Long-run risk is the worst-case scenario: ambiguity aversion and non-parametric estimation of the endowment process

Rhys Bidder and Ian Dew-Becker*

August 29, 2014

Abstract

We study an agent 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 is a simple and tractable way of 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 5.3 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 Stefano Giglio and seminar participants at the Federal Reserve Bank of San Francisco, Duke Fuqua, and the Asset Pricing Retreat in Tilburg. The views expressed in this paper are those of the authors and not necessarily those of the Federal Reserve Bank of San Francisco, the Federal Reserve Board of Governors or the Federal Reserve System.

1 Introduction

Economists do not agree on the dynamic properties of the economy. For example, 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 agent who considers a set of endowment processes 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 specific 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 her utility and how she prices financial assets.

More specifically, the question that we then ask is, for an agent who has Epstein–Zin (1991) preferences, what model of consumption growth would yield the lowest lifetime utility, subject to a constraint on the implausibility of the model? The headline result is that for an agent whose point estimate is that consumption growth is white noise, the worst-case model that she will use for decision making is an ARMA(1,1) with a highly persistent trend – the homoskedastic version of Bansal and Yaron's (2004) long-run risk model. More generally, whatever the agent's point estimate, the worst-case model simply adds a long-run risk component to it.

The long-run risk model 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. 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 – they decay at exactly the rate of pure time preference.[1]

Beyond the specific result about the long-run risk model, we obtain three key sets of results for general models, all of them analytic and transparent. First, we derive formulas for exactly what the worst-case

[1] The volatility shock in Bansal and Yaron (2004) has a higher persistence, with a half-life of 4.5 years. In Bansal, Kiku, and Yaron (2012), the persistence of volatility is even higher, with a monthly autocorrelation of 0.999. But even that calibration is less persistent than the worst-case model that we obtain.

model is 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. The paper's central result is that the idea that agents price assets as though there are long-run risks in consumption growth is natural and to be expected when agents are unsure of the true process driving consumption growth. This finding 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 the agent'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 she is 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.

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 price/dividend ratio, persistence in the price/dividend ratio almost exactly equal to the observed value, 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 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. 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 5.3. 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 10 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.[2] 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.[3] The work of Hansen and Sargent (2010) is perhaps most comparable to ours, in that they study the case of 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.

We take the agent's estimation problem seriously, allowing her to use non-parametric methods that are suitable in situations in which a broad class of data generating processes could describe her consumption process. The non-parametric methods allow the agent to calculate a penalty for almost any ARMA model. Models farther from the point estimate in a statistical sense are viewed as less plausible. The worst-case model is one that is both particularly painful in utility terms and also difficult to reject. We show that, in general, the worst-case model is nearly identical to the agent's point estimate in its high-frequency behavior, but with substantial power added at very low frequencies, which have the largest effects on utility.

The restriction in past work to specific parametric models (e.g. Hansen and Sargent (2010), Ju and Miao (2012)) is therefore potentially important. Limiting the agent to only consider specific ARMA specifications

[2] See, e.g., Kreps and Porteus (1978); Weil (1989); Epstein and Zin (1991); and Hansen and Sargent (2005), among many others.

[3] 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).

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.

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): agents behave as though the endowment has a small highly persistent component, which leads them to excessively extrapolate based on recent observations. Agents overpay for risky assets in good times and underpay in bad times (from the perspective of an econometrician who has the same point estimate as the agents), which leads to countercyclically predictable asset returns. In that regard, our paper is related in its driving mechanisms and implications to the literature on belief distortions and extrapolative expectations.[5] We differ from the papers on extrapolative expectations, though, in that we derive the extrapolative model as a worst-case model, rather than as a result of bounded rationality or limited information processing ability.

The remainder of the paper is organized as follows. section 2 describes the basic structure of the agent's preferences. Section 3 discusses the agent's estimation method, 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 of the model and section 8 concludes.

2 Preferences and endowment models

We study pure endowment economies. Agents form their expectations for future consumption growth using a class of models of the form

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

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

[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, in particular, Barsky and De Long (1993), Cecchetti et al. (1997), Fuster et al. (2011) and Hirshleifer and Yu (2012)

where $a(\cdot)$ 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, conditional on a lag polynomial a and a scaling of the innovation by b_0. The agent is not sure of the values of a and b_0.[6]

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

$$\begin{aligned} \Delta c_t &= b(L)\,\varepsilon_t && (3)\\ b(L) &\equiv (1 - La(L))^{-1} b_0 && (4)\\ &= \sum_{j=1} b_j L^j && (5) \end{aligned}$$

We can thus express the dynamics of consumption equivalently as depending on $\{a, b_0\}$ or just on the polynomial b.[7] 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 $a(L)$ as 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. We discuss necessary constraints on the models below. For now simply assume that they are sufficiently constrained that any quantities we must derive exist.

2.1 Utility for fixed beliefs

The agent faces two sources of uncertainty: the innovations and the propagation mechanism. If the agent is convinced without any doubt that consumption is driven by a particular model $\{a, b_0\}$, her preferences over the innovations ε_t are described by Epstein–Zin (1991) preferences. Her 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 to ensure that we can easily derive analytic results.[8]

[6] We restrict our attention to models with purely linear feedback from past to current consumption growth. We view it as 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. For our purposes, the restriction to the class of linear processes is important mainly as a description of the agent's beliefs, rather than as a restriction on the true process driving consumption growth.

