different AR(1) processes by how far their autocorrelations are from her point estimate. That example, though, imposes a specific parametric specification of consumption growth and rules out all other possible models. In the spirit of modeling the agent as looking for decision rules that are robust to a broad class of potential models, we assume that she estimates the dynamic process driving consumption non-parametrically so as to make only minimal assumptions about the driving process.

Following Berk (1974) and Brockwell and Davis (1988), we assume that the investor estimates a finite-order AR or MA model for consumption growth, but that she does not actually believe that consumption growth necessarily follows a finite-order specification. Instead, it may be driven by an infinite-order model, and her finite-order model is simply an approximation. In terms of asymptotic econometric theory, the way that she expresses her statistical doubts is to imagine that if she were given more data, she would estimate a richer model. That is, the number of lags in her AR or MA model grows asymptotically with the sample size, potentially allowing eventually for a broad class of dynamics.[8] The agent then has a non-parametric confidence set around any particular point estimate that implicitly includes models far more complex than the actual AR or MA model she estimates in any particular sample.

Our analysis of the model takes place in the frequency domain because it will allow us to obtain a tractable and interpretable solution, but it only requires defining a single object. The *transfer*

[8] We use the results from Berk (1974) and Brockwell and Davis (1988) because their method of using models with growing lag lengths delivers distributions for the lag polynomial $b(L)$. The results for other non-parametric spectral estimators are similar to theirs in the sense that the standard deviation of the spectrum is proportional to the level of the spectrum itself, and errors are independent across frequencies (see, e.g., Priestly (1981) and Hashimzade and Vogelsang (2008)). The problem with using other results on the distribution of the spectrum is that factoring the spectrum to obtain the lag polynomial is a complicated and non-analytic process (Priestly (1981)).

function for a model is defined as

$$B(\omega) \equiv b\left(e^{i\omega}\right) \tag{6}$$

for $i = \sqrt{-1}$. The transfer function measures how the filter $b(L)$ transfers power at each frequency, ω, from the white-noise innovations, ε, to consumption growth. $B(\omega)$ and $b(L)$ are a Fourier transform pair. Berk (1974) and Brockwell and Davis (1988) show that under standard conditions, estimates of the transfer function, $\hat{B}(\omega)$, have the following asymptotic distribution[9]

$$\hat{B}(\omega)T^q \sim CN\left(B^{True}(\omega), f^{True}(\omega)\right) \tag{7}$$

$$cov\left(\hat{B}(\omega_1) - B^{True}(\omega_1), \hat{B}(\omega_2) - B^{True}(\omega_2)\right) = 0 \text{ for } \omega_1 \neq \omega_2 \tag{8}$$

$$\text{where } f^{True}(\omega) = \left|B^{True}(\omega)\right|^2 \tag{9}$$

where $B^{True}(\omega)$ is the true dynamic model, CN is the complex normal distribution, T is the length of the sample used for estimation, and q is a positive constant (different values of q correspond to different assumptions about the growth rate of the order of the estimated AR model).[10] $f^{True}(\omega)$, which determines the variance of the estimates, is the spectral density of consumption growth, which measures the variance of the fluctuations in consumption growth with frequency ω.[11]

Now suppose the agent has a point estimate for consumption dynamics of $\bar{B}(\omega)$ (with associated spectral density $\bar{f}(\omega) = \left|\bar{B}(\omega)\right|^2$). Given the independent normal distribution for the transfer function at each frequency, a natural test statistic for a particular model is

$$g(b) = \int \frac{\left|B(\omega) - \bar{B}(\omega)\right|^2}{\bar{f}(\omega)} d\omega \tag{10}$$

[9] The key condition on the dynamic process for consumption growth is that the spectral density, $|B(\omega)|^2$ is finite and bounded away from zero. Alternatively, one may assume that the MA coefficients $\{b_j\}$ are absolutely summable. See Stock (1994) for a discussion of the relationship between such conditions. That condition simply eliminates some pathological processes and also fractional integration. The innovations to consumption growth must have a finite fourth moment, and there are some technical conditions on the growth rate of the lag order of the model that must be satisfied. Bhansali (1997) derives a closely related consistency result (but without a distribution theory) under substantially weaker conditions on the distribution of the innovations.

The summability condition is also needed for the time-domain analysis discussed below.

[10] $q \leq 1/3$. For estimation of the mean of consumption growth, the usual rate would correspond to $q = 1/2$, so the estimate of consumption dynamics converges at an asymptotically slower rate than the estimate of the mean.

[11] The original results from Berk (1974) and Brockwell and Davis (1988) include an extra restriction that we do not impose here. The asymptotics imply that the variance of the innovations, b_0^2, is estimated at a faster asymptotic rate than the other lag coefficients. Were we to impose that part of the result, we would add an extra constraint in the optimization problem that $b_0^w = \bar{b}_0$. The results are essentially unaffected by this constraint.

(Here and below, integrals without limits denote $(2\pi)^{-1}\int_{-\pi}^{\pi}$). $g(b)$ is a χ^2-type test statistic for the null hypothesis that $B = \bar{B}$.[12,13] We are modeling the agent's beliefs about potential models by assuming that she compares possible models to a point estimate $\bar{B}$, which may or may not be the true model, using essentially a Wald test based on the non-parametric asymptotics of Berk (1974) and Brockwell and Davis (1988). Models that deviate from the point estimate by a larger amount on average across frequencies are viewed as less plausible.[14]

2.3 Time-domain interpretation

While not strictly necessary to the analysis below, we now briefly discuss how $g(b)$ can be derived as a limiting case of a standard Wald statistic. Suppose the agent were to calculate a simple Wald test of the first m coefficients of a model b, given the point estimate $\bar{b}$. Denote the row vector of those coefficients as $\mathbf{b}_{1:m}$ and the variance matrix for $\mathbf{b}_{1:m}$ as Σ_m. The Wald statistic is then

$$m^{-1}\left(\mathbf{b}_{1:m} - \bar{\mathbf{b}}_{1:m}\right)\bar{\Sigma}_m\left(\mathbf{b}_{1:m} - \bar{\mathbf{b}}_{1:m}\right)' \tag{12}$$

[12] Hong (1996) studies a closely related distance metric in the context of testing for general deviations from a benchmark spectral density.

[13] Note that the penalty function can be calculated for values of b that are ruled out under the asymptotic assumptions underlying the estimation process. Most interestingly, $g(b)$ is well defined for certain fractionally integrated processes. The agent therefore is allowing for fractional integration in the models she considers, so if (as will be the case) we ultimately find that the worst-case does not involve fractional integration, that is an equilibrium result, rather than an assumption.

[14] One potentially puzzling aspect of the results used here is that it seems as though there is an equal amount of uncertainty at all frequencies. In particular, suppose consumption growth is serially independent so that $f(\omega)$ is flat. Then our results would imply that there is an exactly equal amount of uncertainty about consumption dynamics at all frequencies. But intuition suggests rather that there is more uncertainty at low frequencies. To see how that intuition is actually consistent with our results, note that the wavelength associated with a particular frequency is $2\pi/\omega$. So wavelengths are compressed more closely near frequency zero. To decompress the wavelengths and put them on an equal footing with frequencies, we can consider an integral over log frequencies, noting that the log wavelength is equal to $\log 2\pi - \log\omega$, so integrating over log frequencies is identical to integrating over log wavelengths. It is straightforward to show that $g(b)$ can be rewritten in terms of log frequencies, $\xi = \log\omega$ as

$$\frac{1}{2\pi}\int_0^{2\pi}\frac{\left|B(\omega)-\bar{B}(\omega)\right|^2}{\bar{f}(\omega)}d\omega = \frac{1}{2\pi}\int_{-\infty}^{\log(2\pi)}\frac{\left|B(\exp(\xi))-\bar{B}(\exp(\xi))\right|^2}{\bar{f}(\exp(\xi))/\exp(\xi)}d\xi \tag{11}$$

(the change of limits from $[-\pi,\pi]$ to $[0,2\pi]$ allows for the change of measure but has no effect on the results since the functions all have period 2π). That is, when $g(b)$ is written as an integral over log frequencies, the relevant measure of uncertainty becomes $\bar{f}(\exp(\xi))/\exp(\xi)$, which grows to ∞ at low frequencies. A similar result is obtained if the integral is written in terms of wavelengths. To the extent that economic intuition is stronger with wavelengths, this scaling may be more natural.

Using the results from Berk (1974) and Brockwell and Davis (1988), the appendix shows that the Wald statistic converges, as $m, T \to \infty$ (with $m = o\left(T^{1/3}\right)$), to the statistic,

$$m^{-1}\left(\mathbf{b}_{1:m} - \bar{\mathbf{b}}_{1:m}\right)\bar{\Sigma}_m\left(\mathbf{b}_{1:m} - \bar{\mathbf{b}}_{1:m}\right)' \to m^{-1}\left(\mathbf{B}_m - \bar{\mathbf{B}}_m\right)\bar{F}_m^{-1}\left(\mathbf{B}_m - \bar{\mathbf{B}}_m\right)^* \tag{13}$$

$$\to \int \frac{\left|B(\omega) - \bar{B}(\omega)\right|^2}{\bar{f}(\omega)} d\omega = g(b) \tag{14}$$

where * denotes a transposed complex conjugate, $\mathbf{B}_m$ is the discrete Fourier transform vector of $\mathbf{b}_{1:m}$ and $\bar{F}_m^{-1}$ is the inverse of the diagonal matrix of the spectral density, $\bar{f}(\omega)$, evaluated at the Fourier frequencies under the model $\bar{b}$.[15]

We can therefore interpret the distance measure $g(b)$ as being the limit of a Wald statistic for the MA coefficients of the hypothetical model b. This result justifies the equal weighting across frequencies implicit in $g(b)$. While solving our model is possible without using the Fourier transform step (i.e. simply defining $g(b) = \left(\mathbf{b} - \bar{\mathbf{b}}\right)\bar{\Sigma}^{-1}\left(\mathbf{b} - \bar{\mathbf{b}}\right)'$), we will see below that in that case the results become difficult or impossible to interpret.[16]

3 Preferences

3.1 Utility for dogmatic beliefs

Investors face two sources of uncertainty: the innovations and the propagation mechanism. If the investor is convinced without any doubt that consumption is driven by a particular model $\{a, b_0\}$, her utility over the innovations ε_t is described by Epstein–Zin (1991) preferences. The reason that we use Epstein–Zin preferences is that in order for the analysis to be non-trivial, it is important that the investor's utility actually be affected by consumption dynamics. Under power utility, autocorrelations have no effect on average lifetime utility, whereas under generalized recursive preferences, we will see that they do.

The investor's coefficient of relative risk aversion is α, her time discount parameter is β, and her elasticity of intertemporal substitution (EIS) is equal to 1. We focus on the case of a unit EIS

[15] That is, the j, j element of $\bar{F}_m$ is equal to $\bar{f}(2\pi(j-1)/m)$ and the off-diagonal elements are all equal to zero.

[16] $g(b)$ has three equivalent expressions. The integral in (10); the infinite vector/matrix form, $\left(\mathbf{B} - \bar{\mathbf{B}}\right)\bar{F}^{-1}\left(\mathbf{B} - \bar{\mathbf{B}}\right)^*$, where $\mathbf{B}$, $\bar{\mathbf{B}}$, and $\bar{F}$ represent limits of $\mathbf{B}_m$, $\bar{\mathbf{B}}_m$, and $\bar{F}_m$ as $m \to \infty$; and the pure time-domain form, $\left(\mathbf{b} - \bar{\mathbf{b}}\right)\bar{\Sigma}^{-1}\left(\mathbf{b} - \bar{\mathbf{b}}\right)'$, where $\bar{\Sigma}$ is definted in the appendix and $\mathbf{b}$ and $\bar{\mathbf{b}}$ are infinitely long vectors of the coefficients in the models b and $\bar{b}$.

to ensure that we can easily derive analytic results.[17]

Lifetime utility, v, for a fixed model $\{a, b_0\}$ is

$$v\left(\Delta c^t; a, b_0\right) = (1-\beta)\, c_t + \frac{\beta}{1-\alpha} \log E_t \left[\exp\left(v\left(\Delta c^{t+1}; a, b_0\right)(1-\alpha)\right) | a, b_0\right] \tag{15}$$

where $E_t\left[\cdot | a, b_0\right]$ 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 $\{a, b_0\}$.[18]

It is convenient to rescale lifetime utility by current consumption. The appendix shows that solution to the recursion for lifetime utility is

$$vc\left(\Delta c^t; a, b_0\right) \equiv v\left(\Delta c^t; a, b_0\right) - c_t = \frac{\beta}{1-\alpha} \log E_t \left[\exp\left(\left(vc\left(\Delta c^{t+1}; a, b_0\right) + \Delta c_{t+1}\right)(1-\alpha)\right) | a, b_0\right] \tag{16}$$

$$= \sum_{k=1}^{\infty} \beta^k E_t\left[\Delta c_{t+k} | a, b_0\right] + \frac{\beta}{1-\beta}\frac{1-\alpha}{2} b(\beta)^2 \tag{17}$$

The summation term is what would be obtained if the investor had time-separable log utility, in which case consumption dynamics would not affect lifetime utility on average. The second term is an adjustment 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 $b(\beta)^2$, which measures the variance of the shocks to lifetime utility in each period. $b(\beta)$ 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 $b(\beta)$ that causes people with Epstein–Zin preferences to be averse to long-run risk. So $b(\beta)$ will feature prominently in determining what models an investor would fear most.

3.2 Robustness over dynamics

Equation (16) gives lifetime utility in the case where our investor dogmatically believes in a particular model. We now model her as considering alternative models of propagation.

The investor entertains a set of possible values for the lag polynomial and can associate with any model a measure of its plausibility. Seeking robustness, the agent makes her choices as though consumption growth is driven by worst-case dynamics, denoted b^w. These dynamics are not the

[17] 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

[18] For example, we have $E_t\left[\Delta c_{t+1} | a, b_0\right] = a(L)\,\Delta c_t$.

worst in an unrestricted sense but, rather, are the worst among statistically plausible models. This discipline is captured by the penalty function $g(b)$, derived above, that expresses how the investor downweights a given model according to its implausibility.

The worst-case model is then obtained as the solution to a penalized minimization problem:

$$b^w = \arg\min_b \left\{ E\left[vc\left(\Delta c^t; b\right) |b\right] + \lambda g(b) \right\} \tag{18}$$

b^w is the model that gives the agent the lowest unconditional expected lifetime utility, subject to the penalty $g(b)$. λ is a parameter that determines how much weight the penalty receives. As usual, λ can either be interpreted directly as a parameter to be calibrated or as a Lagrange multiplier on a constraint on the Wald-type statistic $g(b)$.

Our ambiguity-averse agent's utility then takes the form of that of an Epstein-Zin agent but using b^w to form expectations about future consumption growth.

$$vc^w\left(\Delta c^t\right) = vc\left(\Delta c^t; b^w\right) = \frac{\beta}{1-\beta}\frac{1-\alpha}{2} b^w(\beta)^2 + \sum_{k=1}^{\infty} \beta^k E_t\left[\Delta c_{t+k} | b^w\right] \tag{19}$$

That is, lifetime utility is calculated using b^w to form expectations about future consumption growth.[19]

In modeling investors as choosing a single worst-case b^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 b^w is timeless – it is invariant to the time-series evolution of consumption – 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, b^w would be that model. Unlike in some related recent papers, the investor in this model does not change her probability weights every day. She chooses a single pessimistic model to protect against.

A natural question is why we analyze a worst case instead of allowing the agent to average as

[19] Note that since utility is recursive, the agent's preferences are time-consistent, just 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.

a Bayesian across all the possible models. A simple and reasonable answer is that people may not actually be Bayesians, or they may not be able to assign priors to all models.[20] More practically, obtaining analytic solutions in a purely Bayesian model with a non-degenerate distribution over the dynamic process for consumption growth is likely impossible: the distribution of future consumption growth in that case is the product of the distributions for b and ε, which does not take a tractable form.[21,22] The analysis of the worst-case model gives us a tractable view into an agent's decision problem that would otherwise not be available.

4 Deriving and interpreting the worst case

The analysis above leads us to a simple quadratic optimization problem. The worst-case lag polynomial is obtained from (18) (ignoring a term involving mean consumption growth, which the agent knows)

$$b^w(L) = \arg\min_{b(L)} \left\{ \frac{\beta}{1-\beta} \frac{1-\alpha}{2} b(\beta)^2 + \lambda \int \frac{\left| B(\omega) - \bar{B}(\omega) \right|^2}{\bar{f}(\omega)} d\omega \right\} \tag{20}$$

Equivalently, we can write

$$\begin{aligned} b^w(L) &= \arg\min_{b(L)} \frac{\beta}{1-\beta} \frac{1-\alpha}{2} \left(\int Z(\omega) B(\omega) d\omega \right)^2 + \lambda \int \frac{\left| B(\omega) - \bar{B}(\omega) \right|^2}{\bar{f}(\omega)} d\omega & (21) \\ \text{where } Z(\omega) &\equiv \sum_{j=0}^{\infty} \beta^j e^{-i\omega j} & (22) \end{aligned}$$

The fact that $b(\beta) = \int Z(\omega) B(\omega) d\omega$ is a simple application of Parseval's theorem. $Z(\omega)$ measures how much weight the lifetime utility function places on frequency ω while $B(\omega)$ captures how much weight the model for propagation puts on that frequency.

Because $B(\omega)$ is a causal filter, only the Fourier coefficients on the positive side of the origin have non-zero values. Dew-Becker and Giglio (2013) use that fact to show that the integrals above

[20] See Machina and Siniscalchi (2014) for a recent review of the experimental literature on ambiguity aversion.