[7] Note that the b_0 in the $\{a, b_0\}$ form is in fact identical to the constant in the associated polynomial b

[8] 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

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

$$v\left(\Delta c^t; a, b_0\right) = (1-\beta)\, c_t + \frac{\beta}{1-\alpha} \log E_t \left[\exp\left(v\left(\Delta c^{t+1}; a, b_0\right)(1-\alpha)\right) | a, b_0\right] \tag{6}$$

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 $\{a, b_0\}$.[9]

It is often convenient to rescale lifetime utility by current consumption, yielding

$$\begin{aligned} vc\left(\Delta c^t; a, b_0\right) &\equiv v\left(\Delta c^t; a, b_0\right) - c_t = \frac{\beta}{1-\alpha} \log E_t \left[\exp\left(\left(vc\left(\Delta c^{t+1}; a, b_0\right) + \Delta c_{t+1}\right)(1-\alpha)\right) | a, b_0\right] & (7) \\ &= \frac{\beta}{1-\beta} \frac{1-\alpha}{2} b\left(\beta\right)^2 + \sum_{k=1}^{\infty} \beta^k E_t\left[\Delta c_{t+k} | a, b_0\right] & (8) \end{aligned}$$

where the second line follows from the assumption that ε is normally distributed.[10] The summation term is what would be obtained if the agent simply had log utility, in which case consumption dynamics would not affect lifetime utility ex ante. The first term is an adjustment for risk. The agent'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 $b\left(\beta\right)^2$, which measures the variance of the shocks to lifetime utility in each period. $b\left(\beta\right)$ measures the total discounted effect of a unit innovation to ε_{t+1} on consumption growth, and hence utility, in the future. It is the $b\left(\beta\right)$ term that causes agents with Epstein–Zin preferences to be averse to long-run risk. They are obviously also averse to short-run shocks, but it is the total long-run effect of the shocks (discounted by β) that determines the risk adjustment in utility.

2.2 Robustness over dynamics

Equation (7) gives lifetime utility in the case where the agent dogmatically believes in a particular model $\{a, b_0\}$ (or, equivalently, b). We now model the agent as considering alternative models of propagation.

The agent 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 worst in an unrestricted sense

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

[10] See the appendix for a full derivation

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

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

$$b^w = \arg\min_b \left\{ E\left[vc\left(\Delta c^t; b\right) | b \right] + \lambda g(b) \right\} \tag{9}$$

b^w is the model that gives the agent the lowest unconditional expected lifetime utility, subject to a penalty associated with the likelihood of the model, $g(b)$. λ is a parameter that determines how much weight the penalty receives.

With the approach that the agent chooses a single worst-case b^w, we obtain a setup similar to Gilboa and Schmeidler (1989) and Epstein and Schneider (2007) in that we are essentially constructing a set of models and minimizing over that set.[11] The choice of b^w is timeless – it is invariant to the time-series evolution of consumption – which allows us to avoid any issues of learning or filtering. In that sense, we do not provide a complete description of dynamic preferences. What b^w represents is what the unconditional worst-case model looks like: 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.[12]

A natural question is why we need to analyze a worst case instead of allowing the agent to average as a Bayesian across all the possible models. One simple answer is that people may not actually be Bayesians, or they may not be able to assign priors to all models. 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. 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 or liable to change.[13] The key difference between

[11] In Barillas, Hansen, and Sargent (2009), agents are unsure of the distribution of the innovations ε. They obtain a similar result that the worst-case distribution of ε is fixed over time, though in their case, the constancy of the distribution of the innovations is endogenous, rather than assumed.

[12] Other papers that study filtering and learning (e.g. Hansen and Sargent, 2010; and Ju and Miao, 2012) rely on numerical analysis and study a narrow range of possible models (in the case of Hansen and Sargent, 2010, the agent selects between parameterized models, while in Ju and Miao, 2012, the agent has uncertainty about a single hidden state variable). While we have a less sophisticated treatment in that we eliminate learning and filtering, the gain is that we obtain analytic solutions and allow our agent to consider a broader range of potential models.

[13] Sgain, see also Hansen and Sargent (2005) for a related strategy.

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.[14] Our model can thus be viewed as a combination of Kreps's (1998) anticipated utility with ambiguity aversion like that of Gilboa and Schmeidler (1989).[15]

3 Non-parametric estimation

$g(b)$ should capture the agent's assessment of the plausibility of a particular model, so it is connected to the method the agent uses to estimate consumption dynamics. For example, if she estimates a parameterized model, such as an AR(1), she will have a likelihood (or posterior distribution) over the autocorrelation, and she can rank different AR(1) processes by how far their autocorrelations are from her point estimate. That method, 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 agent 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. She allows the number of lags in her estimated model to grow with the sample size, potentially allowing eventually for a broad class of dynamics.[16] The agent then has a non-parametric confidence set around any particular point estimate implicitly includes models far more complex than the actual AR or MA model she estimates in any particular sample.[17]

[14] Kreps (1998) also allows the agent's model to change over time as she gains new information. We shut down the learning channel in this paper.

[15] 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.

[16] We use the results from Berk (1974) and Brockwell and Davis (1988) because their method of using models with growing lag lengths also delivers distributions for the lag polynomial $b(L)$. The results for other non-parametric spectral estimators are similar 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 difference is just that factoring the spectrum to obtain the lag polynomial is a complicated and non-analytic process (Priestly (1981)).

[17] So the agent estimates a parametric model but has non-parametric confidence intervals. That is really how all non-parametric estimation works; you estimate a parametric model, e.g. a piecewise linear regression or a histogram, and imagine the number of parameters growing with the sample size asymptotically.