[21] It is also worth noting that it would be impossible to obtain numerical solutions when the distribution is infinite-dimensional, as it is here. By assuming that the agent behaves as if she places probability 1 on a single model, we avoid the problem of having to integrate over the infinite-dimensional distribution of possible models when forming expectations.

[22] Kreps (1998) discusses related issues in models with learning and argues for an anticipated utility modeling strategy, which, similarly to our setup, models agents as making decisions assuming a specific model of the world holds, even though in reality they understand that their current beliefs may be wrong and likely to change. The key difference between this paper and Kreps (1998) is that the model our agent uses for pricing is a worst-case model rather than the current point estimate.

can generally be written in terms of real-valued functions. Specifically, define

$$Z_r(\omega) \equiv 1 + 2\sum_{j=1}^{\infty} \beta^j \cos(\omega j) \tag{23}$$

$$B_r(\omega) \equiv \text{real}(B(\omega)) \tag{24}$$

We then have

$$b(\beta) = \int Z(\omega) B(\omega) d\omega = \int Z_r(\omega) B_r(\omega) d\omega \tag{25}$$

Figure 1 plots $Z_r(\omega)$ 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. In other words, the agent's fears are primarily about very low-frequency fluctuations in consumption growth.

(21) is a quadratic optimization, so it has a unique minimum. The appendix shows that the solution, in terms of the worst-case transfer function, is

$$B^w(\omega) = \bar{B}(\omega) + \lambda^{-1}\frac{\beta}{1-\beta}\frac{\alpha-1}{2} b^w(\beta) \times \bar{f}(\omega) \times Z^*(\omega) \tag{26}$$

where a * denotes a complex conjugate and $b^w(\beta)$ is given by

$$b^w(\beta) = \frac{\bar{b}(\beta)}{1 - \lambda^{-1}\frac{\beta}{1-\beta}\frac{\alpha-1}{2}\int Z(\omega)^* Z(\omega)\bar{f}(\omega) d\omega} \tag{27}$$

We thus have a closed form expression for the worst-case model, described in the frequency domain.[23] The worst-case transfer function $B^w(\omega)$ in (26) is equal to the true transfer function plus a term that depends on three factors. First, $\lambda^{-1}\frac{\beta}{1-\beta}\frac{\alpha-1}{2}b^w(\beta)$ represents the ratio of the utility losses from a marginal increase in $b^w(\beta)$ to the cost of normalized deviations from the point estimate, λ. Second, $\bar{f}(\omega)$ represents the amount of uncertainty the agent has about consumption dynamics at frequency ω. Where $\bar{f}(\omega)$ is high there is relatively more uncertainty. Finally, $Z(\omega)$ determines how much weight the lifetime utility function places on frequency ω. Since the mass of $Z(\omega)$ lies very close to frequency 0, the worst case tends to shift power to very low frequencies. In

[23] In the case where ε_t is not normally distributed, the solution is determined by the condition $B^w(\omega) = \bar{B}(\omega) - \frac{\lambda^{-1}}{2}\frac{\beta}{1-\beta}\Gamma'(b^w(\beta)(1-\alpha))$, where Γ' is the derivative of the cumulant-generating function of ε.

the limit where $\beta \to 1$, $Z(\omega)$ becomes a point mass at zero.[24,25]

The worst-case lag polynomial $b^w(L)$ is obtained as the inverse Fourier transform of $B^w(\omega)$,

$$b_j^w = \int \exp(-i\omega j) B^w(\omega) d\omega \tag{30}$$

Equation (26) represents the completion of the solution to the model. To summarize, given a point estimate for dynamics, $\bar{B}$ (estimated from a finite-order model that we need not specify here), the agent selects a worst-case model $B^w(\omega)$, which is associated with a specific $b^w(L)$ through equation (30). She then uses the worst-case model when calculating expectations and pricing assets.

4.1 Long-run risk is the worst case scenario

Suppose the agent's point estimate is that consumption growth is white noise, with

$$\bar{b}(L) = \bar{b}_0 \tag{31}$$

Using the formula for $B^w(\omega)$ from (26) and the inverse Fourier transform, we obtain

$$b_j^w = \varphi\beta^j \text{ for } j > 0 \tag{32}$$

$$b_0^w = \bar{b}_0 + \varphi \tag{33}$$

$$\varphi \equiv \frac{\beta}{1-\beta}\frac{\alpha-1}{2} b^w(\beta)\bar{b}_0^2\lambda^{-1} \tag{34}$$

The MA process in (32-34) has an equivalent state-space representation

$$\Delta c_t = \mu + x_{t-1} + \eta_t \tag{35}$$

$$x_t = \beta x_{t-1} + v_t \tag{36}$$

[24] See Dew-Becker and Giglio (2013), for a thorough discussion of $Z(\omega)$ and related functions

[25] The equivalent time-domain version of the analysis is

$$\mathbf{b}^w = \arg\min_b \left\{ -\beta(1-\alpha)\mathbf{b}'zz'\mathbf{b} + \lambda\left(\mathbf{b}-\bar{\mathbf{b}}\right)\bar{\Sigma}^{-1}\left(\mathbf{b}-\bar{\mathbf{b}}\right)' \right\} \tag{28}$$

where $z \equiv \left[1, \beta, \beta^2, ...\right]'$ and $\bar{\Sigma}$ is the asymptotic covariance from Brockwell and Davis (1988) (also defined in the appendix below). The solution is

$$\left(\mathbf{b}^w - \bar{\mathbf{b}}\right)' = \lambda^{-1}\bar{\Sigma}\beta(1-\alpha) z b^w(\beta) \tag{29}$$

The interpretation of this result is far less clear than that of the spectral version due to $\bar{\Sigma}$. To move from this to the spectral result, we simply use the facts that $b = B\Lambda^*$ and $\bar{\Sigma} = \Lambda\bar{F}\Lambda^*$ to replace $\mathbf{b}^w$, $\bar{\mathbf{b}}$, and $\bar{\Sigma}$ with $\mathbf{B}^w$, $\bar{\mathbf{B}}$, and $\bar{F}$.

where

$$\eta_t \sim N\left(0, \theta\beta^{-1}\left(\bar{b}_0+\varphi\right)^2\right) \tag{37}$$

$$v_t \sim N\left(0, \frac{(1-\beta\theta)(\beta-\theta)}{\beta}\left(\bar{b}_0+\varphi\right)^2\right) \tag{38}$$

and η_t and v_t are independent. The state-space form in equations (35–38) is observationally equivalent to the original MA process in the sense that they have identical autocovariances, and it is exactly case I from Bansal and Yaron (2004), the homoskedastic long-run risk model. So when the agent's point estimate is that consumption growth is white noise, her worst-case model is literally the long-run risk model.[26]

The worst-case process exhibits a small but highly persistent trend component, where 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 at rate β spreads its effects as evenly as possible across future dates, scaled by their weight in utility. By spreading out the effects of the shock over time, the worst-case model limits its effects on the transfer function to a very narrow range of frequencies near zero, thus minimizing the penalty $g(b)$. Essentially, that is what the long-run risk model was designed to do originally: it generates large risk prices with only a minimal impact on the measurable dynamics of consumption growth. The worst-case model is 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.[27]

Figure 2 plots the real transfer function for the white-noise benchmark and the worst-case model. The transfer function for white noise is totally flat, while the worst case has substantial power at the very lowest frequencies, exactly as we would expect from figure 1.

[26] The result in this section can also be derived using the time-domain analysis. When the point estimate for consumption growth is white noise, the covariance of the MA coefficients is $\bar{\Sigma}=\bar{b}_0^2 I$. So we have $(\mathbf{b}^w-[1,0,0,...])'=\lambda^{-1}\bar{b}_0^2\beta(1-\alpha)zb^w(\beta)$ with $z=\left[1,\beta,\beta^2,...\right]'$, which yields the same solution as in (32-34).

[27] Bidder and Smith (2013) show that an agent who fears model misspecification (as captured by her having multiplier preferences) in a heteroskedastic environment will price assets as if there is autocorrelation in consumption growth even if there is none in the actual data generating process. Important in that story is the explicit presence of stochastic volatility and their finding that the agent's pessimism manifests itself especially strongly in periods of high volatility. In our current context we obtain a similar result, but without invoking stochastic volatility. Instead we exploit the agent's particular sensitivity to low frequency variation in consumption growth.

5 The behavior of asset prices

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

$$\begin{aligned} 1 &= E_t\left[R_{t+1}M_{t+1}|b^w\right] && (39) \\ \text{where } M_{t+1} &\equiv \beta\exp\left(-\Delta c_{t+1}\right)\frac{\exp\left(v\left(\Delta c^{t+1};b^w\right)\times(1-\alpha)\right)}{E_t\left[\exp\left(v\left(\Delta c^{t+1};b^w\right)\times(1-\alpha)\right)|b^w\right]} && (40) \end{aligned}$$

M_{t+1} is the stochastic discount factor.

It is straightforward, given that log consumption follows a linear Gaussian process, to derive 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. From the perspective of an econometrician who has the same point estimate for consumption dynamics as the agent, $\bar{b}$, the expected excess log return on the levered consumption claim is

$$\begin{aligned} E_t\left[r_{t+1}-r_{f,t+1}|\bar{a},\bar{b}_0\right] &= \frac{\gamma-\delta a^w(\delta)}{1-\delta a^w(\delta)}\left(\bar{a}(L)-a^w(L)\right)\left(\Delta c_t-\mu\right)+ \\ &\quad \underbrace{-\frac{\gamma-\delta a^w(\delta)}{1-\delta a^w(\delta)}b_0^w\left[b_0^w-(1-\alpha)\,b^w(\beta)\right]}_{cov^w(r_{t+1},m_{t+1})}-\underbrace{\frac{1}{2}\left(\frac{\gamma-\delta a^w(\delta)}{1-\delta a^w(\delta)}\right)^2\left(b_0^w\right)^2}_{\frac{1}{2}var_t^w(r_{t+1})} && (41) \end{aligned}$$

where δ is a linearization parameter (from the Campbell–Shiller approximation) that depends on the steady-state price/dividend ratio.[28]

The second line, which is equal to $-cov_t^w\left(r_{t+1},\log M_{t+1}\right)-\frac{1}{2}var_t^w\left(r_{t+1}\right)$ (i.e. a conditional covariance and variance measured under the worst-case dynamics), is the standard risk premium, and it is calculated now under the worst-case model, rather than the point estimate. 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 higher. That covariance, in turn, is higher for two reasons. First, since the agent believes that shocks to consumption growth are highly persistent, they have large effects on lifetime utility, thus making the SDF very volatile (the term $b_0^w-(1-\alpha)\,b^w(\beta)$). Second, again because of the persistence of consumption growth under the worst-case, shocks to

[28] See the appendix for a full derivation.

consumption have large effects on expected long-run dividend growth, so the return on the levered consumption claim is also very sensitive to shocks (through $\frac{\gamma-\delta a^w(\delta)}{1-\delta a^w(\delta)} b_0^w$). These two effects cause the consumption claim to strongly negatively covary with the SDF and generate a high risk premium.

The second difference between the risk premium in this model from a setting where the agent prices assets under the point estimate for consumption dynamics is the term $\frac{\gamma-\delta a^w(\delta)}{1-\delta a^w(\delta)} \left(\bar{a}\left(L\right) - a^w\left(L\right)\right)\left(\Delta c_t - \mu\right)$, which reflects the difference in forecasts of dividend growth between the point estimate, used by the econometrician, and the worst-case model used by investors. Since $\Delta c_t - \mu$ is zero on average, this term is also zero on average. But it induces predictability in returns. When the worst-case implies higher consumption growth, investors pay relatively more for equity, thus raising asset prices and lowering expected returns. So while we might expect the agent's fears of unfavorable dynamics would lead asset prices to be low, we see that sometimes they can actually lead to high prices. This channel leads to procyclical asset prices and countercyclical expected returns when $a^w\left(L\right)$ implies more persistent dynamics than $\bar{a}\left(L\right)$, similarly to Fuster et al. (2011).

It is also straightforward to solve for the risk-free rate,

$$r_{f,t+1} = -\log\beta + \mu + a^w\left(L\right)\left(\Delta c_t - \mu\right) - \frac{1}{2}\left(b_0^w\right)^2 + b_0^w b^w\left(\beta\right)\left(1-\alpha\right) \tag{42}$$

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 $\mu + a^w\left(L\right)\left(\Delta c_t - \mu\right)$, which is the expectation under the worst-case model.

6 Calibration

We now parameterize the model to analyze its quantitative implications. Most of our analysis will be 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 our parsimonious approach, we obtain striking empirical success in terms of matching important asset pricing moments. In addition we also briefly demonstrate that the essential elements of our results go through even if the agent's point estimate is not white noise and, in so doing, also emphasize the importance of not limiting the agent to consider only a restricted parametric class of processes.

A number 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. The steady-state dividend/price ratio used in the Campbell–Shiller approximation is 5 percent per year, as in Campbell and Vuolteenaho (2004), which implies $\delta = 0.95^{1/4}$. 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 4.63 to generate mean annualized equity returns of 6.33 percent.

The appendix shows that when α is interpreted as constraining a worst-case distribution of ε, as in Hansen and Sargent (2005) and Barillas, Hansen, and Sargent (2009), we can directly link it to λ through the formula

$$\alpha = 1 + \frac{1}{2\lambda\left(1-\beta\right)} \tag{43}$$

In Hansen and Sargent (2005) and Barillas, Hansen, and Sargent (2009), agents form expectations as though the innovation ε_t is drawn from a worst-case distribution. That distribution is chosen to minimize lifetime utility subject to a penalty on its distance from the benchmark of a standard normal, similarly to how we choose b^w here, and that distance depends on α. The coefficient of relative risk aversion in Epstein–Zin preferences, α, can therefore alternatively be interpreted as a penalty on a distance measure analogous to λ.

We calibrate λ to equal 52.24 to match the observed Sharpe ratio on equities. Formula (43) then implies α should equal 4.81. That level of risk aversion is extraordinarily small in the context of the consumption-based asset pricing literature with Gaussian innovations. It is only half the value used by Bansal and Yaron (2004), for example, who themselves are notable for using a relatively low value. α therefore immediately seems to take on a plausible value in its own right, separate from any connection it has to λ.

To further investigate how reasonable λ is, in the next section we study how easy it would be to reject the worst-case model given data generated under the true model in simulated samples and find that the worst-case model is extremely difficult to reject.

7 Quantitative implications

7.1 The white noise case

In this section, we return to the case where the agent's point estimate is that consumption growth is white noise, where we previously saw that the worst-case model is a long-run risk model. We now examine quantitative implications for asset prices.

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). However, $b^w(\beta)$ the relevant measure of the total risk in the economy, is 0.039, at the quarterly frequency, in both our model and theirs. The two models thus both have the same quantity of long-run risk, but in our case the long-run risk comes from a smaller but more persistent shock.

Note also that φ is very small, implying that b_0^w is only 2 percent larger than $\bar{b}_0$. So the variance of consumption growth under the worst-case model is essentially identical to under the benchmark. However, because the worst-case model is so persistent, $b^w(\beta)$ is 2.6 times higher than $\bar{b}(\beta)$, thus implying that the worst-case economy is far riskier than the point estimate.

7.1.1 Asset prices

Table 1 reports key asset pricing moments for the white-noise benchmark model. 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 is low and stable, with a mean and standard deviation of 1.89 and 0.33 percent, respectively.

The second column in the bottom section of table 1 shows that would happen if we set $\lambda = \infty$ but held α fixed at 4.81, 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 6.3 to 1.9 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 growth has far smaller effects on asset prices. The standard deviation of returns falls from 19.4 to 13.6 percent and the standard deviation of the price/dividend ratio falls from 20 percent to exactly zero. The agent's fear of a model with long-run risk thus raises the average equity premium by a factor of more than 4 and doubles the volatility of returns.

Going back to the first column, we see that there are large and persistent movements in the price/dividend ratio in our model. The autocorrelation of the price/dividend ratio at 0.96 is somewhat higher than the empirical autocorrelation, while the standard deviation is 0.20. 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 that moment. 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, believing dividends will continue to grow in the future. From the perspective of an econometrician, these 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 agent (or her worst-case model), there is almost no discount-rate news. Rather, she prices the equity claim differently over time due to beliefs about cash flows.

So from the perspective of an econometrician, this model seems to generate predictability in stock returns. To see that, figure 3 plots percentiles of sample R^2s from regressions of returns on price dividend ratios in 240-quarter samples (the approximate length of the post-war period). The gray line is the set of corresponding values from the empirical post-war (1950–2010) sample. We report R^2s for horizons of 1 quarter to 10 years. At both short and long horizons the model matches well. The median R^2 from the predictive regressions at the ten-year horizon is 37 percent, while in the data it is 29 percent. Again, even though we have no parameter that allows us to directly calibrate the model to match predictability regressions, the model compares favorably to the data.

A final feature of the data that many papers in the asset pricing literature 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 Campbell and Mankiw (1989), will generate EIS estimates of zero on average.

To summarize, in a model where agents have Epstein–Zin preferences and price assets under a worst-case model, we are able to match the mean and standard deviation of equity returns, the autocorrelation of the price/dividend ratio, and R^2s from forecasting regressions; we generate low and stable interest rates and substantial volatility in the price/dividend ratio (though still less than what is observed empirically); and we match the standard result that regressions of consumption growth on interest rates yield a coefficient of zero. All of these results are obtained without any extra free parameters compared to other models. Consumption growth is white noise, and the only parameter we can choose to manipulate asset prices is the coefficient of relative risk aversion.

Furthermore, our asset pricing results are actually richer than the homoskedastic long-run risk model, so it is not the case that ambiguity aversion is just a complicated way of obtaining the results from the long-run risks literature. The results on stock return predictability and the EIS regressions would not hold in the homoskedastic long-run risk model since they come from the fact that the true process driving consumption growth in this calibration is white noise.