Denote true spectral density as

$$f^{True}(\omega) \equiv b^{True}\left(e^{i\omega}\right) b^{True}\left(e^{-i\omega}\right) \tag{10}$$

where b^{True} is the true model driving consumption. The transfer function for a model, which we will return to repeatedly, is defined as

$$B(\omega) \equiv b\left(e^{i\omega}\right) \tag{11}$$

The transfer function measures how the filter $b(L)$ transfers power at each frequency from the white-noise innovations, ε, to consumption growth. Now assume that consumption growth follows a linear Gaussian process with a spectral density that is finite and bounded away from zero.[18] Berk (1974) and Brockwell and Davis (1988) show that, under those conditions, estimates of the transfer function, $\hat{B}(\omega)$, have the following asymptotic distribution

$$\hat{B}(\omega) \sim CN\left(B^{True}(\omega), f^{True}(\omega)/T^{q}\right) \tag{12}$$

$$cov\left(\hat{B}(\omega_1) - B^{True}(\omega_1), \hat{B}(\omega_2) - B^{True}(\omega_2)\right) = 0 \text{ for } \omega_1 \neq \omega_2 \tag{13}$$

where CN is the complex normal distribution 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). So the variance of the estimate $\hat{B}(\omega)$ is proportional to $f^{True}(\omega)$ and estimation errors are independent across frequencies.[19]

The result that projecting the lag polynomial into the frequency domain orthogonalizes the estimates is a common finding, and similar to the result that periodogram ordinates are asymptotically independent. Because uncertainty is independent across frequencies, the agent's problem of choosing a worst-case model will also be independent across frequencies. That makes analyzing the solution relatively simple and it is why we work in the frequency domain in this paper.

[18] The assumption of normality is not strictly necessary. The results of Berk (1974) and Brockwell and Davis (1988) do not use that restriction. We only need to assume that the agent models innovations as normally distributed in order to obtain simple expressions for lifetime utility. More generally, though, the results go through similarly for other distributions (including ones featuring disasters).

[19] 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 since the worst-case dynamics differ from the benchmark almost purely in their long-run behavior rather than in the innovation variance.

Denote the agent's point estimates of the transfer function and spectrum in a given sample as $\bar{B}$ and $\bar{f}$. Given the complex normal distribution for $\hat{B}(\omega)$, we can construct a penalty function $g(b)$ by integrating a test statistic across frequencies,

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

which is a χ^2-type test statistic for the null hypothesis that $B = \bar{B}$.[20][21] Here and below, integrals without limits denote $(2\pi)^{-1} \int_{-\pi}^{\pi}$.[22]

3.1 Time-domain interpretation

While $g(b)$ will be useful for solving the agent's optimization problem, it is perhaps somewhat difficult to interpret since most time series analysis is performed in the time domain. This section shows that $g(b)$ has a natural time-domain interpretation and also gives an expression for its statistical distribution.

Suppose the agent observes a sample of consumption growth of length T, $\{\Delta c_1, \Delta c_2, ...\Delta c_T\}$. As an alternative to using the frequency domain test in (14), she could test whether the consumption is driven by the model $\bar{b}$ using a time-domain model specification test. Consider filtering the realized consumption data by $\bar{b}(L)^{-1}$ to create a set of filtered innovations,

$$\hat{\varepsilon}_t = \bar{b}(L)^{-1} \Delta c_t \tag{15}$$

If $\bar{b}$ is correctly specified then $\hat{\varepsilon}$ should be white noise. The statistical test of white noise that she uses is a portmanteau test of model fit similar to that of Box and Pierce (1970) and Hong (1996). She estimates an $MA(m)$ model on the filtered sample (using, for example, the estimator from Brockwell and Davis, 1988),

[20] Hong (1996) studies a closely related distance metric (defined on the spectrum instead of the transfer function) in the context of testing for general deviations from a benchmark spectral density and explicitly derives its asymptotic distribution. Yu (2012) considers a similar distance metric in a reduced-form manner.

[21] Intuitively, the asymptotic normal distribution for $B(\omega)$ depends on the assumed smoothness of the true transfer function. So using $g(b)$ as our test statistic implicitly imposes a reasonable prior on the possible transfer functions, which is also what allows the agent to rank transfer functions that are potentially indistinguishable in finite samples. See Kitigawa and Gersch (1985) for a method of directly imposing smoothness priors on the spectral density and transfer function. The drawback of their approach for our purposes is that they do not provide an asymptotic distribution.

[22] 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, even though Berk (1974) and Brockwell and Davis (1988) rule them out by assuming that the spectral density is bounded. 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.

denoting the estimated MA coefficients $\hat{d}_j$, $j \in \{1, ..., m\}$. As $T \to \infty$, $m \to \infty$. The test statistic is

$$\sum_{j=1}^{m} \hat{d}_j^2 \tag{16}$$

Under the null hypothesis that $\bar{b}(L)$ is the true model and $\hat{\varepsilon}_t$ is thus white noise, the true values of the MA coefficients are zero and they are asymptotically normal with variance proportional to T^{-1}.[23] The appendix shows, using the asymptotic results of Brockwell and Davis (1988) that under the null hypothesis, $\sum_{j=1}^{m} \hat{d}_j^2 \sim \chi_m^2$. The appendix also derives the following result for the value of the test statistic if instead the data was generated by the alternative model b:

Result 1 *Suppose a sample of consumption growth of length T is generated with dynamics $b(L)$ and the agent filters it using $\bar{b}(L)^{-1}$. As $T, m \to \infty$, the test statistic $\sum_{j=1}^{m} \hat{d}_j^2$ will be approximately distributed as a non-central χ_m^2 random variable with non-centrality parameter $Tg(b)$, denoted $\chi_m^2(Tg(b))$*

$$\sum_{j=1}^{m} \hat{d}_j^2 \sim \chi_m^2(Tg(b)) \tag{17}$$

where $g(b)$ is the penalty function defined in (14).

So the spectral penalty function $g(b)$ is proportional to the non-centrality parameter in a time-domain specification test. The probability that the agent rejects the null is monotonically increasing in that non-centrality parameter. When $g(b)$ is higher, b is farther from $\bar{b}$ in the sense that the agent is more likely to reject the null that the data was driven by $\bar{b}$ after observing a sample driven by b and using a portmanteau test. So $g(b)$ has intuitive interpretations in both the frequency and time domains.

[23] Under the alternative hypothesis that $\hat{\varepsilon}_t$ is not white noise, specifically that the data was driven by $b(L) \neq \hat{b}(L)$, the estimated coefficients are still asymptotically normal, but they have a non-trivial covariance structure. To avoid the change in the covariance matrix, result 1 is derived using a local (Pitman-type) alternative, where $b(L) - \hat{b}(L)$ is proportional to $T^{-1/2}$ as the sample size grows.

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 (9)

$$b^w(L) = \arg\min_{b(L)} \left\{ -\beta(1-\alpha) b(\beta)^2 + \lambda \int \frac{\left|B(\omega) - \bar{B}(\omega)\right|^2}{\bar{f}(\omega)} d\omega \right\} \tag{18}$$

Equivalently, we can write

$$b^w(L) = \arg\min_{b(L)} -\beta(1-\alpha) \left(\int Z(\omega) B(\omega) d\omega \right)^2 + \lambda \int \frac{\left|B(\omega) - \bar{B}(\omega)\right|^2}{\bar{f}(\omega)} d\omega \tag{19}$$

$$\text{where } Z(\omega) \equiv \sum_{j=0}^{\infty} \beta^j e^{-i\omega j} \tag{20}$$

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. (19) 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

$$B^w(\omega) = \bar{B}(\omega) + \lambda^{-1}\beta(1-\alpha) b(\beta)^w \times \bar{f}(\omega) \times Z^*(\omega) \tag{21}$$

where a * denotes a complex conjugate. A linear solution is to be expected given that the optimization problem is quadratic.

We thus have a closed form expression for the worst-case model, described in the frequency domain. The worst-case lag polynomial $b^w(L)$ is obtained as the inverse Fourier transform of $B^w(\omega)$,

$$b_j^w = \int \exp(-i\omega j) B^w(\omega) d\omega \tag{22}$$

The worst-case transfer function $B^w(\omega)$ in (21) is equal to the true transfer function plus a term that depends on three factors. First, $\lambda^{-1}\beta(1-\alpha) b^w(\beta)$ represents the ratio of the utility losses from a marginal increase in $b^w(\beta)$ to the cost of normalized deviations from the benchmark, λ. Second, $\bar{f}(\omega)$ represents the amount of uncertainty the agent has about consumption dynamics at frequency ω. When $\bar{f}(\omega)$ is high there will be 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 the limit where $\beta \to 1$, $Z(\omega)$ becomes a point mass at zero.[24]

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 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 $Z_r(\omega)$. It is strikingly peaked near frequency zero for any plausible value of the time discount factor β. In other words, the agent's fears are primarily about very low-frequency fluctuations in consumption growth.

It is also straightforward to solve for $b^w(\beta)$, which determines the size of risk premia,

$$b^w(\beta) = \frac{\bar{b}(\beta)}{1 - (\alpha - 1)\lambda^{-1} \int Z(\omega)^* Z(\omega) \bar{f}(\omega) d\omega} \tag{26}$$

Equation (21) represents the completion of the solution to the model. To summarize, the agent first selects a worst-case model $B^w(\omega)$, which is associated with a specific $b^w(L)$. She then uses the worst-case model when calculating expectations and pricing assets (i.e. using the operator $E_t[\cdot|b^w]$).

4.1 Long-run risk is the worst case

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

$$\bar{b}(L) = \bar{b}_0 \tag{27}$$

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

Using the formula for $B^w(\omega)$ from (21) and the inverse Fourier transform, we obtain

$$b_j^w = \varphi\beta^j \text{ for } j > 0 \tag{28}$$

$$b_0^w = \bar{b}_0 + \varphi \tag{29}$$

$$\varphi \equiv \beta(\alpha - 1) b^w(\beta) \bar{b}_0^2 \lambda^{-1} \tag{30}$$

So the worst-case model for consumption growth is an ARMA(1,1), which can be expressed as

$$\Delta c_t = \beta\Delta c_{t-1} + \left(\bar{b}_0 + \varphi\right)\varepsilon_t - \theta\left(\bar{b}_0 + \varphi\right)\varepsilon_{t-1} \tag{31}$$

$$\varepsilon_t \backsim \mathcal{N}(0,1) \tag{32}$$

$$\theta = \beta\left(1 - \varphi\bar{b}_0^{-1}\right) \tag{33}$$

In general, a given ARMA model has many latent factor representations. An intuitive representation here is to define an underlying state variable representing the conditional mean of consumption growth. We can then equivalently express the ARMA process (31) as

$$\Delta c_t = x_{t-1} + \mu_t \tag{34}$$

$$x_t = \beta x_{t-1} + v_t \tag{35}$$

where

$$\mu_t \sim N\left(0, \theta\beta^{-1}\left(\bar{b}_0 + \varphi\right)^2\right) \tag{36}$$

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

and μ_t and v_t are independent. The state-space form in equations (34–37) is observationally equivalent to the ARMA(1,1) 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.[25] 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.