7.1.2 Probability of rejecting the worst-case dynamics

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). The Ljung–Box test is based on a weighted sum of squared autocorrelations and is valid against a broad range of alternative models, which makes it robust, but also likely to have low power. As an alternative, Andrews and Ploberger (1996) argue that testing the null of white noise against the alternative of an ARMA(1,1) model can increase power. The ARMA(1,1) is parametrically parsimonious, but still allows for a wide range of alternative dynamics. Moreover, since the worst-case model is literally an ARMA(1,1), the Andrews–Ploberger test is in fact the correctly specified likelihood ratio test, implying it should be asymptotically most powerful.

We obtain small-sample critical values for the two test statistics by simulating their distributions under the null that the observed innovations in consumption growth are white noise. The reported rejection probabilities are then based on the simulated small-sample critical values.

Table 1 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. As expected, the Ljung–Box test is the weaker of the two, with rejection rates of 5 percent in both cases, while the ARMA(1,1) likelihood ratio test performs only slightly better, with rates of 5.3 and 6.3 percent in the 50- and 100-year samples, respectively. Table 1 thus shows that the worst-case model, while having economically large differences from the benchmark in terms of its asset pricing implications, can barely be distinguished from the benchmark in long samples. 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.

The results on rejection probabilities are critical to our claim that λ is calibrated reasonably. If λ were too small, the worst-case model that the investors consider would be easier to distinguish from the benchmark. What we find, though, is that λ seems to imply worst-case beliefs that are entirely plausible. Thus both λ and α (which were calibrated jointly with only a single degree of freedom) take on independently reasonable values.

7.2 Historical aggregate price/dividend ratios

To try to compare the model more directly to historical data, we now ask how the price/dividend ratio implied by the model compares to what we observe empirically. A natural benchmark is to treat the point estimate for consumption growth as a white-noise process and then use our model to construct the historical price/dividend ratio on a levered consumption claim and compare that to the price/dividend ratio on the aggregate equity market. As an alternative, we also try replacing consumption growth with dividend growth, which can be motivated either by treating expectations about dividend growth as being formed in the same way as those for consumption growth, or, alternatively, if dividends are viewed as the consumption flow of a representative investor.

Since the average level of the price/dividend ratio depends on average dividend growth, which we have not yet needed to calibrate, we simply set it so that the mean price/dividend ratio from the model matches the data.

Figure 4 plots the historical price/dividend ratio on the S&P 500 from Robert Shiller against the price/dividend ratios implied by consumption and dividend growth from 1882 to 2012 derived from the model (see the appendix). The consumption growth data is from Barro and Ursua (2010) (and extended by us to 2012), while the dividends also come from Shiller.[29]

Both the consumption- and dividend-implied price/dividend ratios perform well in matching historical price/dividend ratios up to the late 1970's, matching the declines in 1920 and 1929 particularly well. After 1975, the consumption-based measure no longer seems to match the true measure as well, while the dividend-based measure works until the enormous rise in valuations in the late 1990's. The full-sample correlations between the consumption- and dividend-based measures with the historical price/dividend ratio are 39 and 50 percent respectively. If we remove the post-1995 period, the correlation rises to 60 percent for the dividend-based measure, while it falls to 32 percent for the consumption based measure. One possible explanation for the change in the late 1990s is that discount rates fell persistently, as argued by Lettau, Ludvigson, and Wachter (2008). In the end, though, figure 5 shows that our simple model with white-noise consumption growth performs well in matching price/dividend ratios, at least up to the 1990s.

7.3 An AR(2) point estimate

To emphasize the robustness of the result that the ambiguity aversion we study here only affects the very low-frequency features of consumption growth, we now analyze a case where the point estimate for consumption dynamics is an AR(2) model. The AR(2) process we examine in this section implies a hump in the spectral density at an intermediate frequency. The point estimate generates power at business cycle frequencies as in many economic time series of interest (see Baxter and King (1999), for example), as illustrated by the spectral density shown in figure 5. The density is maximized at a frequency implying periodicity of approximately eight quarters. Thus the term $\bar{f}(\omega)$ in (26) varies across frequencies, potentially also generating extra variation in the worst-case model across frequencies compared to the white-noise case.

We first consider what worst-case model the agent would derive in solving problem (20) if she were constrained to minimize utility with respect to a transfer function implied by an AR(2). That

[29] The dividend-based series is highly similar to the exercise carried out by Barsky and De Long (1993), who also treat expected dividend growth as a geometrically weighted moving average of past growth rates, though without the ambiguity motivation used here, while the series based on consumption is similar to that used by Campbell and Cochrane (1999b), but without the non-linearities they included.

is, the problem is solved by choosing a worst-case $\{a_1, a_2, b_0\}$ in the model

$$\Delta c_t = \mu + a_1 (\Delta c_{t-1} - \mu) + a_2 (\Delta c_{t-2} - \mu) + b_0 \varepsilon_t \tag{44}$$

This setup is similar to a more standard type of problem where the agent only has parameter uncertainty, conditional on a tightly specified parametric form.[30] We hold the value of λ fixed at its calibration from the previous section, to ensure that the results are comparable. We also assume that when the agent chooses a worst-case model, she still uses the penalty function $g(b)$. The only difference is that b^w must be an AR(2), so the optimization problem is highly restricted.

The worst case that emerges implies the real transfer function, $B_r^w(\omega)$, plotted in figure 6; we refer to it as the parametric worst case. The parametric worst case is essentially indistinguishable from the point estimate – it is the gray line that lines up nearly perfectly with the black line representing the benchmark model in the figure. Intuitively, since there are only two free parameters, 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 non-parametric case, the parametric worst-case puts very small deviations on a wide range of frequencies.[31]

When we allow the agent to solve the more general problem (20) in an unrestricted manner, allowing for arbitrary dynamics and giving rise to the conditions in equation (26) the worst-case is very similar to what we obtained for the white-noise benchmark, as shown in figure 6. The figure is dominated by the non-parametric worst-case mainly deviating from the benchmark at very low frequencies. Again this reflects $Z(\omega)$ being small at all but the lowest frequencies, which implies that the worst-case leaves $B_r(\omega)$ essentially unchanged at all but very low frequencies. The worst-case thus inherits the local peak in power at middle frequencies that we observe in the benchmark AR(2).

[30] 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.

[31] The specific parameters in the benchmark and parametric worst-case models are (dropping μ for legibility)

$$\text{Benchmark:} \quad \Delta c_t = (0.70) \times \Delta c_{t-1} + (-0.35) \times \Delta c_{t-2} + (0.0051) \times \varepsilon_t \tag{45}$$

$$\text{Parametric worst case:} \quad \Delta c_t = (0.69) \times \Delta c_{t-1} + (-0.34)\, \Delta c_{t-2} + (0.0053) \times \varepsilon_t \tag{46}$$

The parametric worst-case model is thus clearly nearly identical to the benchmark model. Under the benchmark, $\bar{b}(\beta) = 0.0078$, while under the worst case, $b^w(\beta) = 0.0081$. 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 in the AR(2) model.

7.4 Endogenous consumption

Throughout the paper so far, we have taken consumption as exogenous. While the analysis of endowment economies is standard in the literature, it is a natural question in our case whether what investors would really be worried about is the risk that their consumption path is shifted by forces beyond their control to a bad path. It seems more natural to think that in the face of uncertainty, people would choose policies to ensure that consumption does not actually have persistently low growth. That is, a person who believes income growth will be persistently low in the future might simply choose to consume less now and smooth the level of consumption. In at least one important case with endogenous consumption, though, that intuition turns out to be incorrect and our results are unchanged.

Suppose investors have the same Epstein–Zin preferences over fixed consumption streams as above given a known model. Rather than taking consumption as exogenous, though, they choose it optimally. In each period, investors may either consume their wealth or invest it in a project with a random return r_{t+1}. The project may be thought of as either a real or financial investment, the only requirement is that it have constant returns to scale. The budget constraint is then

$$W_{t+1} = \exp(r_t) W_t - C_t \tag{47}$$

Investors perceive the return process as taking the same $MA(\infty)$ form as above

$$r_t = \mu_{ret} + b_{ret}(L)\varepsilon_t \tag{48}$$

As is well known, under log utility, an agent will always consume a constant fraction of current wealth, regardless of expectations for future returns, $C_t = (1-\beta)\exp(r_t) W_t$. That is, the income effect from higher future expected returns is exactly offset by the substitution effect caused by the increased price of consumption in the current period compared to future periods. Consumption itself then directly inherits the dynamics of returns: if investors believe that returns are persistent, then they also believe that consumption growth is persistent.

A common metaphor in the ambiguity aversion literature is that people play a game with an evil agent who chooses the worst-case model for returns conditional on the consumption policy the agent chooses. If that game has a Nash equilibrium, then our investor's consumption policy must be optimal taking the worst-case model chosen by the evil agent as given. So the investor understand that the evil agent will choose a process $b_{ret}^{w}(L)$, and she chooses an optimal consumption policy, taking that process as given. That optimal policy is to consume a constant fraction of wealth, so that the model driving consumption growth is exactly $b_{ret}^{w}(L)$.

Is is then straightforward to show that lifetime utility again takes the form,

$$v_t - c_t = \log(1-\beta) + \frac{\beta}{1-\beta}\log\beta + \frac{\beta}{1-\beta}\frac{(1-\alpha)}{2}b_{ret}(\beta)^2 \tag{49}$$

for a constant $\bar{v}$. If the worst-case process is again chosen as a minimization over lifetime utility, the worst-case model $b_{ret}(L)$ takes exactly the same form as in (26). Our analysis of worst-case persistence in consumption growth can thus also be interpreted as worst-case persistence in the returns to financial or real investment.

7.5 Stochastic volatility

So far, the analysis has purely addressed a homoskedastic model. The reason that the worst-case model is homoskedastic is that the investor considers only models of that form. Her estimation gives no description of how volatility varies over time. We now show that if the investor is allowed to explicitly estimate a model of stochastic volatility, the worst-case model will feature long-run risk in both consumption growth and volatility, as in the heteroskedastic version of the long-run risk model (Bansal and Yaron's (2004) case II). The results here thus further emphasize the extent to which equilibrium outcomes depend on assumptions about the precise form of model uncertainty that the investor has, in addition to showing that the full long-run risk model from Bansal and Yaron (2004) can be obtained as an outcome of model uncertainty.

Suppose that in addition to observing consumption growth, the investor also observes the volatility of the innovations to consumption growth in each period. She considers models of the form

$$\Delta c_t = b(L)\, s_{t-1}\varepsilon_t \tag{50}$$

$$s_t^2 = 1 + h(L)\,\mu_t \tag{51}$$

$$[\varepsilon_t, \mu_t]' \sim N(0_{2\times1}, I_2) \tag{52}$$

s_t controls the volatility of the innovations to consumption.[32] We assume that the investor uses the same non-parametric estimation method as above, but now she estimates the dynamics of both Δc_t and s_t. So she has point estimates $\bar{b}(L)$ and $\bar{h}(L)$. Using the same quadratic penalty function on deviations from the point estimate as before, $\int \frac{\left|b(e^{i\omega})-\bar{b}(e^{i\omega})\right|^2}{\left|\bar{b}(e^{i\omega})\right|^2} d\omega$ and $\int \frac{\left|h(e^{i\omega})-h(e^{i\omega})\right|^2}{\left|\bar{h}(e^{i\omega})\right|^2} d\omega$, we obtain worst-case transfer functions for consumption growth and its conditional variance,

$$B^w(\omega) - \bar{B}(\omega) = \frac{\alpha-1}{2}\frac{\beta}{1-\beta}\lambda_b^{-1}\bar{F}_b(\omega)\left[1+2\left(\frac{1-\alpha}{2}\beta b^w(\beta) h^w(\beta)\right)^2\right] b^w(\beta) Z(\omega) \quad (53)$$

$$H^w(\omega) - \bar{H}(\omega) = \frac{\alpha-1}{2}\frac{\beta}{1-\beta}\lambda_h^{-1}\bar{F}_h(\omega)\left(\frac{1-\alpha}{2}\beta b^w(\beta)\right)^2 h^w(\beta) Z(\omega) \quad (54)$$

where λ_b and λ_h are the multipliers on the penalties for deviations from the point estimates $\bar{b}$ and $\bar{h}$, respectively. The worst-case model for variance dynamics takes exactly the same form as the worst-case model for consumption dynamics in the homoskedastic model (equation XX). So, again, whatever the point estimate is for variance dynamics, the worst-case model adds a low-frequency component.

The worst-case model for consumption growth, $B^w(\omega)$, is now slightly modified from equation (XX). There is an additional term, $2\left(\frac{1-\alpha}{2}\beta b^w(\beta) h^w(\beta)\right)^2$. This term reflects the fact that an increase in the persistence of consumption growth has two effects. First, it increases the amount of long-run risk, lowering lifetime utility. That effect appears in the homoskedastic model. The second, new effect is that an increase in the amount of long-run risk increases the effects of a shock to s_t^2 on lifetime utility. That is, when consumption is more volatile to begin with, stochastic volatility is more painful. So the investor is now even more averse to long-run risk. However, this new term is constant across frequencies, so the types of models of consumption growth that she fears is unchanged.

As before, if the point estimates for consumption and variance dynamics are white noise, then the worst-case models both have an ARMA(1,1) form, which is observationally equivalent to the long-run risk specification. Now, since there is stochastic volatility, the model becomes analogous

[32] One way to motivate the idea that s_t can be observed directly would be that the innovation ε_t might be realized as a Brownian motion, the volatility of which can be estimated with arbitrary accuracy from high-frequency data. In other words, information is revealed slowly within each period, t, and the agent can observe volatility throughout the period.

Given the specification, s_t^2 can potentially become negative. Following the literature, we ignore that issue for the sake of simplicity here.

to Bansal and Yaron's (2004) case II, which is the most commonly used long-run risk specification.

8 Conclusion

This paper studies asset pricing when agents are unsure about the endowment process. The fundamental insight that the long-run risk model, precisely because it is difficult to test for empirically, represents a natural benchmark for investors to use when modeling the economy. More technically, for an agent with Epstein–Zin preferences who estimates consumption dynamics non-parametrically, 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 non-parametric 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 price/dividend ratio, identical persistence for the price/dividend ratio as what is observed empirically, 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: is rejected at the 5% 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

Daniel Andrei, Bruce Carlin, and Michael Hasler. Model disagreement, volatility, and trading volume. Working paper, 2013.

Donald W. K. Andrews and Werner Ploberger. Testing for serial correlation against an arma(1, 1,) process. *Journal of the American Statistical Association*, 91(435):pp. 1331–1342, 1996. ISSN 01621459. URL `http://www.jstor.org/stable/2291751`.

Ravi Bansal and Amir Yaron. Risks for the long-run: A potential resolution of asset pricing puzzles. *Journal of Finance*, 59(4):1481–1509, August 2004.

Ravi Bansal, Dana Kiku, and Amir Yaron. An empirical evaluation of the long-run risks model for asset prices. *Critical Finance Review*, 1(1):183–221, January 2012. URL `http://ideas.repec.org/a/now/jnlcfr/104.00000005.html`.

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

Robert J. Barro and Jose F. Ursua. Barro–ursua macroeconomic dataset. http://scholar.harvard.edu/barro/publications/barro-ursua-macroeconomic-data, 2010.

Robert B. Barsky and Bradford De Long. Why does the stock market fluctuate? *Quarterly Journal of Economics*, 108(2):291–311, May 1993.

Marianne Baxter and Robert G. King. Measuring business cycles: Approximate band-pass filters for economic time series. *The Review of Economics and Statistics*, 81(4):575–593, November 1999. URL `http://ideas.repec.org/a/tpr/restat/v81y1999i4p575-593.html`.

Jason Beeler and John Y. Campbell. The long-run risks model and aggregate asset prices: An empirical assessment. *Critical Finance Review*, 1(1):141–182, January 2012. URL `http://ideas.repec.org/a/now/jnlcfr/104.00000004.html`.

Kenneth N. Berk. Consistent autoregressive spectral estimates. *The Annals of Statistics*, 2(3):pp. 489–502, 1974. ISSN 00905364. URL `http://www.jstor.org/stable/2958135`.

R. J. Bhansali. Robustness of the autoregressive spectral estimate for linear processes with infinite variance. *Journal of Time Series Analysis*, 18(3):213–229, 1997.

Rhys Bidder and Matthew E. Smith. Doubts and variability: A robust perspective on exotic consumption series. Federal Reserve Bank of San Francisco, Working Paper No. 2013-28, 2013.

Michael J. Brennan and Yihong Xia. Stock price volatility and the equity premium. *Journal of Monetary Economics*, 47:249–283, 2001.

John Y. Campbell and John H. Cochrane. By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy*, 107(2):205–251, 1999a.

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

John Y. Campbell and Gregory Mankiw. Consumption, income, and interest rates: Reinterpreting the time series evidence. In Oliver Blanchard and Stanley Fischer, editors, *NBER Macroeconomics Annual: 1989*, pages 185–216. MIT Press, Cambridge, MA, 1989.

John Y. Campbell and Tuomo Vuolteenaho. Bad beta, good beta. *American Economic Review*, 94(5):1249–1275, 2004. doi: 10.1257/0002828043052240. URL `http://www.aeaweb.org/articles.php?doi=10.1257/0002828043052240`.

Stephen G. Cecchetti, Poksang Lam, and Nelson C. Mark. Asset pricing with distorted beliefs: Are equity returns to good to be true? *American Economic Review*, 90:787–805, 1997.

Timothy Cogley and Thomas J. Sargent. The market price of risk and the equity premium: A legacy of the great depression? *Journal of Monetary Economics*, 55(3):454–476, April 2008. URL `http://ideas.repec.org/a/eee/moneco/v55y2008i3p454-476.html`.

Pierre Collin-Dufrese, Michael Johannes, and Lars A. Lochstoer. Parameter learning in general equilibrium: The asset pricing implications. Working paper, 2013.

Pierre Collin-Dufresne, Michael Johannes, and Lars A. Lochstoer. Parameter learning in general equilibrium: The asset pricing implications. Working paper, 2013.

Massimiliano Croce, Martin Lettau, and Sydney C. Ludvigson. Investor information, long-run risk, and the term structure of equity. *Review of Financial Studies*, 2014.

Ian Dew-Becker and Stefano Giglio. Asset pricing in the frequency domain: theory and empirics. Working Paper, 2013.

Itamar Drechsler. Uncertainty, time-varying fear and asset prices. Unpublished paper, New York University, Stern School of Business, 2010.

Larry Epstein and Stan Zin. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical investigation. *Journal of Political Economy*, 99:555–576, 1991.

Larry G. Epstein and Martin Schneider. Learning under ambiguity. *Review of Economic Studies*, 74:1275–1303, 2007.

Andreas Fuster, Benjamin Hebert, and David Laibson. Natural expectations, macroeconomic dynamics, and asset pricing. Working Paper 17301, National Bureau of Economic Research, August 2011. URL `http://www.nber.org/papers/w17301`.

Itzhak Gilboa and David Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153, April 1989. URL `http://ideas.repec.org/a/eee/mateco/v18y1989i2p141-153.html`.

Robert M. Gray. Toeplitz and circulant matrices: A review. *Foundations and Trends in Communications and Information Theory*, 2(3):155–239, 2000.

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

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

Lars Peter Hansen and Thomas J. Sargent. Fragile beliefs and the price of uncertainty. *Quantitative Economics*, 1(1):129–162, 2010. ISSN 1759-7331. doi: 10.3982/QE9. URL `http://dx.doi.org/10.3982/QE9`.

Nigar Hashimzade and Timothy J. Vogelsang. Fixed-b asymptotic approximation of the sampling behaviour of nonparametric spectral density estimators. *Journal of Time Series Analysis*, 29(1): 142–162, 01 2008. URL `http://ideas.repec.org/a/bla/jtsera/v29y2008i1p142-162.html`.

David Hirshleifer and Jianfeng Yu. Asset pricing in production economies with extrapolative expectations. Working Paper, 2012.

Yongmiao Hong. Consistent testing for serial correlation of unknown form. *Econometrica*, 64(4): 837–864, 1996.

Nengjiu Ju and Jianjun Miao. Ambiguity, learning, and asset returns. *Econometrica*, 80(2):559–591, 2012.

David M. Kreps. Anticipated utility and dynamic choice. In Donald P. Jacobs, Edhud Kalai, Morton I. Kamien, and Nancy L. Schwartz, editors, *Frontiers of Research in Economic Theory: The Nancy L. Schwartz Memorial Lectures, 1983Ű1997*. Cambridge Univerity Press, 1998.

David M. Kreps and Evan L. Porteus. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica*, 46(1):185–200, 1978.

Markus Leippold, Fabio Trojani, and Paolo Vanini. Learning and asset prices under ambiguous information. *Review of Financial Studies*, 21(6):2656–2597, 2008.

Jun Liu, Jun Pan, and Tan Wang. An equilibrium model of rare-event premia and its implication for option smirks. *Review of Financial Studies*, 18(1):131–164, 2004.

G. M. Ljung and G. E. P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65 (2):pp. 297–303, 1978. ISSN 00063444. URL `http://www.jstor.org/stable/2335207`.

Fabio Maccheroni, Massimo Marinacci, and Aldo Rustichini. Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 74(6):1447–1498, 2006.

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

Pascal J. Maenhout. Robuuncertainty rules and asset pricing. *Review of Financial Studies*, 17(4): 951–983, 2004.

Ulrich K. Müller and Mark W. Watson. Measuring uncertainty about long-run predictions. Working paper, 2013.

M. B. Priestly. *Spectral Analysis and Time Series*. Academic Press, Harcourt Brace and Company, London, United Kingdom, 1981.

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

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

Costis Skiadas. Smooth ambiguity aversion toward small risks and continuous-time recursive utility. *Journal of Political Economy*, 121(4):775–792, 2013.

James H. Stock. Unit roots, structural breaks, and trends. In *Handbook of Econometrics*. Elsevier, 1994.

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

Pietro Veronesi. How does information quality affect stock returns? *Journal of Finance*, 55(2): 807–837, 2000.

Philippe Weil. The equity premium puzzle and the risk-free rate puzzle. *Journal of Monetary Economics*, 24(3):401–421, 1989.

A Asymptotic distribution of the transfer function for a non-parametric estimator

The goal is to estimate the transfer function $B(\omega)$. Berk (1974) studies estimates of the spectral density based on AR(p) processes. He shows that if p grows with the sample size T, such that $p \to \infty$ as $T \to \infty$ and $p^3/T \to \infty$, then the inverse of $B(\omega)$ will be normally distributed around its true value. Specifically,

$$
\begin{aligned}
A(\omega) &\equiv B(\omega)^{-1} && (55)\\
A(\omega) &= a_r(\omega) + i a_i(\omega) && (56)\\
\begin{bmatrix} a_r(\omega) \\ a_i(\omega) \end{bmatrix} &\sim N\left(\begin{bmatrix} a_r^{True}(\omega) \\ a_i^{True}(\omega) \end{bmatrix}, I_2 f^{True}(\omega)^{-1} \right) && (57)
\end{aligned}
$$

where $a_r(\omega)$ and $a_i(\omega)$ are the real and complex components of estimates of A and $f(\omega)$ is the spectral density, $f(\omega) = B(\omega) B(\omega)^* = \frac{1}{A(\omega)A(\omega)^*}$, with an asterisk denoting the complex conjugate.

Now we have

$$
\begin{aligned}
B(\omega) &= A^{-1}(\omega) && (58)\\
&= \frac{A(\omega)^*}{A(\omega) A(\omega)^*} && (59)\\
&= \frac{a_r(\omega) - i a_i(\omega)}{a_r(\omega)^2 + a_i(\omega)^2} && (60)
\end{aligned}
$$

We can use the delta method then to find the asymptotic distribution of the real and complex components of $B(\omega)$, which we denote $b_r(\omega)$ and $b_i(\omega)$. For compactness, drop the (ω) notation and the ^{True} superscripts for now. We have

$$B = b_r + i b_i \tag{61}$$

$$b_r = \frac{a_r}{a_r^2 + a_i^2} \quad \text{and} \quad b_i = \frac{-a_i}{a_r^2 + a_i^2}$$

The derivatives with respect to a_i and a_r are

$$\frac{db_r}{da_r} = \frac{a_i^2 - a_r^2}{\left(a_r^2 + a_i^2\right)^2} \tag{62}$$

$$\frac{db_r}{da_i} = \frac{-2a_r a_i}{\left(a_r^2 + a_i^2\right)^2} \tag{63}$$

$$\frac{db_i}{da_r} = \frac{2a_i a_r}{\left(a_r^2 + a_i^2\right)^2} \tag{64}$$

$$\frac{db_i}{da_i} = \frac{a_i^2 - a_r^2}{\left(a_r^2 + a_i^2\right)^2} \tag{65}$$

The covariance matrix of $[b_r, b_i]'$ is then

$$\left[\begin{array}{cc} \left(\frac{a_i^2-a_r^2}{\left(a_r^2+a_i^2\right)^2}\right)^2 + \left(\frac{-2a_r a_i}{\left(a_r^2+a_i^2\right)^2}\right)^2 & \frac{-2a_r a_i}{\left(a_r^2+a_i^2\right)^2}\frac{a_i^2-a_r^2}{\left(a_r^2+a_i^2\right)^2} + \frac{2a_i a_r}{\left(a_r^2+a_i^2\right)^2}\frac{a_i^2-a_r^2}{\left(a_r^2+a_i^2\right)^2} \\ \frac{-2a_r a_i}{\left(a_r^2+a_i^2\right)^2}\frac{a_i^2-a_r^2}{\left(a_r^2+a_i^2\right)^2} + \frac{2a_i a_r}{\left(a_r^2+a_i^2\right)^2}\frac{a_i^2-a_r^2}{\left(a_r^2+a_i^2\right)^2} & \left(\frac{a_i^2-a_r^2}{\left(a_r^2+a_i^2\right)^2}\right)^2 + \left(\frac{2a_r a_i}{\left(a_r^2+a_i^2\right)^2}\right)^2 \end{array}\right] f(\omega)^{-1} \tag{66}$$

$$= \left[\begin{array}{cc} \frac{1}{\left(a_r^2+a_i^2\right)} & 0 \\ 0 & \frac{1}{\left(a_r^2+a_i^2\right)} \end{array}\right] = I_2 f(\omega) \tag{67}$$

where I_2 is a 2×2 identity matrix and $\frac{1}{\left(a_r^2+a_i^2\right)} = \frac{1}{A(\omega)A(\omega)^*} = B(\omega)\,B(\omega)^* = f(\omega)$. The two components of $B(\omega)$ are thus independent with variances $f^{True}(\omega)$. Finally, then, a Wald statistic for a particular $B(\omega)$ is

$$\frac{\left(B(\omega) - B^{True}(\omega)\right)\left(B(\omega)^* - B^{True}(\omega)^*\right)}{f^{True}(\omega)} \tag{68}$$

B Time-domain results

The MA coefficients have a covariance matrix denoted Σ_m. Brockwell and Davis (1988) show that Σ_m converges, as $m \to \infty$, to a product,[33]

$$\Sigma_m \quad \to \quad J_m^{True} J_m^{True\prime} \tag{69}$$

$$\text{where } J_m^{True} \quad \equiv \quad \begin{bmatrix} b_0^{True} & b_1^{True} & \cdots & b_m^{True} \\ 0 & b_0^{True} & \cdots & b_{m-1}^{True} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_0^{True} \end{bmatrix} \tag{70}$$

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 coefficients (ignoring scale factors) then is

$$m^{-1} \left(\mathbf{b}_{1:m} - \bar{\mathbf{b}}_{1:m}\right) \left(\bar{J}_m \bar{J}_m'\right)^{-1} \left(\mathbf{b}_{1:m} - \bar{\mathbf{b}}_{1:m}\right)' \tag{71}$$

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 (Grenader and Szegő (1958) and Gray (2006)). So $\bar{\Sigma}_m^{-1}$ has an approximate orthogonal decomposition, converging as $m \to \infty$, such that[34]

$$\bar{\Sigma}_m^{-1} \approx \Lambda_m \bar{F}_m^{-1} \Lambda_m^* \tag{72}$$

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

[33] 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.

[34] 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 $\left[\bar{b}_0, \bar{b}_1, ..., \bar{b}_m\right]$.

$$m^{-1}\left(\mathbf{b}_{1:m}-\bar{\mathbf{b}}_{1:m}\right)\bar{\Sigma}_m^{-1}\left(\mathbf{b}_{1:m}-\bar{\mathbf{b}}_{1:m}\right)' \approx m^{-1}\left(\mathbf{B}_m\Lambda_m^*-\bar{\mathbf{B}}_m\Lambda_m^*\right)\Lambda_m\bar{F}_m^{-1}\Lambda_m^*\left(\mathbf{B}_m^{*\prime}\Lambda_m'-\bar{\mathbf{B}}_m^{*\prime}\Lambda_m'\right) \quad (73)$$

$$= m^{-1}\left(\mathbf{B}_m-\bar{\mathbf{B}}_m\right)\bar{F}_m^{-1}\left(\mathbf{B}_m-\bar{\mathbf{B}}_m\right)^* \quad (74)$$

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

$$m^{-1}\left(\mathbf{B}_m-\bar{\mathbf{B}}_m\right)\bar{F}_m^{-1}\left(\mathbf{B}_m-\bar{\mathbf{B}}_m\right)^*\to\int\frac{\left|B(\omega)-\bar{B}(\omega)\right|^2}{\bar{f}(\omega)}d\omega \quad (75)$$

C Lifetime utility

As discussed in the text, the agent's expectation of future consumption growth, $E_t\left[\Delta c_{t+j}|a,b_0\right]$ 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 $\{a,b_0\}$ drives consumption growth, we can write the innovations implied by that model as

$$\varepsilon_t^{\{a,b_0\}}=\left(\Delta c_t-\mu-a(L)\left(\Delta c_{t-1}-\mu\right)\right)/b_0 \quad (76)$$

The agent's subjective expectations for future consumption growth are then

$$E_t\left[\Delta c_{t+j}|a,b_0\right]=\mu+\sum_{j=0}^{\infty}b_{k+j}\varepsilon_{t-j}^{\{a,b_0\}} \quad (77)$$

with subjective distribution

$$\frac{\Delta c_{t+1}-E_t\left[\Delta c_{t+1}|a,b_0\right]}{b_0}\sim N(0,1) \quad (78)$$

We guess that vc takes the form

$$vc\left(\Delta c^t;a,b_0\right)=\bar{k}+\sum_{j=0}^{\infty}k_j\varepsilon_{t-j}^{\{a,b_0\}} \quad (79)$$

Inserting into the recursion for lifetime utility yields

$$\bar{k}+\sum_{j=0}^{\infty}k_j\varepsilon_{t-j}^{\{a,b_0\}} = \frac{\beta}{1-\alpha}\log E_t\left[\exp\left(\left(\bar{k}+\mu+\sum_{j=0}^{\infty}(k_j+b_j)\,\varepsilon_{t-j+1}^{\{a,b_0\}}\right)(1-\alpha)\right)|a,b_0\right] \tag{80}$$

$$= \beta\left(\bar{k}+\mu\right)+\beta\sum_{j=0}^{\infty}(k_{j+1}+b_{j+1})\,\varepsilon_{t-j}^{\{a,b_0\}}+\beta\frac{1-\alpha}{2}(k_0+b_0)^2 \tag{81}$$

Matching the coefficients on each side of the equality yields

$$vc\left(\Delta c^t;b\right) = \frac{\beta}{1-\beta}\frac{1-\alpha}{2}b(\beta)^2+\frac{\beta}{1-\beta}\mu+\sum_{k=1}^{\infty}\beta^k\sum_{j=0}^{\infty}b_{j+k}\varepsilon_{t-j}^{\{a,b_0\}} \tag{82}$$

$$= \frac{\beta}{1-\beta}\frac{1-\alpha}{2}b(\beta)^2+\frac{\beta}{1-\beta}\mu+\sum_{j=0}^{\infty}\left(\sum_{k=1}^{\infty}\beta^k b_{j+k}\right)\varepsilon_{t-j}^{\{a,b_0\}} \tag{83}$$

$$= \frac{\beta}{1-\beta}\frac{1-\alpha}{2}b(\beta)^2+\sum_{k=1}^{\infty}\beta^k E_t\left[\Delta c_{t+k}|a,b_0\right] \tag{84}$$

D Solving for the worst-case transfer function

The optimization problem is

$$B^w(\omega) = \arg\min_{b(L)}\frac{\beta}{1-\beta}\frac{1-\alpha}{2}\left(\sum_{j=0}^{\infty}b_j\beta^j\right)^2+\lambda\int\frac{\left|B(\omega)-\bar{B}(\omega)\right|^2}{\bar{f}(\omega)}d\omega \tag{85}$$

$$= \arg\min_{b(L)}\frac{\beta}{1-\beta}\frac{1-\alpha}{2}\left(\sum_{j=0}^{\infty}b_j\beta^j\right)^2+\lambda\int\frac{\left(\sum_{j=0}^{\infty}\exp(i\omega j)\left(b_j-\bar{b}_j\right)\right)\left(\sum_{j=0}^{\infty}\exp(-i\omega j)\left(b_j-\bar{b}_j\right)\right)}{\bar{f}(\omega)}d\omega \tag{86}$$

We guess that

$$B^w(\omega)=\bar{B}(\omega)+k\bar{f}(\omega)Z(\omega)^* \tag{87}$$

for a real constant k. The first-order condition for b_j is

$$0 = 2\frac{\beta}{1-\beta}\frac{1-\alpha}{2}b^w(\beta)\beta^j+\lambda\int\frac{\exp(i\omega j)\left(B(\omega)-\bar{B}(\omega)\right)^*+\exp(-i\omega j)\left(B(\omega)-\bar{B}(\omega)\right)}{\bar{f}(\omega)}d\omega \tag{88}$$

$$= 2\frac{\beta}{1-\beta}\frac{1-\alpha}{2}b^w(\beta)\beta^j+\lambda\int\frac{\exp(i\omega j)\,k\bar{f}(\omega)Z(\omega)+\exp(-i\omega j)\,k\bar{f}(\omega)Z(\omega)^*}{\bar{f}(\omega)}d\omega \tag{89}$$

$$= 2\frac{\beta}{1-\beta}\frac{1-\alpha}{2}b^w(\beta)\beta^j+2\lambda k\beta^j \tag{90}$$

where the third line follows from the definition of $Z(\omega) = \sum_{j=0}^{\infty} \beta^j \exp(-i\omega j)$. Clearly, then, our guess is a valid solution if

$$k = -\lambda^{-1} \frac{\beta}{1-\beta} \frac{1-\alpha}{2} b^w(\beta) \tag{91}$$

which is the result from the text,

$$B^w(\omega) = \bar{B}(\omega) + \lambda^{-1} \frac{\beta}{1-\beta} \frac{\alpha-1}{2} b^w(\beta) \bar{f}(\omega) Z(\omega)^* \tag{92}$$