[25] If we allowed the agent to estimate volatility dynamics similarly to how we allow her to estimate consumption dynamics, we would likely obtain heteroskedasticity in the worst-case model. Accounting for stochastic volatility substantially complicates the analysis, though, so we leave it for future work.

The worst-case process exhibits a small but highly persistent trend component, where the degree of 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.[26]

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

5 The behavior of asset prices

Utility under the worst-case model is

$$vc^w\left(\Delta c^t\right) = vc\left(\Delta c^t; b^w\right) = \frac{\beta}{1-\beta}\frac{1-\alpha}{2}b^w(\beta)^2 + \sum_{k=1}^{\infty}\beta^k E_t\left[\Delta c_{t+k}|b^w\right] \tag{38}$$

That is, lifetime utility is calculated using b^w to form expectations about future consumption growth. The Euler equation in this model is then simply the standard Euler equation under Epstein–Zin preferences, but again calculated under the worst-case dynamics. As usual under Epstein–Zin preferences, the household's

[26] 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.

Euler equation for an asset return R_{t+1} is

$$1 = E_t\left[R_{t+1}M_{t+1}|b^w\right] \tag{39}$$

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

where M_{t+1} is the stochastic discount factor. The key implication of the model is that assets are priced under the worst-case model, rather than the true model.

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 return on the levered consumption claim is

$$\begin{aligned} E_t\left[r_{t+1}-r_{f,t+1}|\bar{a},\bar{b}_0\right] &= \frac{\gamma-\delta a^w\left(\delta\right)}{1-\delta a^w\left(\delta\right)}\left(\bar{a}\left(L\right)-a^w\left(L\right)\right)\Delta c_t + \\ &\quad -\frac{1}{2}\left(\left(\gamma-1\right)b^w\left(\delta\right)+b_0^w\right)^2-\left(\left(\gamma-1\right)b^w\left(\delta\right)+b_0^w\right)\left(\left(1-\alpha\right)b^w\left(\beta\right)-b_0^w\right) \quad (41) \\ &= \frac{\gamma-\delta a^w\left(\delta\right)}{1-\delta a^w\left(\delta\right)}\left(\bar{a}\left(L\right)-a^w\left(L\right)\right)\Delta c_t - cov\left(r_{t+1},m_{t+1}\right)-\frac{1}{2}var\left(r_{t+1}\right) \quad (42) \end{aligned}$$

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

The middle line, which is equal to $-cov\left(r_{t+1},m_{t+1}\right)-\frac{1}{2}var\left(r_{t+1}\right)$, 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 (through a high $b^w\left(\beta\right)$), they have large effects on lifetime utility, thus making the SDF very volatile. Second, again because of the persistence of consumption growth under the worst-case, shocks to consumption have large effects on expected long-run dividend growth, so the return on the consumption claim is also very sensitive to shocks. 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 $\frac{\gamma-\delta a^w(\delta)}{1-\delta a^w(\delta)}\left(\bar{a}\left(L\right)-a^w\left(L\right)\right)\Delta c_t$, 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 consumption growth 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,

$$r_{f,t+1}=-\log\beta+a^w\left(L\right)\Delta c_t-\frac{1}{2}\left(b_0^w\right)^2+b_0^w b^w\left(\beta\right)\left(1-\alpha\right) \tag{43}$$

Under log utility, interest rates move one for one with expected consumption growth. In the present model, the relevant measure of expected consumption growth is $a^w\left(L\right)\Delta c_t$, which is the expectation under the worst-case model.

6 Calibration

We now parameterize the model to analyze its quantitative implications. 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 $\delta=0.95^{1/4}$. The models are all calibrated so that consumption growth has a quarterly standard deviation of 1 percent under the true model, which we assume is also the agent's point estimate. Finally, the leverage parameter for the consumption claim, γ, is set to 5.62 to generate mean annualized equity returns of 7 percent.

We now show that α can be linked to the parameter λ that controls the plausibility of the models the agent considers. 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.

The appendix shows that when α is interpreted as constraining the worst-case distribution of ε, we can directly link it to λ through the formula

$$\alpha = 1 + \frac{1}{2\lambda} \tag{44}$$

We calibrate λ to equal 0.0785 to match the observed Sharpe ratio on equities. Formula (44) then implies α should equal 5.31. That level of risk aversion is extraordinarily small in the context of the consumption-based asset pricing literature. 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 (31) in table 1. The model is calibrated so that the point estimate for the standard deviation of quarterly consumption growth is 1 percent. 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). Furthermore, in their calibration, the trend component of consumption growth accounts for 4.5 percent of the variance of consumption growth,

while in our calibration, it accounts for 7.9 percent. The long-run risk in the worst-case model is thus far larger than in Bansal and Yaron's calibration, both in terms of persistence and in terms of the fraction of consumption growth accounted for by the trend.

Secondly, note 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 3.9 times higher than $\bar{b}(\beta)$, thus implying that the worst-case economy is far riskier than the benchmark.