E Solution for non-normal innovations

E.1 Lifetime utility

The recursion remains,

$$\begin{aligned}
\bar{k} + \sum_{j=0}^{\infty} k_j \varepsilon_{t-j}^{\{a,b_0\}} &= \frac{\beta}{1-\alpha} \log E_t \left[\exp\left(\left(\bar{k} + \mu + \sum_{j=0}^{\infty} (k_j + b_j) \varepsilon_{t-j+1}^{\{a,b_0\}} \right) (1-\alpha) \right) | a, b_0 \right] && (93) \\
&= \beta \left(\bar{k} + \mu \right) + \beta \sum_{j=0}^{\infty} (k_{j+1} + b_{j+1}) \varepsilon_{t-j}^{\{a,b_0\}} + \frac{\beta}{1-\alpha} \log E_t \left[\exp\left((k_0 + b_0) \varepsilon_{t+1}^{\{a,b_0\}} (1-\alpha) \right) \right] && (94) \\
&= \beta \left(\bar{k} + \mu \right) + \beta \sum_{j=0}^{\infty} (k_{j+1} + b_{j+1}) \varepsilon_{t-j}^{\{a,b_0\}} + \frac{\beta}{1-\alpha} \Gamma\left((k_0 + b_0)(1-\alpha) \right) && (95)
\end{aligned}$$

where Γ is the cumulant-generating function for ε. Again, by matching coefficients, we obtain

$$vc\left(\Delta c^t; b\right) = \frac{\beta}{1-\beta} \frac{1}{1-\alpha} \Gamma\left(b^w(\beta)(1-\alpha)\right) + \sum_{k=1}^{\infty} \beta^k E_t\left[\Delta c_{t+k} | a, b_0\right]$$

E.2 Worst-case transfer function

The optimization problem is now

$$\begin{aligned}
B^w(\omega) &= \arg\min_{b(L)} \frac{\beta}{1-\beta} \frac{1}{1-\alpha} \Gamma\left(b^w(\beta)(1-\alpha)\right) + \lambda \int \frac{\left|B(\omega) - \bar{B}(\omega)\right|^2}{\bar{f}(\omega)} d\omega && (9 \\
&= \arg\min_{b(L)} \frac{\beta}{1-\beta} \frac{1}{1-\alpha} \Gamma\left(b^w(\beta)(1-\alpha)\right) + \lambda \int \frac{\left(\sum_{j=0}^{\infty} \exp(i\omega j) \left(b_j - \bar{b}_j\right) \right) \left(\sum_{j=0}^{\infty} \exp(-i\omega j) \left(b_j - \bar{b}_j\right) \right)}{\bar{f}(\omega)} && (9
\end{aligned}$$

We again guess that

$$B^w(\omega) = \bar{B}(\omega) + k\bar{f}(\omega) Z(\omega)^* \tag{98}$$

for a real constant k. The first-order condition for b_j is

$$\begin{aligned}
0 &= \frac{\beta}{1-\beta}\Gamma'(b^w(\beta)(1-\alpha))\beta^j + \lambda\int \frac{\exp(i\omega j)\left(B(\omega)-\bar{B}(\omega)\right)^* + \exp(-i\omega j)\left(B(\omega)-\bar{B}(\omega)\right)}{\bar{f}(\omega)}d\omega \quad (99)\\
&= \frac{\beta}{1-\beta}\Gamma'(b^w(\beta)(1-\alpha))\beta^j + \lambda\int \frac{\exp(i\omega j)k\bar{f}(\omega)Z(\omega) + \exp(-i\omega j)k\bar{f}(\omega)Z(\omega)^*}{\bar{f}(\omega)}d\omega \quad (100)\\
&= \frac{\beta}{1-\beta}\Gamma'(b^w(\beta)(1-\alpha))\beta^j + 2\lambda k\beta^j \quad (101)
\end{aligned}$$

where the third line follows from the definition of $Z(\omega) = \sum_{j=0}^{\infty}\beta^j \exp(i\omega j)$. Clearly, then, our guess is a valid solution if

$$\begin{aligned}
k &= \frac{-\lambda^{-1}}{2}\frac{\beta}{1-\beta}\Gamma'(b^w(\beta)(1-\alpha)) \quad (102)\\
B^w(\omega) &= \bar{B}(\omega) - \frac{\lambda^{-1}}{2}\frac{\beta}{1-\beta}\Gamma'(b^w(\beta)(1-\alpha)) \quad (103)
\end{aligned}$$

Note also that the CGF for the standard normal distribution is $\Gamma(x) = x^2/2$, so $\Gamma'(x) = x$, so (102) reduces to (91) when ε is a standard normal.

F Asset prices and expected returns

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)

$$r_{t+1} = \delta_0 + \delta pd_{t+1} + \gamma\Delta c_{t+1} - pd_t \tag{104}$$

where δ is a linearization parameter slightly less than 1.

We guess that

$$pd_t = \bar{h} + \sum_{j=0}^{\infty} h_j \Delta c_{t-j} \tag{105}$$

for a set of coefficients $\bar{h}$ and h_j. Using the formula for lifetime utility from the text, we have

$$\begin{aligned} v_{t+1} - E_t v_{t+1} &= \sum_{k=0}^{\infty} \beta^k \Delta E_{t+1} \left[\Delta c_{t+k} | b^w \right] & (106) \\ &= b^w \left(\beta \right) \varepsilon_{t+1} & (107) \end{aligned}$$

So the pricing kernel can be written as

$$M_{t+1} = \beta \exp \left(-a^w \left(L \right) \Delta c_t - b_w^0 \varepsilon_{t+1} + \left(1 - \alpha \right) b^w \left(\beta \right) \varepsilon_{t+1} - \frac{\left(1 - \alpha \right)^2}{2} b^w \left(\beta \right)^2 \right) \tag{108}$$

The pricing equation for the levered consumption claim is

$$\begin{aligned} 0 &= \log E_t \left[\beta \exp \left(\begin{array}{c} \delta_0 + \left(\delta - 1 \right) \bar{h} + \left(\delta h_0 + \gamma - 1 \right) \Delta c_{t+1} + \sum_{j=0}^{\infty} \left(\delta h_{j+1} - h_j \right) \Delta c_{t-j} \\ + \left(1 - \alpha \right) b^w \left(\beta \right) \varepsilon_{t+1} - \frac{(1-\alpha)^2}{2} b^w \left(\beta \right)^2 \end{array} \right) \right] & (109) \\ &= \left(\delta h_0 + \gamma - 1 \right) a^w \left(L \right) \Delta c_t + \sum_{j=0}^{\infty} \left(\delta h_{j+1} - h_j \right) \Delta c_{t-j} & (110) \\ &\quad + \delta_0 + \left(\frac{1}{2} \left(\delta h_0 + \gamma - 1 \right)^2 \left(b_0^w \right)^2 + \left(\delta h_0 + \gamma - 1 \right) \left(1 - \alpha \right) b^w \left(\beta \right) b_0^w \right) + \left(\delta - 1 \right) \bar{h} + \log \beta & (111) \end{aligned}$$

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

$$\begin{aligned} \left(\delta - 1 \right) \bar{h} + \log \beta &= -\delta_0 - \left(\frac{1}{2} \left(\delta h_0 + \gamma - 1 \right)^2 \left(b_0^w \right)^2 + \left(\delta h_0 + \gamma - 1 \right) \left(1 - \alpha \right) b^w \left(\beta \right) b_0^w \right) & (112) \\ \left(\delta h_{j+1} - h_j \right) &= - \left(\delta h_0 + \gamma - 1 \right) a_j^w & (113) \end{aligned}$$

And thus

$$h_0 = \frac{\left(\gamma - 1 \right) a^w \left(\delta \right)}{1 - \delta a^w \left(\delta \right)}$$

$$\begin{aligned} \left(\delta - 1 \right) \bar{h} + \log \beta &= -\delta_0 - \left(\frac{1}{2} \left(\delta h_0 + \gamma - 1 \right)^2 \left(b_0^w \right)^2 + \left(\delta h_0 + \gamma - 1 \right) \left(1 - \alpha \right) b^w \left(\beta \right) b_0^w \right) & (114) \\ \bar{h} &= \frac{1}{1 - \delta} \left[\log \beta + \delta_0 + \left(\frac{1}{2} \left(\delta h_0 + \gamma - 1 \right)^2 \left(b_0^w \right)^2 + \left(\delta h_0 + \gamma - 1 \right) \left(1 - \alpha \right) b^w \left(\beta \right) b_0^w \right) \right] & (115) \end{aligned}$$

For the risk-free rate, we have

$$
\begin{aligned}
r_{f,t+1} &= -\log E_t\left[\beta \exp\left(-a^w(L)\,\Delta c_t - b_0^w \varepsilon_{t+1} + (1-\alpha)\, b^w(\beta)\, \varepsilon_{t+1} - \frac{(1-\alpha)^2}{2} b^w(\beta)^2\right)\right] && (116)\\
&= -\log \beta + a^w(L)\,\Delta c_t - \frac{1}{2}\left(b_0^w\right)^2 + b_0^w (1-\alpha)\, b^w(\beta) && (117)
\end{aligned}
$$

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

$$
\begin{aligned}
E_t\left[r_{t+1} - r_{f,t+1} | \bar{b}\right] &= E_t\left[\begin{array}{c} \delta_0 + (\delta - 1)\,\bar{h} + (\delta h_0 + \gamma)\,\Delta c_{t+1} - \sum_{j=0}^{\infty} (\delta h_0 + \gamma - 1)\, a_j^w \Delta c_{t-j} \\ + \log\beta - a^w(L)\,\Delta c_t + \frac{1}{2}\left(b_0^w\right)^2 - b_0^w (1-\alpha)\, b^w(\beta) \end{array} |\bar{b}\right] \\
&= (\delta h_0 + \gamma)\,(a(L) - a^w(L))\,\Delta c_t + \frac{1}{2}\left(b_0^w\right)^2 + \delta_0 + (\delta-1)\,\bar{h} + \log\beta - b_0^w (1-\alpha)\, b^w(\beta) \\
&= (\delta h_0 + \gamma)\,(a(L) - a^w(L))\,\Delta c_t + \frac{1}{2}\left(b_0^w\right)^2 + \delta_0 \\
&\quad -\delta_0 - \left(\frac{1}{2}(\delta h_0 + \gamma - 1)^2 \left(b_0^w\right)^2 + (\delta h_0 + \gamma - 1)(1-\alpha)\, b^w(\beta)\, b_0^w\right) - b_0^w (1-\alpha)\, b^w(\beta) \\
&= (\delta h_0 + \gamma)\,(a(L) - a^w(L))\,\Delta c_t + \frac{1}{2}\left(b_0^w\right)^2 \\
&\quad - \left(\frac{1}{2}(\delta h_0 + \gamma)^2 \left(b_0^w\right)^2 + \frac{1}{2}\left(b_0^w\right)^2 - (\delta h_0 + \gamma)\left(b_0^w\right)^2 + (\delta h_0 + \gamma - 1)(1-\alpha)\, b^w(\beta)\, b_0^w\right) - b_0^w
\end{aligned}
$$

$$
\begin{aligned}
&= (\delta h_0 + \gamma)\,(a(L) - a^w(L))\,\Delta c_t && (124)\\
&\quad - \left(\frac{1}{2}(\delta h_0 + \gamma)^2 \left(b_0^w\right)^2 - (\delta h_0 + \gamma)\left(b_0^w\right)^2 + (\delta h_0 + \gamma)(1-\alpha)\, b^w(\beta)\, b_0^w\right) && (125)\\
&= (\delta h_0 + \gamma)\,(a(L) - a^w(L))\,\Delta c_t - cov\,(m_{t+1}, r_{t+1}) - \frac{1}{2} var\,(r_{t+1}) && (126)
\end{aligned}
$$

Where

$$
\begin{aligned}
var_t^w\,(r_{t+1}) &= (\delta h_0 + \gamma)^2 \left(b_0^w\right)^2 && (127)\\
cov\,(r_{t+1}, m_{t+1}) &= (\delta h_0 + \gamma)\, b_0^w \left(-b_0^w + (1-\alpha)\, b^w(\beta)\right) && (128)\\
&= -(\delta h_0 + \gamma)\left(b_0^w\right)^2 + (\delta h_0 + \gamma)(1-\alpha)\, b^w(\beta)\, b_0^w && (129)
\end{aligned}
$$

Now to get the numerical solution, we need to be able to get $\delta h_0 + \gamma$. We have

$$\begin{aligned} \frac{b_0^w}{1-\delta a^w(\delta)} &= b^w(\delta) && (130)\\ 1-\delta a^w(\delta) &= \frac{b_0^w}{b^w(\delta)} && (131) \end{aligned}$$

And

$$a^w(\delta) = \delta^{-1}\left(1-\frac{b_0^w}{b^w(\delta)}\right) \quad (132)$$

So then

$$\begin{aligned} \delta h_0 + \gamma &= \delta\frac{(\gamma-1)\,a^w(\delta)}{1-\delta a^w(\delta)} + \gamma = \frac{\gamma-\delta a^w(\delta)}{1-\delta a^w(\delta)} && (133)\\ &= \delta\frac{(\gamma-1)\,\delta^{-1}\left(1-\frac{b_0^w}{b^w(\delta)}\right)}{\frac{b_0^w}{b^w(\delta)}} + \gamma && (134)\\ &= (\gamma-1)\,\frac{b^w(\delta)}{b_0^w} + 1 && (135) \end{aligned}$$

Finally, mean of the risk-free rate is

$$-\log\beta - \frac{1}{2}\left(b_0^w\right)^2 + b_0^w\,(1-\alpha)\,b^w(\beta) \quad (136)$$

And the standard deviation is

$$std\left(a^w(L)\,\Delta c_t\right) \quad (137)$$

When consumption growth is white noise, this is

$$\begin{aligned} std\left(a^w(L)\,\Delta c_t\right) &= std\left((\beta-\theta)\sum_{j=0}^{\infty}\theta^j\Delta c_{t-j}\right) && (138)\\ &= (\beta-\theta)\,\frac{\sigma_{\Delta c}}{\sqrt{1-\theta^2}} && (139) \end{aligned}$$

Finally, then

$$
\begin{aligned}
&= (\delta h_0 + \gamma)(a(L) - a^w(L))\Delta c_t && (140)\\
&\quad -\frac{1}{2}\left(\frac{\gamma - \delta a^w(\delta)}{1-\delta a^w(\delta)}\right)^2 (b_0^w)^2 - \frac{\gamma - \delta a^w(\delta)}{1-\delta a^w(\delta)} b_0^w \left[b_0^w - (1-\alpha) b^w(\beta)\right] && (141)\\
&= (\delta h_0 + \gamma)(a(L) - a^w(L))\Delta c_t - cov(m_{t+1}, r_{t+1}) - \frac{1}{2} var(r_{t+1}) && (142)
\end{aligned}
$$

$$
(\delta h_0 + \gamma)(b_0^w)^2 + (\delta h_0 + \gamma)(1-\alpha) b^w(\beta) b_0^w \qquad (143)
$$

F.1 Results for the white-noise benchmark

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

$$
\begin{aligned}
\Delta c_t &= \beta \Delta c_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1} && (144)\\
a^w(L) &= (\beta - \theta)\sum_{j=0}^{\infty} \theta^j L^j && (145)
\end{aligned}
$$

$$
a^w(\delta) = \frac{\beta - \theta}{1 - \theta\delta} \qquad (146)
$$

$$
b_j = \beta^j (\beta - \theta) \qquad (147)
$$

The price/dividend ratio is

$$
pd_t = \bar{k} + \frac{(\delta k_0 + \gamma - 1)(\beta - \theta)}{1 - \delta\theta}\sum_{j=0}^{\infty} \theta^j \Delta c_{t-j}
$$

and its standard deviation is

$$
\frac{(\delta k_0 + \gamma - 1)(\beta - \theta)}{1 - \delta\theta}\frac{\sigma_{\Delta c}}{\sqrt{1-\theta^2}} \qquad (148)
$$

G Test statistics and entropy constraints

G.1 KL distance

It is well known that Epstein–Zin preferences can be reinterpreted as the outcome of a robust-control model (see Hansen and Sargent (2005) and Barillas, Hansen, and Sargent (2009)). In those models, agents form expectations as though the innovation ε_t is drawn from a worst-case distribution. That distribution is chosen to minimize lifetime utility subject to a penalty on its distance from the benchmark of a standard normal, similarly to how we choose b^w here, and that distance depends on α. The coefficient of relative risk aversion in Epstein–Zin preferences, α, can therefore alternatively be interpreted as a penalty on a distance measure. We take advantage of that interpretation of recursive preferences here.

Formally, Barillas, Hansen, and Sargent (2009) model lifetime utility as

$$v^{BHS}\left(\Delta c^t; a, b_0\right) = (1-\beta)\, c_t + \beta \min_{\tilde{h}(\varepsilon_{t+1})} \left\{ \begin{array}{c} E_t\left[\frac{\tilde{h}(\varepsilon_{t+1})}{h(\varepsilon_{t+1})} \exp\left(\left(v^{BHS}\left(\Delta c^{t+1}; a, b_0\right)\right)\right) | a, b_0\right] \\ +\frac{1}{(\alpha-1)} E_t\left[\frac{\tilde{h}(\varepsilon_{t+1})}{h(\varepsilon_{t+1})} \log \frac{\tilde{h}(\varepsilon_{t+1})}{h(\varepsilon_{t+1})}\right] \end{array} \right\} \tag{149}$$

where $h(\varepsilon_{t+1})$ is the benchmark (standard normal) probability density for ε_{t+1}, and $\tilde{h}(\varepsilon_{t+1})$ is the worst-case density. The penalty function is

$$KL\left(h, \tilde{h}\right) \equiv E_t\left[\frac{\tilde{h}(\varepsilon_{t+1})}{h(\varepsilon_{t+1})} \log \frac{\tilde{h}(\varepsilon_{t+1})}{h(\varepsilon_{t+1})}\right] \tag{150}$$

Barillas, Hansen, and Sargent (2009) show that $v^{BHS}\left(\Delta c^t; a, b_0\right) = v\left(\Delta c^t; a, b_0\right)$, where $v\left(\Delta c^t; a, b_0\right)$ is the utility function under Epstein–Zin preferences defined in equation (15). We now show that the Kullback–Leibler (KL) distance can be interpreted as a penalty on a χ^2 test statistic.

Consider the case where

$$\varepsilon_t \sim {}_h N\left(0, \sigma^2\right) \tag{151}$$

$$\varepsilon_t \sim {}_{\tilde{h}} N\left(\mu, \sigma^2\right) \tag{152}$$

The KL distance is then $\frac{1}{2}\frac{\mu^2}{\sigma^2}$. The KL distance here is related to to a rejection probability. Suppose we take a sample of length T and split it into m equal-length groups (so we assume T is an integer multiple of m for the sake of simplicity). Denote the sum of ε_t in each of those m pieces as ε_i. We have

$$\varepsilon_i \sim {}_h N\left(0, \frac{T}{m}\sigma^2\right) \tag{153}$$

$$\varepsilon_i \sim {}_{\tilde{h}} N\left(\frac{T}{m}\mu, \frac{T}{m}\sigma^2\right) \tag{154}$$

So then

$$\sum_{i=1}^{m}\left(\frac{\varepsilon_i}{\left(\frac{T}{m}\right)^{1/2}\sigma}\right)^2 \sim {}_h\chi_m^2\,(0) \tag{155}$$

$$\sum_{i=1}^{m}\left(\frac{\varepsilon_i}{\left(\frac{T}{m}\right)^{1/2}\sigma}\right)^2 \sim {}_{\tilde{h}}\chi_m^2\left(m\left(\frac{\frac{T}{m}\mu}{\left(\frac{T}{m}\right)^{1/2}\sigma}\right)^2\right) = \chi_m^2\left(T\frac{\mu^2}{\sigma^2}\right) \tag{156}$$

where $\chi_m^2\,(k)$ denotes a non-central χ^2 variable with m degrees of freedom and non-centrality parameter k. Therefore, the standard χ^2 test statistic $\sum_{i=1}^{m}\left(\frac{\varepsilon_i}{\left(\frac{T}{m}\right)^{1/2}\sigma}\right)^2$, if the data is generated by $\tilde{h}$, is a χ_m^2 with non-centrality parameter $T\frac{\mu^2}{\sigma^2}$. Note that the non-centrality parameter is proportional to the sample size, showing how larger samples make it easier to reject the null given a fixed alternative. Moreover, the non-centrality parameter does not depend on m.

Finally, we have

$$\frac{1}{(\alpha-1)}KL\left(h,\tilde{h}\right) = \frac{1}{(\alpha-1)}\frac{1}{2}\frac{\mu^2}{\sigma^2} \tag{157}$$

$$= \frac{1}{(\alpha-1)}\frac{1}{T}\frac{1}{2}\left(T\frac{\mu^2}{\sigma^2}\right) \tag{158}$$

So the multiplier on the non-centrality parameter in the utility function is $\frac{1}{(\alpha-1)}\frac{1}{T}\frac{1}{2}$.

G.2 Spectral distance

We first simply define a measure of the goodness of fit. This applies to general time series models, and is basically a portmanteau test like Box and Pierce (1970), Ljung and Box (1978), and, most importantly, Hong (1996).

Suppose we have data generated by $b\,(L)$,

$$x_t = b\,(L)\,\varepsilon_t \tag{159}$$

where ε_t is i.i.d. standard normal white noise. For our measure of fit, we will filter by $\bar{b}\,(L)^{-1}$, where $\bar{b}\,(L)$ represents the null hypothesis of the point estimate. We then have

$$\frac{1}{\bar{b}\,(L)}x_t = \frac{b\,(L)}{\bar{b}\,(L)}\varepsilon_t \tag{160}$$

The way we test the fit of the model $\bar{b}$ is by measuring whether $\bar{b}(L)^{-1} x_t$ is white noise.

To test for white noise, we fit an $MA(m)$ to a sample of $\frac{1}{\bar{b}(L)} x_t$. m will grow with the sample size. Call the sample lag polynomial $\hat{d}(L)$. If $b(L) = \bar{b}(L)$, i.e., if the null hypothesis is correct, then $\hat{d}(L)$ should equal 1 on average. We also know the distribution of the $\hat{d}$ coefficients from Brockwell and Davis (1988).

Denote $\bar{d}(L) = \frac{b(L)}{\bar{b}(L)}$. The test statistic is

$$\frac{1}{m} \sum_{j=1}^{m} \hat{d}_j^2 \tag{161}$$

In the case where $\bar{d}(L) = 1$ (i.e. under the null), we have, using the results from Brockwell and Davis (1988), for $j > 1$, and $m, T \to \infty$ (at the appropriate joint rate),

$$T^{1/2} \hat{d}_j \Rightarrow N(0, 1) \tag{162}$$

and thus

$$\frac{1}{m} \sum_{j=1}^{m} T \hat{d}_j^2 \Rightarrow \chi_m^2 \tag{163}$$

under the null.

More generally, Brockwell and Davis (1988) show that $\hat{d}_j \Rightarrow N\left(\bar{d}_j, var\left(\bar{d}(L)\varepsilon_t\right)\right)$. So $\frac{1}{m}\sum_{j=1}^{m} T\hat{d}_j^2$ is not exactly a non-central χ^2 under the alternative hypothesis that $b(L) \neq \bar{b}(L)$. The difference appears because under the alternative hypothesis the variance of the test statistic is not the same as under the null. However, if $\bar{d}(L)$ is close enough to 1 asymptotically (i.e. b is close enough to $\bar{b}$), we can treat the deviation in the variance as small.

We now define vectors, $\hat{\mathbf{d}} \equiv \left[\hat{d}_1, \hat{d}_2, ..., \hat{d}_m\right]$, $\bar{\mathbf{d}} \equiv \left[\bar{d}_1, \bar{d}_2, ...\right]$. The null hypothesis is that $\bar{\mathbf{d}} = 0$. We model $\bar{\mathbf{d}}$ as local to zero by defining $\bar{\mathbf{d}}^T$ to be the value of $\bar{\mathbf{d}}$ in a sample of size T, and setting $\bar{\mathbf{d}}^T = \gamma T^{-1/.2}$ for some vector γ (with elements γ_j). A sample value of the MA polynomial in a sample of size T is denoted analogously as $\hat{\mathbf{d}}^T$. By scaling by $T^{-1/2}$, we are using a Pitman drift to study local power.

We have, from Brockwell and Davis (1988),

$$\begin{aligned} T^{1/2}\left(\hat{\mathbf{d}}^T - \bar{\mathbf{d}}^T\right) &\Rightarrow N\left(0, \Sigma_T\right) && (164) \\ T^{1/2}\hat{\mathbf{d}}^T &\Rightarrow N\left(T^{1/2}\bar{\mathbf{d}}^T, \Sigma_T\right) = N\left(\gamma, \Sigma_T\right) && (165) \end{aligned}$$

where Σ_T is the covariance matrix of $[x_t, x_{t-1}, ..., x_{t-m}]$ when $x_t = \bar{d}^T(L)\,\varepsilon_t$. Element i, j of Σ_T is $cov\left(x_{t-i}, x_{t-j}\right)$. For $j \neq 0$, we have

$$\begin{aligned} cov\left(x_t, x_{t-j}\right) &= \sum_{k=0}^{\infty} \bar{d}_k^T \bar{d}_{k+j}^T && (166) \\ &= \gamma_j T^{-1/2} + T^{-1}\sum_{k=1}^{\infty} \gamma_k \gamma_{k+j} && (167) \end{aligned}$$

where the first term comes from the fact that $\bar{d}_0^T = 1$. For $j = 0$,

$$\begin{aligned} cov\left(x_t, x_t\right) &= 1 + \sum_{k=1}^{\infty}\left(\bar{d}_k^T\right)^2 && (168) \\ &= 1 + T^{-1}\sum_{k=1}^{\infty}\gamma_k^2 && (169) \end{aligned}$$

We then define two matrices. Ω_1 has element (h, j) equal to $\sum_{k=1}^{\infty}\gamma_k\gamma_{k+|h-j|}$ and Ω_2 has element (h, j) equal to $\gamma_{|h-j|}$ if $h - j \neq 0$, and 0 if $h = j$. Finally, then

$$\Sigma_T = I + T^{-1}\Omega_1 + T^{-1/2}\Omega_2 \qquad (170)$$

We then have

$$\begin{aligned} T^{1/2}\hat{\mathbf{d}}^T &\Rightarrow N\left(T^{1/2}\bar{\mathbf{d}}^T, I + \Omega_1 T^{-1} + \Omega_2 T^{-1/2}\right) && (171) \\ T^{1/2}\left(\hat{\mathbf{d}}^T - \bar{\mathbf{d}}^T\right) &\Rightarrow \varepsilon_I + \varepsilon_{\Omega_1} T^{-1/2} + \varepsilon_{\Omega_2} T^{-1/4} && (172) \end{aligned}$$

where ε_x is a mean-zero normally distributed vector of innovations with covariance matrix x, for

$x \in \{I, \Omega_1, \Omega_2\}$. Therefore

$$
\begin{aligned}
T\left(\hat{\mathbf{d}}^T - \bar{\mathbf{d}}^T\right)\left(\hat{\mathbf{d}}^T - \bar{\mathbf{d}}^T\right)' &= \varepsilon_I \varepsilon_I' + \varepsilon_{\Omega_1}\varepsilon_{\Omega_1}' T^{-1} + \varepsilon_{\Omega_2}\varepsilon_{\Omega_2}' T^{-1/2} && (173)\\
&\quad + 2\varepsilon_I \varepsilon_{\Omega_1}' T^{-1/2} + 2\varepsilon_I \varepsilon_{\Omega_2}' T^{-1/4} + 2\varepsilon_{\Omega_1}\varepsilon_{\Omega_2}' T^{-3/4} && (174)
\end{aligned}
$$

The terms involving negative powers of T approach zero asymptotically, so we have

$$
\begin{aligned}
\left(T^{1/2}\hat{\mathbf{d}}^T - T^{1/2}\bar{\mathbf{d}}^T\right)\left(T^{1/2}\hat{\mathbf{d}}^T - T^{1/2}\bar{\mathbf{d}}^T\right)' &\approx \varepsilon_I \varepsilon_I' = \chi_m^2(0) && (175)\\
\left(T^{1/2}\hat{\mathbf{d}}^T\right)\left(T^{1/2}\hat{\mathbf{d}}^T\right)' &\approx \chi_m^2\left(T\bar{\mathbf{d}}^T\bar{\mathbf{d}}^{T\prime}\right) && (176)
\end{aligned}
$$

That is, the test statistic is a non-central χ_m^2, with non-centrality parameter $T\bar{\mathbf{d}}^T\bar{\mathbf{d}}^{T\prime}$.

Now, finally, we want to compute the test statistic. For any alternative hypothesis $\bar{\mathbf{d}}^T$, and defining $\bar{D}^T(\omega) = \bar{d}^T\left(e^{i\omega}\right)$, we have

$$
\begin{aligned}
\int \left|\bar{D}^T(\omega) - 1\right|^2 d\omega &= \int \left|\bar{D}^T(\omega)\right|^2 - \bar{D}^T(\omega) - \bar{D}^T(\omega)^* + 1 d\omega && (177)\\
&= \int \left|\bar{D}^T(\omega)\right|^2 - 1 d\omega && (178)\\
&= \sum_{j=1}^{\infty}\left(\bar{d}_j^T\right)^2 = \bar{\mathbf{d}}^T\bar{\mathbf{d}}^{T\prime} && (179)
\end{aligned}
$$

where the second line follows from the fact that $\int \bar{D}^T(\omega) = \bar{d}_0^T = 1$, and the third line is then just Parseval's theorem. So we have that for a given m, $\int \left|\bar{D}^T(\omega) - 1\right|^2 d\omega$ is equal to the non-centrality parameter in the χ_m^2 when you generate data under the model B^w and test the null that the data was driven by $\bar{B}$.

Now note that $\int \left|\bar{D}^T(\omega) - 1\right|^2 d\omega$ is exactly our $g(b)$ from the text. Specifically,

$$
\begin{aligned}
\int \left|\bar{D}^T(\omega) - 1\right|^2 d\omega &= \int \left|\frac{b(L)}{\bar{b}(L)} - 1\right|^2 d\omega && (180)\\
&= \int \left|\frac{b(L) - \bar{b}(L)}{\bar{b}(L)}\right|^2 d\omega && (181)\\
&= \int \frac{\left|b(L) - \bar{b}(L)\right|^2}{\bar{f}(L)} d\omega && (182)
\end{aligned}
$$

So we have that

$$g(b) = \sum_{j=1}^{\infty} \left(\bar{d}_j^T\right)^2 = \bar{\mathbf{d}}^T \bar{\mathbf{d}}^{T\prime} \tag{183}$$

$$\lambda g(b) = \lambda \bar{\mathbf{d}}^T \bar{\mathbf{d}}^{T\prime} \tag{184}$$

$$= \frac{\lambda}{T} \left(T \bar{\mathbf{d}}^T \bar{\mathbf{d}}^{T\prime}\right) \tag{185}$$

Therefore λT^{-1} is what multiplies the non-centrality parameter in a χ^2 specification test, the same as $\frac{1}{(\alpha-1)} \frac{1}{T} \frac{1}{2}$ multiplies a non-centrality parameter on the KL distance. To equate them, we say

$$\frac{1}{1-\beta} \frac{1}{(\alpha-1)} \frac{1}{T} \frac{1}{2} = \lambda T^{-1}$$

The extra term multiplying the left-hand side reflects the fact that the penalty on the dynamic model ambiguity is paid only once, while the penalty on the ε ambiguity is paid in every period.

Solving for α yields

$$\alpha = 1 + \frac{1}{2\lambda(1-\beta)} \tag{186}$$

G.3 Connecting α and λ

If we adopt the robust control reinterpretation of Epstein-Zin preferences, as in Barillas, Hansen and Sargent (2009) we can argue that one should connect α and λ through a single underlying assertion of the agent's preparedness to entertain models with a certain distance from the benchmark - in terms of b and the distribution of ε. The KL distance and spectral distance are thus both multiples of non-centrality parameters in χ_m^2 tests. We can then think of the agent's distance penalty for both being in terms of the same units. To equate them, we must set

$$\alpha = 1 + \frac{1}{2\lambda(1-\beta)} \tag{187}$$

We therefore have a direct mapping between α and λ. We can calibrate either α or λ to help fit the data, but then the other is pinned down immediately.[35]

[35] Implicit in these results is the thought experiment of generating data under the worst-case model and asking how likely one would be to reject the benchmark. There are two reasons for running the analysis in this direction, rather than the reverse. The first is that it is what we do when using the KL distance in standard robust-control models. $KL\left(h, \tilde{h}\right)$ is an expectation of a test statistic when data is generated by $\tilde{h}$. So to maintain the analogy, we calculate

Equation 187 asserts that there is a single fundamental parameter driving risk preferences. The ultimate focus of this paper is not, however, on econometric testing and the results derived here rely on assumptions that are in some cases special to this paper (such as the Gaussianity of ε_{t+1}). What this section does (again adopting the robustness interpretation of the Epstein-Zin recursion) is let us remove one degree of freedom from the parameterization of the model. Rather than having to calibrate both α and λ, we now only have a single degree of freedom in determining risk preferences, unlike Hansen and Sargent (2010) and Ju and Miao (2012) who calibrate *separate* parameters determining risk preferences for the different sources of uncertainty that they study.

That said, the remaining results in the paper hold even ignoring the content of this section if one wishes to take λ and α as independent parameters to be calibrated. We will also argue below that our calibrated values of α and λ are both reasonable, independently of each other. Indeed, through our choice of an extremely low α we have essentially bound our hands and rely on λ as our only free parameter and even this, in fact, is disciplined with the analysis of section 7.1.2.

H Endogenous consumption

Suppose the agent can invest in a single asset that faces log-normal shocks. The recursion for lifetime utility is

$$v_t = \max_{c_t} (1-\beta) c_t + \frac{\beta}{1-\alpha} \log E_t \exp((1-\alpha) v_{t+1}) \tag{188}$$

Wealth follows

$$W_{t+1} = R_t W_t - C_t \tag{189}$$

Think of W_t measuring investment in some technology that shifts consumption across dates. It might be a financial asset or it might be a real investment project with payoff R_t. It might also represent storage.

Now suppose

$$r_t \equiv \log R_t = b(L) \varepsilon_t \tag{190}$$

$g(b)$ also generating data under the worst-case model. The second reason is more practical. The distribution under which the data is generated is what determines the distribution of the test statistic. If the worst case model were to constitute the null of the test, then not only would the value of the test statistic change for different possible values of b considered by the agent, but also its distribution. By generating data from b and making the null always $\bar{b}$, different values of b are always compared to the same null distribution.