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 7 and 19.7 percent respectively, nearly identical to the data. The real risk-free rate is low and stable, with a mean and standard deviation of 0.31 percentage points.

The second column in the bottom section of table 1 shows that would happen if we set $\lambda = \infty$ (but held α fixed at 5.31), 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 7 to 1.6 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 19.6 to 9.5 percent and the standard deviation of the price/dividend ratio falls from 20 percent to 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 0.94 nearly exactly matches the empirical autocorrelation, which is 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, again, 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 grey 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 four-year horizon is 35 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 EIS regressions 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, R^2s from forecasting regressions, generate low and stable interest rates, generate substantial volatility in the price/dividend ratio (though still less than what we observe empirically), and 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 subsection is white noise.

7.1.2 Probability of rejecting the worst-case dynamics

An obvious question is how difficult it would be for an agent to believe that the worst-case model is actually what generated a particular sample. That is, if the agent's null hypothesis is that the worst-case model generated the data, what is the probability that she would reject that hypothesis for a given sample of data generated by the benchmark model? All of the successes we have in matching asset pricing moments would be less notable if they relied on agents believing in a model of consumption growth that is massively at odds with the data. This section shows that the worst-case model is only weakly distinguishable from the true model.

We consider two tests of the fit of the worst-case model to data generated under the benchmark: 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 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. 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 4.6 and 4.9 percent (*lower* than the nominal test size) in the 50- and 100-year samples, respectively, while the ARMA(1,1) likelihood ratio test performs only slightly better, with rates of 5.4 and 6.3 percent.[27] Table 1 thus shows that the worst-case model, while having economically large differences from the benchmark

[27] The low rejection rates likely comes from the fact that under the alternative hypothesis, consumption growth is negatively autocorrelated, which causes the test statistic to have a low variance, which decreases the frequency of rejection.

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 agent 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 agent considers 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 An AR(2) benchmark

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 4. The density is maximized at a frequency implying periodicity of approximately eight quarters. Thus the term $\bar{f}(\omega)$ in (21) 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 (18) if she were constrained to minimize utility with respect to a transfer function implied by an AR(2). That is, the problem is solved by choosing a worst-case $\{a_1, a_2, b_0\}$ in the model

$$\Delta c_t = a_1 \Delta c_{t-1} + a_2 \Delta c_{t-2} + b_0 \varepsilon_t \tag{45}$$

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.[28] 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

[28] See Collin-Dufresne, Johannes and Lochstoer (2013) for a deep analysis of the case of parameter uncertainty with Epstein–Zin preferences.

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 5; we refer to it as the parametric worst case. The parametric worst case is essentially indistinguishable from the point estimate – it is the grey 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.

The specific parameters in the benchmark and parametric worst-case models are

$$\text{Benchmark:} \qquad \Delta c_t = (0.70) \times \Delta c_{t-1} + (-0.35) \times \Delta c_{t-2} + (0.0051) \times \varepsilon_t \tag{46}$$

$$\text{Parametric worst case:} \qquad \Delta c_t = (0.69) \times \Delta c_{t-1} + (-0.34)\, \Delta c_{t-2} + (0.0053) \times \varepsilon_t \tag{47}$$

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.

When we allow the agent to solve the more general problem (18) in an unrestricted manner, allowing for arbitrary dynamics and giving rise to the conditions in equation (21) the worst-case is very similar to what we obtained for the white-noise benchmark, as shown in figure 5. 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).

The qualitative difference in the shape of the benchmark and worst-case transfer functions can be reinterpreted by considering the MA representation of each process. In figure 6 we display, up to a truncation point, these MA coefficients. Several points that are worth noting. First, we see the coefficients associated with the benchmark die out rather quickly, whereas the worst-case coefficients remain noticeably

positive, if small, for very many lags, which, of course, underpins the considerable power observed at low frequencies. Second, although it is not entirely obvious from the diagram, the rate of decay of the worst-case MA coefficients from around lag 17 onwards is $\frac{b_{j+1}}{b_j} \approx 0.994 = \beta$. That is, the rate of decline is apparently equal to the rate of time preference, which is also what we obtain in the white-noise case. So we observe, after a small number of lags, properties very similar to those exhibited under the worst case that emerged from our tractable i.i.d. benchmark example in section 7.1.

7.3 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 7 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 (2008) (and extended by us to 2012), while the dividends also come from Shiller. 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 (1999) (but without some of the non-linearities they included).

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 4 shows that our simple model with white-noise consumption growth performs well in matching price/dividend ratios, at least up to the 1990s.

8 Conclusion

This paper studies asset pricing when agents are unsure about the endowment process. The fundamental insight of the paper is that 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 for consumption dynamics 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 a consumer 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

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. Macroeconomic crises since 1870. NBER Working Papers 13940, National Bureau of Economic Research, Inc, April 2008. URL http://ideas.repec.org/p/nbr/nberwo/13940.html.

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.

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.

G. E. P. Box and David A. Pierce. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332):1509–1526, 1970. doi: 10.1080/01621459.1970.10481180. URL http://www.tandfonline.com/doi/abs/10.1080/01621459.1970.10481180.

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

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

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, 1999.

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

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(4): 1275–1303, 2007. URL http://ideas.repec.org/a/oup/restud/v74y2007i4p1275-1303.html.

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.

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):pp. 837–864, 1996. ISSN 00129682. URL http://www.jstor.org/stable/2171847.