Lower-case letters are logs. We guess that $v_t = \bar{v} + v_w (w_t + r_t) + \sum_{j=0}^{\infty} v_j \varepsilon_{t-j}$. The optimization problem is then

$$\max_{c_t} (1-\beta) c_t + \frac{\beta}{1-\alpha} \log E_t \exp \left((1-\alpha) \left(\bar{v} + v_w \left(\log (R_t W_t - C_t) + r_{t+1} \right) + \sum_{j=0}^{\infty} v_j \varepsilon_{t+1-j} \right) \right) \tag{191}$$

$$= \max_{c_t} (1-\beta) c_t + \beta \left(\bar{v} + v_w \log (R_t W_t - C_t) + \sum_{j=1}^{\infty} (v_w b_j + v_j) \varepsilon_{t+1-j} \right) + \frac{1}{2} \beta (1-\alpha) (v_w + v_0)^2 \tag{192}$$

The first-order condition for consumption is

$$\frac{1-\beta}{C_t} = \frac{\beta v_w}{R_t W_t - C_t} \tag{193}$$

$$v_w \beta C_t = (1-\beta)(R_t W_t - C_t) \tag{194}$$

$$C_t = \frac{(1-\beta) R_t W_t}{v_w \beta + (1-\beta)} \tag{195}$$

So then,

$$R_t W_t - C_t = R_t W_t - \frac{(1-\beta) R_t W_t}{v_w \beta + (1-\beta)} \tag{196}$$

$$= R_t W_t \frac{v_w \beta}{v_w \beta + (1-\beta)} \tag{197}$$

We can then plug optimal consumption back into the equation for lifetime utility (starting from the point above where we already took the log-normal expectation)

$$\bar{v} + v_w (w_t + r_t) + \sum_{j=0}^{\infty} v_j \varepsilon_{t-j} = (1-\beta) \log \frac{(1-\beta) R_t W_t}{v_w \beta + (1-\beta)} + \frac{\beta (1-\alpha)}{2} (v_w b_0 + v_0)^2 + \tag{198}$$

$$\beta \left(\bar{v} + v_w \log \left(R_t W_t \frac{v_w \beta}{v_w \beta + (1-\beta)} \right) + \sum_{j=1}^{\infty} (v_w b_j + v_j) \varepsilon_{t+1-j} \right) \tag{199}$$

Matching coefficients (noting that $(w_t + r_t) = \log W_t R_t$)

$$v_w = (1-\beta) + \beta v_w = 1 \tag{200}$$

$$v_j = \beta (v_w b_{j+1} + v_{j+1}) \tag{201}$$

So for v_0,

$$v_0 = \beta b_1 + \beta v_1 \tag{202}$$
$$= \sum_{j=1}^{\infty} \beta^j b_j \tag{203}$$

We then have

$$\bar{v} = (1-\beta)\log(1-\beta) + \beta(\bar{v} + \log\beta) + \frac{\beta(1-\alpha)}{2} b(\beta)^2 \tag{204}$$
$$= \log(1-\beta) + \frac{\beta}{1-\beta}\log\beta + \frac{\beta(1-\alpha)}{(1-\beta)2} b(\beta)^2 \tag{205}$$

I Stochastic volatility

Suppose consumption follows an MA(∞) with volatility driven by a variable s_t,

$$\Delta c_t = b(L) s_{t-1}\varepsilon_t \tag{206}$$
$$s_t^2 = 1 + h(L)\mu_t \tag{207}$$

We guess that

$$vc_t = \bar{v} + \sum_{j=0}^{\infty}\left(v_{bj} s_{t-1-j}\varepsilon_{t-j} + v_{sj}\mu_{t-j}\right) \tag{208}$$

The recursion for lifetime utility is then

$$vc_t = \frac{\beta}{1-\alpha}\log E_t \exp\left((1-\alpha)\left(\bar{v} + \sum_{j=0}^{\infty}\left(v_{bj}s_{t-j}\varepsilon_{t+1-j} + v_{sj}\mu_{t+1-j}\right) + b(L)s_t\varepsilon_{t+1}\right)\right) \tag{209}$$
$$= \beta\bar{v} + \beta\sum_{j=1}^{\infty}\left((v_{bj}+b_j)s_{t-j}\varepsilon_{t+1-j} + v_{sj}\mu_{t+1-j}\right) + \frac{1-\alpha}{2}\beta\left((v_{b0}+b_0)^2 s_t^2 + v_{s0}^2\right) \tag{210}$$
$$= \beta\bar{v} + \beta\sum_{j=1}^{\infty}\left((v_{bj}+b_j)s_{t-j}\varepsilon_{t+1-j} + v_{sj}\mu_{t+1-j}\right) + \frac{1-\alpha}{2}\beta\left((v_{b0}+b_0)^2 + (v_{b0}+b_0)^2 h(L)\mu_t + v_{s0}^2\right) \tag{211}$$

Matching coefficients yields

$$\begin{aligned}\bar{v} &= \beta\bar{v}+\frac{1-\alpha}{2}\beta\left(\left(v_{b0}+b_0\right)^2+v_{s0}^2\right) &(212)\\ &= \frac{1-\alpha}{2}\frac{\beta}{1-\beta}\left(\left(v_{b0}+b_0\right)^2+v_{s0}^2\right) &(213)\end{aligned}$$

$$\begin{aligned}v_{bj} &= \beta\left(v_{b,j+1}+b_{j+1}\right) &(214)\\ v_{b0} &= \sum_{j=1}^{\infty}\beta^j b_j &(215)\end{aligned}$$

$$\begin{aligned}v_{sj} &= \beta v_{sj+1}+\frac{1-\alpha}{2}\beta\left(v_{b0}+b_0\right)^2 h_j &(216)\\ &= \beta v_{sj+1}+\frac{1-\alpha}{2}\beta b\left(\beta\right)^2 h_j &(217)\end{aligned}$$

$$\begin{aligned}v_{s,0} &= \beta v_{s,1}+\frac{1-\alpha}{2}\beta b\left(\beta\right)^2 h_0 &(218)\\ &= \frac{1-\alpha}{2}\beta b\left(\beta\right)^2 h_0+\beta\left(\beta v_{s,2}+\frac{1-\alpha}{2}\beta b\left(\beta\right)^2 h_1\right) &(219)\\ &= \frac{1-\alpha}{2}\beta b\left(\beta\right)^2 h\left(\beta\right) &(220)\end{aligned}$$

So we end up with

$$\bar{v}=\frac{1-\alpha}{2}\frac{\beta}{1-\beta}\left(b\left(\beta\right)^2+\left(\frac{1-\alpha}{2}\beta b\left(\beta\right)^2 h\left(\beta\right)\right)^2\right) \tag{221}$$

The easiest way to solve this is using the matrix form. We define $z=\left[1,\beta,\beta^2,...\right]$. The optimization problem is to minimize over the vectors of coefficients b and h

$$\frac{1-\alpha}{2}\frac{\beta}{1-\beta}b'zz'b+\frac{1-\alpha}{2}\frac{\beta}{1-\beta}\left(\frac{1-\alpha}{2}\beta\left(b'zz'b\right)z'h\right)^2+\lambda_b\left(b-\bar{b}\right)'\Sigma_b^{-1}\left(b-\bar{b}\right)+\lambda_h\left(h-\bar{h}\right)'\Sigma_h^{-1}\left(h-\bar{h}\right) \tag{222}$$

The FOCs are

$$\begin{aligned}b &: 0=2\frac{1-\alpha}{2}\frac{\beta}{1-\beta}zz'b+4\frac{1-\alpha}{2}\frac{\beta}{1-\beta}\left(\frac{1-\alpha}{2}\beta z'h\right)^2\left(b'zz'b\right)zz'b+2\lambda_b\Sigma_b^{-1}\left(b-\bar{b}\right) &(223)\\ h &: 0=2\frac{1-\alpha}{2}\frac{\beta}{1-\beta}\left(\frac{1-\alpha}{2}\beta\left(b'zz'b\right)\right)^2 zz'h+2\lambda_h\Sigma_h^{-1}\left(h-\bar{h}\right) &(224)\end{aligned}$$

Solving for b,

$$-\lambda_b \Sigma_b^{-1} \left(b - \bar{b}\right) = \frac{1-\alpha}{2} \frac{\beta}{1-\beta} \left(1 + 2\left(\frac{1-\alpha}{2} \beta h^w (\beta) b^w (\beta)\right)^2\right) zz'b \tag{225}$$

$$\left(b - \bar{b}\right) = \frac{\alpha - 1}{2} \frac{\beta}{1-\beta} \lambda_b^{-1} \Sigma_b b^w (\beta) \left[1 + 2\left(\frac{1-\alpha}{2} \beta b^w (\beta) h^w (\beta)\right)^2\right] z \tag{226}$$

To shift this to the frequency domain, we just recall that

$$\Sigma_b = \Lambda \bar{F}_b \Lambda' \tag{227}$$

so that

$$\begin{aligned} \left(b - \bar{b}\right) &= \frac{\alpha - 1}{2} \frac{\beta}{1-\beta} \lambda_b^{-1} \Lambda \bar{F}_b \Lambda' b^w (\beta) \left[1 + 2\left(\frac{1-\alpha}{2} \beta b^w (\beta) h^w (\beta)\right)^2\right] z & (228) \\ \Lambda' b - \Lambda' \bar{b} &= \frac{\alpha - 1}{2} \frac{\beta}{1-\beta} \lambda_b^{-1} \bar{F}_b b^w (\beta) \left[1 + 2\left(\frac{1-\alpha}{2} \beta b^w (\beta) h^w (\beta)\right)^2\right] \Lambda' z & (229) \\ B(\omega) - \bar{B}(\omega) &= \frac{\alpha - 1}{2} \frac{\beta}{1-\beta} \lambda_b^{-1} \bar{F}_b (\omega) b^w (\beta) \left[1 + 2\left(\frac{1-\alpha}{2} \beta b^w (\beta) h^w (\beta)\right)^2\right] Z(\omega)^* & (230) \end{aligned}$$

Solving the FOC for h,

$$\begin{aligned} -\lambda_h \Sigma_h^{-1} \left(h - \bar{h}\right) &= \frac{1-\alpha}{2} \frac{\beta}{1-\beta} \left(\frac{1-\alpha}{2} \beta b^w (\beta)\right)^2 z h^w (\beta) & (231) \\ h - \bar{h} &= \frac{\alpha - 1}{2} \frac{\beta}{1-\beta} \lambda_h^{-1} \Sigma_h \left(\frac{1-\alpha}{2} \beta b^w (\beta)\right)^2 h^w (\beta) z & (232) \end{aligned}$$

Again, we can decompose Σ_h using the spectrum of h,

$$H(\omega) - \bar{H}(\omega) = \frac{\alpha - 1}{2} \frac{\beta}{1-\beta} \lambda_h^{-1} \bar{F}_h (\omega) \left(\frac{1-\alpha}{2} \beta b^w (\beta)\right)^2 h^w (\beta) Z(\omega) \tag{233}$$

If the point estimate is that both consumption and volatility follow independent white noise, then the two worst-case models are ARMA(1,1)'s with persistence equal to β.

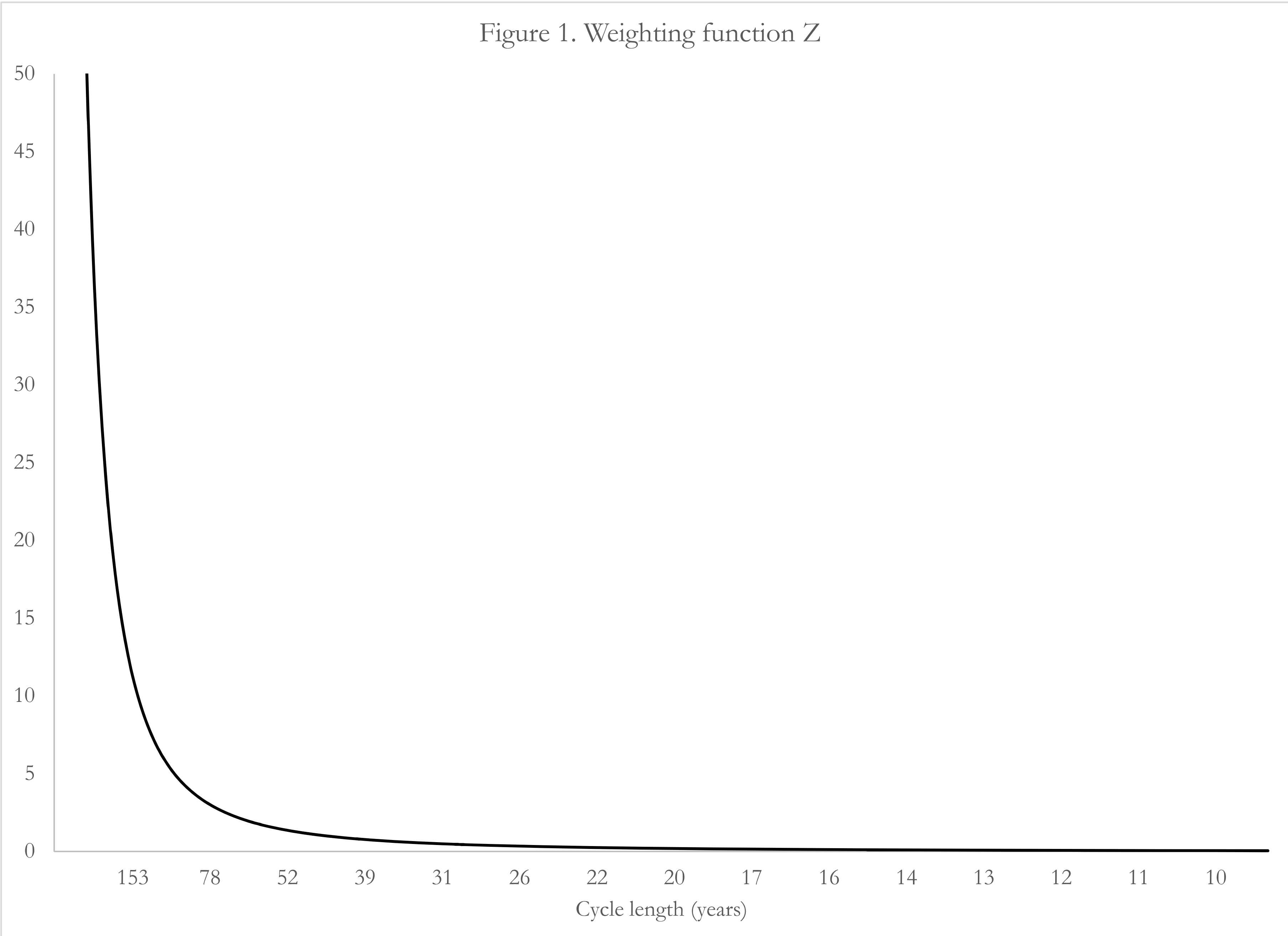


Figure 1. Weighting function Z
50
45
40
35
30
25
20
15
10
5
0
153
78
52
39
31
26
22
20
17
16
14
13
12
11
10
Cycle length (years)

Figure 2. Transfer function under benchmark and worst case for white-noise benchmark

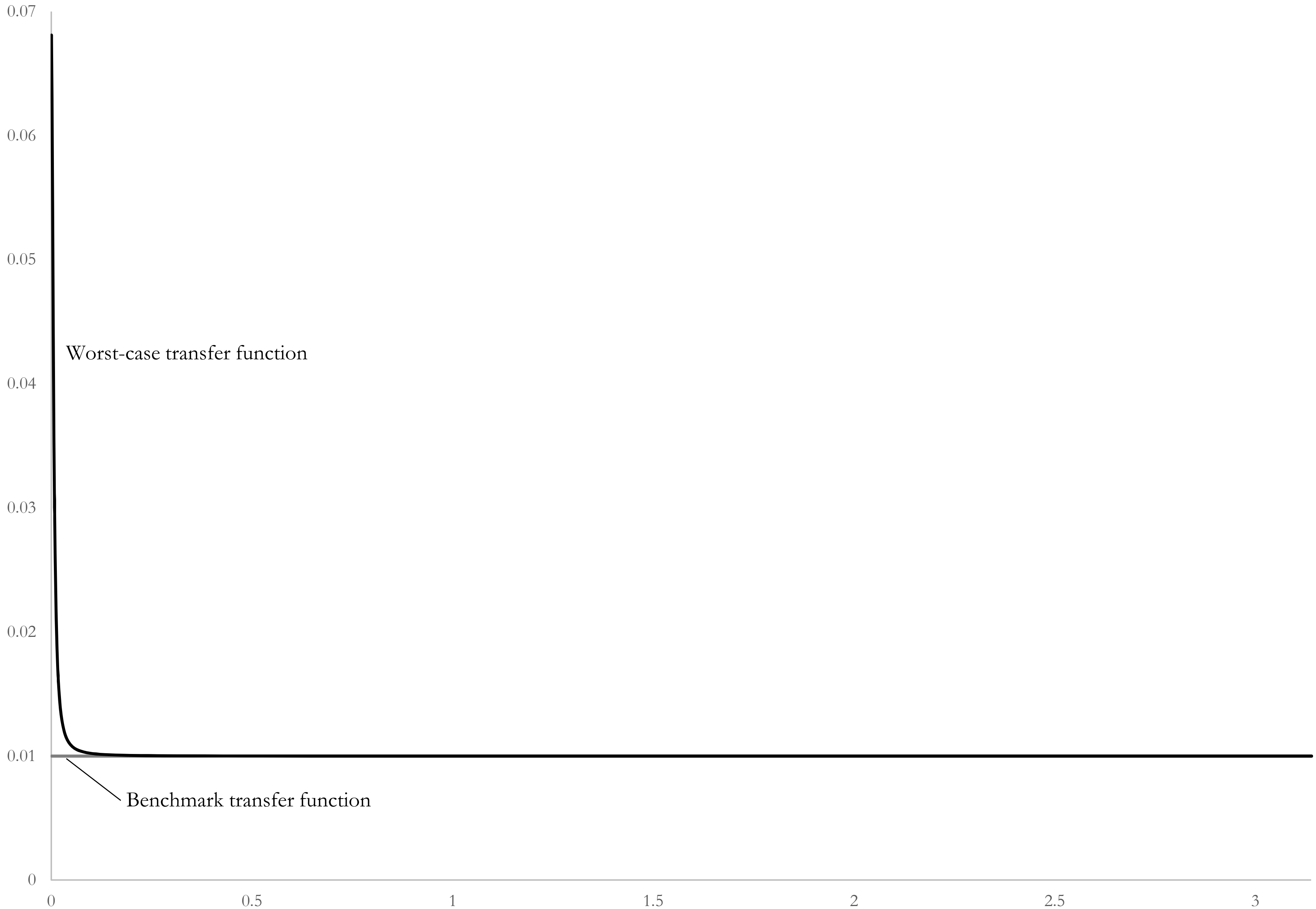
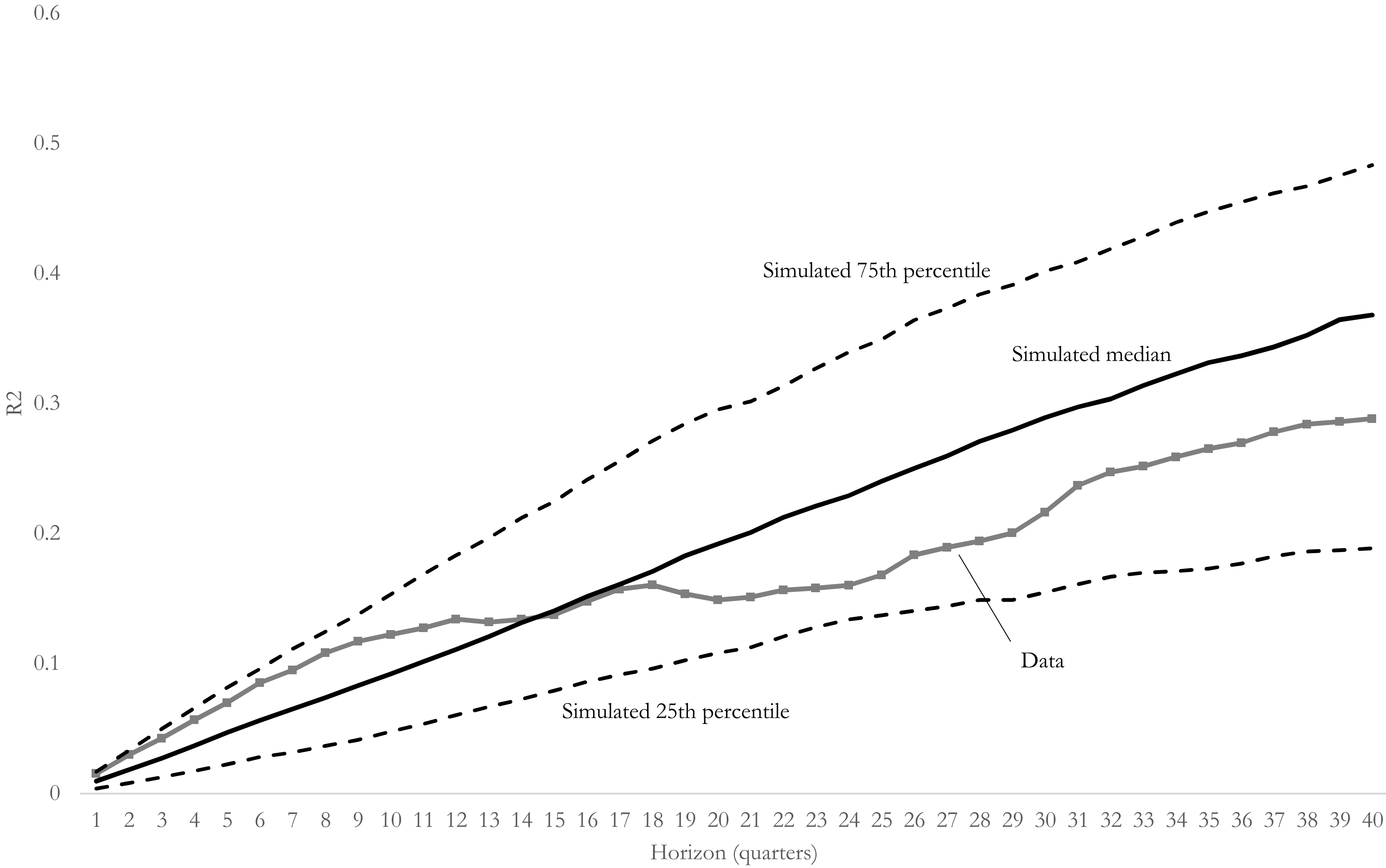
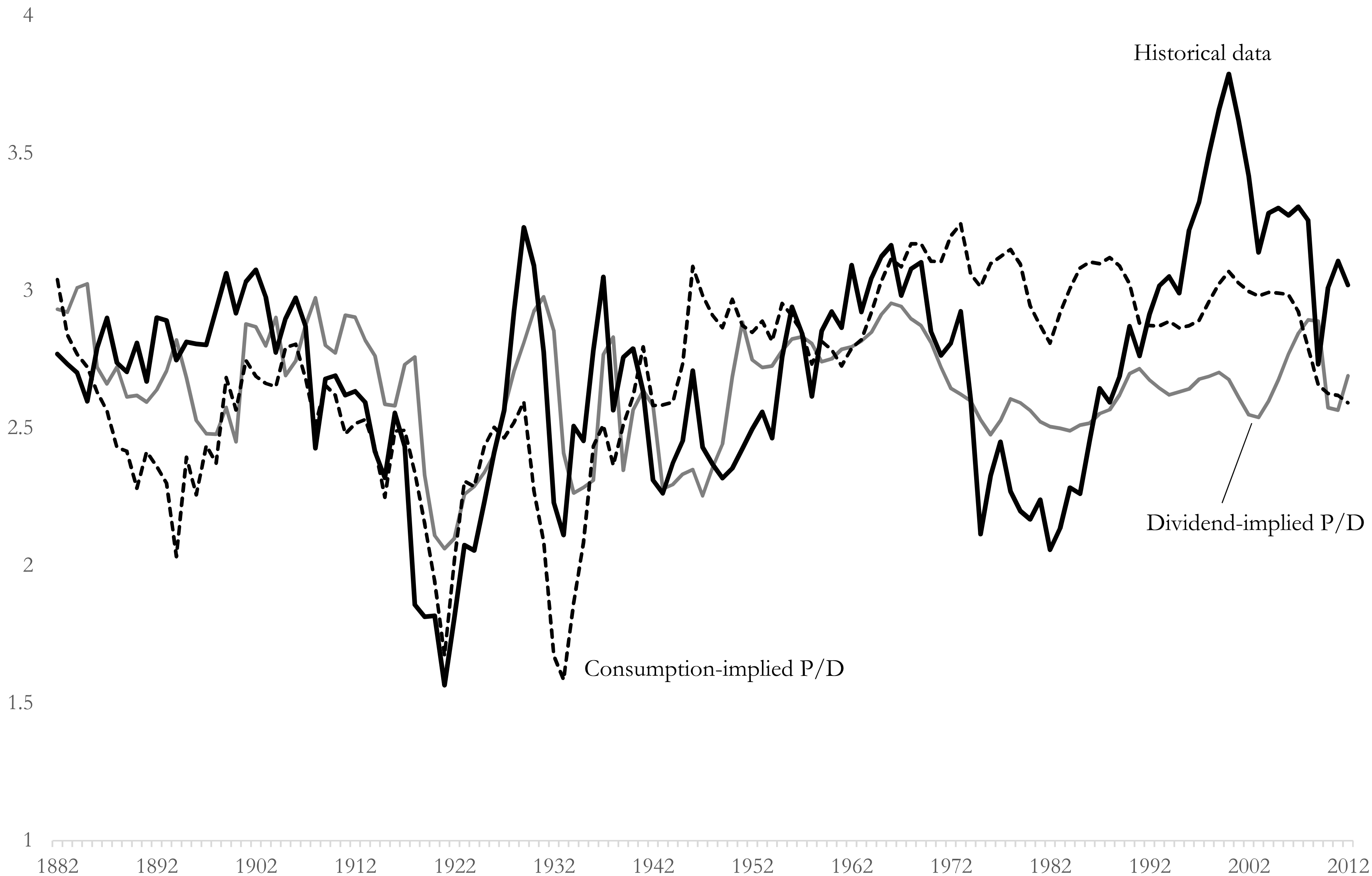


Figure 3. Empirical and model-implied R2's from return forecasting regressions



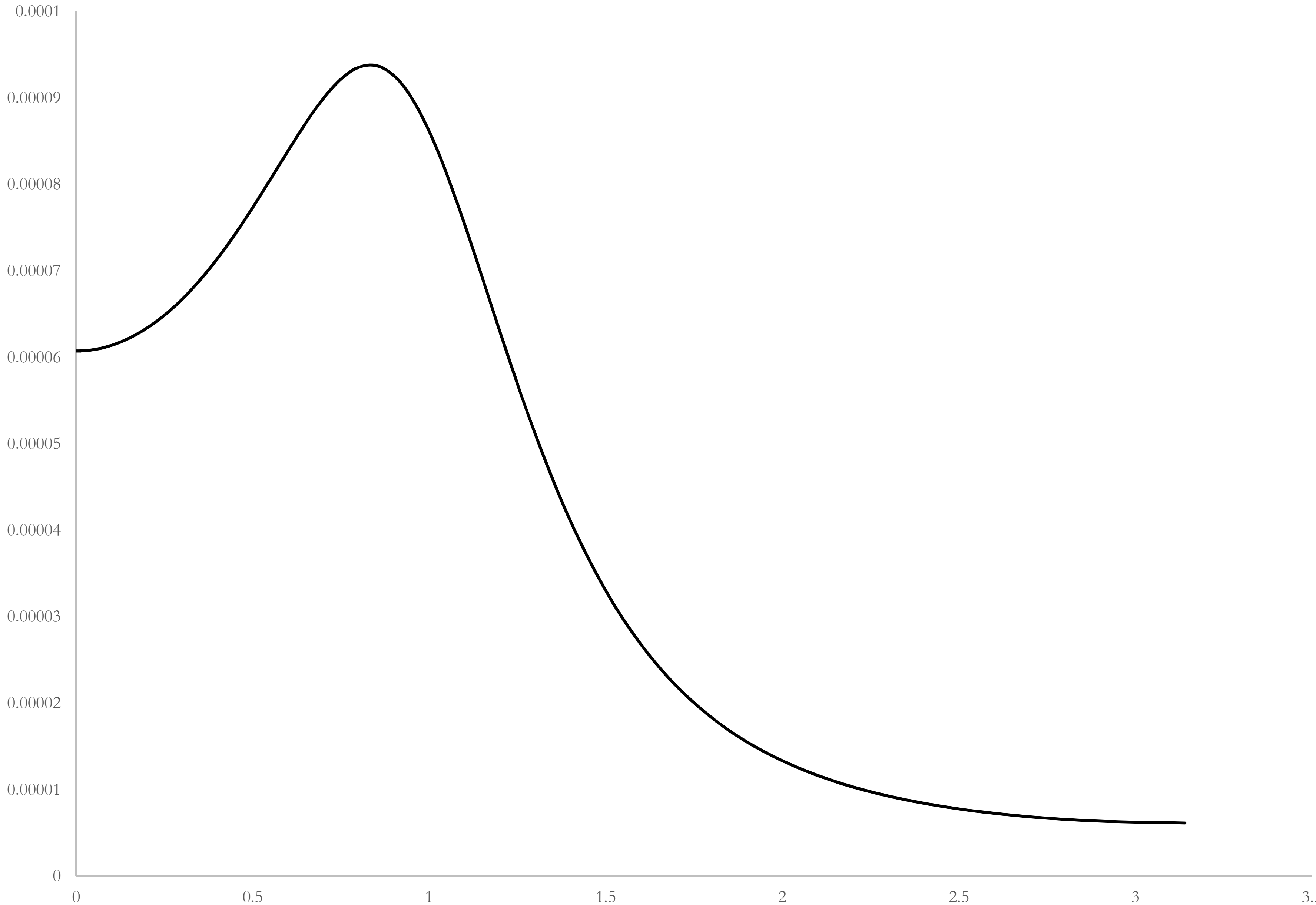
Notes: Black lines give results from simulated regressions on 50-year samples. The grey line plots R2s from regressions of aggregate equity returns on the price/dividend ratio in the post-war sample.

Figure 4. Historical and model-implied log price/dividend ratios



Notes: The historical price/dividend ratio is for the S&P 500 from Robert Shiller. The consumption-based measure uses data from Barro and Ursua (2008). The dividend-based measure uses Shiller's data on dividends. All three series have the same mean by construction.

Figure 5. Spectral density of benchmark AR(2) process (coefficients = {0.70, -0.35})



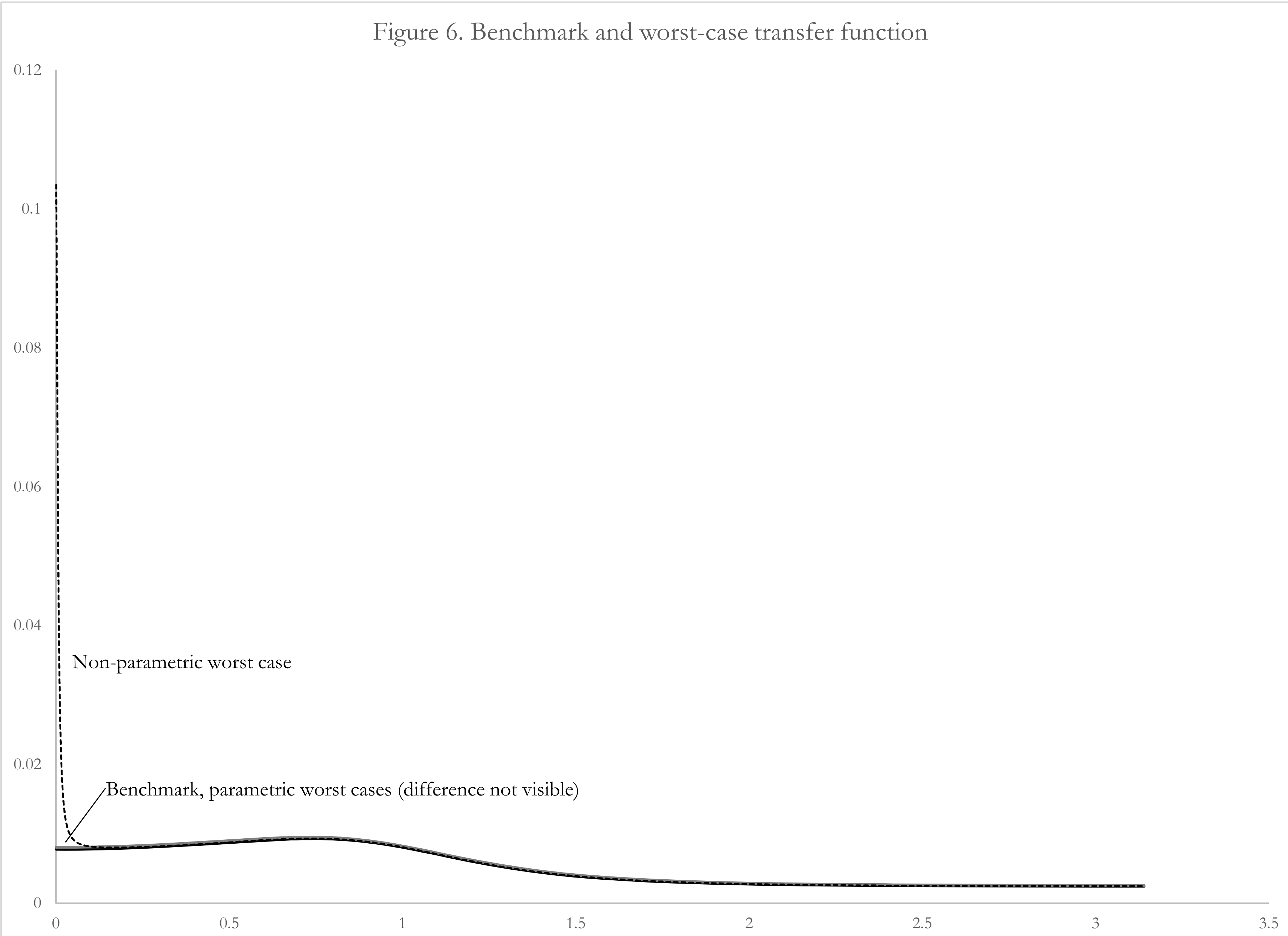


Figure 6. Benchmark and worst-case transfer function
0.12
0.1
0.08
0.06
0.04
0.02
0
0
0.5
1
1.5
2
2.5
3
3.5
Non-parametric worst case
Benchmark, parametric worst cases (difference not visible)

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

Fundamental parameters			*Implied worst-case*	
$\overline{b_0}$	Cons. vol. point est.	0.0147	b_0^w	0.015
$\overline{b}(\beta)$	Long-run vol point est.	0.0147	$b^w(\beta)$	0.039
μ	Mean cons. growth	0.0045	θ	0.98932
β	Time discount	0.997		
λ	Ambiguity aversion	52.24	*Standard Epstein–Zin / robust-control*	
α	RRA (implied by λ)	4.81	b_0	0.0147
γ	Leverage	4.63	$b(\beta)$	0.0147

Asset pricing moments (annualized)

	Model	Standard RC	Data
std(M)	0.33	0.14	
E[r-rf]	6.33	1.91	6.33
std(r)	19.42	13.55	19.42
E[rf]	1.89	2.43	0.86
std(rf)	0.33	0	0.97
AC1(PD)	0.96	N/A	0.81
std(P/D)	0.20	0	0.29

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

	50-year sample	100-year sample
Ljung–Box	4.9%	5.1%
ARMA(1,1)	5.3%	6.3%

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. For the asset pricing moments, r is the log return on the levered consumption claim, and rf is the risk-free rate. 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 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 test statistics in the latter simulation are outside the 95% range in the former simulation.