Nengjiu Ju and Jianjun Miao. Ambiguity, learning, and asset returns. *Econometrica*, 80(2):559–591, 2012. ISSN 1468-0262. doi: 10.3982/ECTA7618. URL http://dx.doi.org/10.3982/ECTA7618.

Genshiro Kitigawa and Will Gersch. A smoothness priors long ar model method for spectral estimation. *IEEE Transactions on Automatic Control*, AC-30(1):57–65, 1985.

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.

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`.

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

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.

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.

Xiangrong Yu. A note on the formulation and solution concepts of robustness games. Working Paper, University of Wisconsin - Madison, 2012.

A 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

$$\varepsilon_t^{\{a,b_0\}} = \left(\Delta c_t - a\left(L\right)\Delta c_{t-1}\right)/b_0 \tag{48}$$

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

$$E_t\left[\Delta c_{t+j}|a,b_0\right] = \sum_{j=0}^{\infty} b_{k+j}\varepsilon_{t-j}^{\{a,b_0\}} \tag{49}$$

with subjective distribution

$$\frac{\Delta c_{t+1} - E_t\left[\Delta c_{t+1}|a,b_0\right]}{b_0} \sim N\left(0,1\right) \tag{50}$$

We guess that vc takes the form

$$vc\left(\Delta c^t;a,b_0\right) = \bar{k} + \sum_{j=0}^{\infty} k_j \varepsilon_{t-j}^{\{a,b_0\}} \tag{51}$$

Inserting into the recursion for lifetime utility yields

$$\bar{k} + \sum_{j=0}^{\infty} k_j \varepsilon_{t-j}^{\{a,b_0\}} = \frac{\beta}{1-\alpha}\log E_t\left[\exp\left(\left(\bar{k} + \sum_{j=0}^{\infty}\left(k_j + b_j\right)\varepsilon_{t-j+1}^{\{a,b_0\}}\right)\left(1-\alpha\right)\right)|a,b_0\right] \tag{52}$$

$$= \beta\bar{k} + \beta\sum_{j=0}^{\infty}\left(k_{j+1} + b_{j+1}\right)\varepsilon_{t-j}^{\{a,b_0\}} + \beta\frac{1-\alpha}{2}\left(k_0 + b_0\right)^2 \tag{53}$$

By matching coefficients we obtain

$$\begin{aligned} vc\left(\Delta c^t; b\right) &= \frac{\beta}{1-\beta}\frac{1-\alpha}{2} b\left(\beta\right)^2 \sigma_\varepsilon^2 + \sum_{k=1}^{\infty}\beta^k \sum_{j=0}^{\infty} b_{j+k}\varepsilon_{t-j}^{\{a,b_0\}} && (54)\\ &= \frac{\beta}{1-\beta}\frac{1-\alpha}{2} b\left(\beta\right)^2 \sigma_\varepsilon^2 + \sum_{j=0}^{\infty}\left(\sum_{k=1}^{\infty}\beta^k b_{j+k}\right)\varepsilon_{t-j}^{\{a,b_0\}} && (55)\\ &= \frac{\beta}{1-\beta}\frac{1-\alpha}{2} b\left(\beta\right)^2 \sigma_\varepsilon^2 + \sum_{k=1}^{\infty}\beta^k E_t\left[\Delta c_{t+k}|a,b_0\right] && (56)\end{aligned}$$

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

The goal is to estimate the transfer function $B\left(\omega\right)$. 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 $B\left(\omega\right)$ will be normally distributed around its true value. Specifically,

$$\begin{aligned} A\left(\omega\right) &\equiv B\left(\omega\right)^{-1} && (57)\\ A\left(\omega\right) &= a_r\left(\omega\right) + i a_i\left(\omega\right) && (58)\\ \begin{bmatrix} a_r\left(\omega\right) \\ a_i\left(\omega\right) \end{bmatrix} &\sim N\left(\begin{bmatrix} a_r^{True}\left(\omega\right) \\ a_i^{True}\left(\omega\right) \end{bmatrix}, I_2 f^{True}\left(\omega\right)^{-1}\right) && (59)\end{aligned}$$

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

Now we have

$$\begin{aligned} B\left(\omega\right) &= A^{-1}\left(\omega\right) && (60)\\ &= \frac{A\left(\omega\right)^*}{A\left(\omega\right)A\left(\omega\right)^*} && (61)\\ &= \frac{a_r\left(\omega\right) - i a_i\left(\omega\right)}{a_r\left(\omega\right)^2 + a_i\left(\omega\right)^2} && (62)\end{aligned}$$

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

for now. We have

$$B = b_r + ib_i \tag{63}$$

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

The derivatives with respect to a_i and a_r are

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

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

$$\frac{db_i}{da_r} = \frac{2a_i a_r}{\left(a_r^2 + a_i^2\right)^2} \tag{66}$$

$$\frac{db_i}{da_i} = \frac{a_i^2 - a_r^2}{\left(a_r^2 + a_i^2\right)^2} \tag{67}$$

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

$$\begin{bmatrix} \left(\frac{a_i^2-a_r^2}{\left(a_r^2+a_i^2\right)^2}\right)^2 + \left(\frac{-2a_r a_i}{\left(a_r^2+a_i^2\right)^2}\right)^2 & \frac{-2a_r a_i}{\left(a_r^2+a_i^2\right)^2}\frac{a_i^2-a_r^2}{\left(a_r^2+a_i^2\right)^2} + \frac{2a_i a_r}{\left(a_r^2+a_i^2\right)^2}\frac{a_i^2-a_r^2}{\left(a_r^2+a_i^2\right)^2} \\ \frac{-2a_r a_i}{\left(a_r^2+a_i^2\right)^2}\frac{a_i^2-a_r^2}{\left(a_r^2+a_i^2\right)^2} + \frac{2a_i a_r}{\left(a_r^2+a_i^2\right)^2}\frac{a_i^2-a_r^2}{\left(a_r^2+a_i^2\right)^2} & \left(\frac{a_i^2-a_r^2}{\left(a_r^2+a_i^2\right)^2}\right)^2 + \left(\frac{2a_r a_i}{\left(a_r^2+a_i^2\right)^2}\right)^2 \end{bmatrix} f(\omega)^{-1} \tag{68}$$

$$= \begin{bmatrix} \frac{1}{\left(a_r^2+a_i^2\right)} & 0 \\ 0 & \frac{1}{\left(a_r^2+a_i^2\right)} \end{bmatrix} = I_2 f(\omega) \tag{69}$$

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, the log likelihood for a particular $B(\omega)$ is

$$\frac{\left(B(\omega) - B^{True}(\omega)\right)\left(B(\omega)^* - B^{True}(\omega)^*\right)}{f^{True}(\omega)} \tag{70}$$

C Solving for the worst-case transfer function

The optimization problem is

$$B^w(\omega) = \arg\min_{b(L)} -\beta(1-\alpha)\left(\sum_{j=0}^{\infty} b_j\beta^j\right)^2 + \lambda\int \frac{\left|B(\omega)-\bar{B}(\omega)\right|^2}{\bar{f}(\omega)}d\omega \tag{71}$$

$$= \arg\min_{b(L)} -\beta(1-\alpha)\left(\sum_{j=0}^{\infty} b_j\beta^j\right)^2 + \lambda\int \frac{\left(\sum_{j=0}^{\infty}\exp(i\omega j)\left(b_j-\bar{b}_j\right)\right)\left(\sum_{j=0}^{\infty}\exp(-i\omega j)\left(b_j-\bar{b}_j\right)\right)}{\bar{f}(\omega)} \tag{72}$$

We guess that

$$B^w(\omega) = \bar{B}(\omega) + k\bar{f}(\omega)Z(\omega)^* \tag{73}$$

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

$$0 = -2\beta(1-\alpha)b^w(\beta)\beta^j + \lambda\int \frac{\exp(i\omega j)\left(B(\omega)-\bar{B}(\omega)\right)^* + \exp(-i\omega j)\left(B(\omega)-\bar{B}(\omega)\right)}{\bar{f}(\omega)}d\omega \tag{74}$$

$$= -2\beta(1-\alpha)b^w(\beta)\beta^j + \lambda\int \frac{\exp(i\omega j)k\bar{f}(\omega)Z(\omega) + \exp(-i\omega j)k\bar{f}(\omega)Z(\omega)^*}{\bar{f}(\omega)}d\omega \tag{75}$$

$$= -2\beta(1-\alpha)b^w(\beta)\beta^j + 2\lambda k\beta^j \tag{76}$$

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

$$k = \lambda^{-1}\beta(1-\alpha)b^w(\beta) \tag{77}$$

which is the result from the text.

D 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)

$$r_{t+1} = \delta pd_{t+1} + \gamma\Delta c_{t+1} - pd_t \tag{78}$$

where δ is a linearization parameter slightly less than 1.

We guess that

$$pd_t = \bar{k} + \sum_{j=0}^{\infty} k_j \Delta c_{t-j} \tag{79}$$

Using the formula for lifetime utility from the text, we have

$$\begin{aligned} v_{t+1} - E_t v_{t+1} &= \sum_{k=0}^{\infty} \beta^k \Delta E_{t+1} \left[\Delta c_{t+k} | b^w \right] && (80) \\ &= b^w(\beta) \varepsilon_{t+1} && (81) \end{aligned}$$

So the pricing kernel can be written as

$$M_{t+1} = \beta \exp\left(-a^w(L) \Delta c_t - b_w^0 \varepsilon_{t+1} + (1-\alpha) b^w(\beta) \varepsilon_{t+1} - \frac{(1-\alpha)^2}{2} b^w(\beta)^2 \right) \tag{82}$$

The pricing equation for the levered consumption claim is

$$\begin{aligned} 0 &= \log E_t \left[\beta \exp\left(\begin{array}{c} \delta_0 + (\delta - 1)\bar{k} + (\delta k_0 + \gamma - 1)\Delta c_{t+1} + \sum_{j=0}^{\infty} (\delta k_{j+1} - k_j) \Delta c_{t-j} \\ + (1-\alpha) b^w(\beta) \varepsilon_{t+1} - \frac{(1-\alpha)^2}{2} b^w(\beta)^2 \end{array} \right) \right] && (83) \\ &= (\delta k_0 + \gamma - 1) a^w(L) \Delta c_t + \sum_{j=0}^{\infty} (\delta k_{j+1} - k_j) \Delta c_{t-j} && (84) \\ &\quad + \delta_0 + \left(\frac{1}{2} (\delta k_0 + \gamma - 1)^2 (b_0^w)^2 + (\delta k_0 + \gamma - 1)(1-\alpha) b^w(\beta) b_0^w \right) + (\delta - 1)\bar{k} + \log \beta && (85) \end{aligned}$$

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

$$\begin{aligned} (\delta - 1)\bar{k} + \log \beta &= -\delta_0 - \left(\frac{1}{2} (\delta k_0 + \gamma - 1)^2 (b_0^w)^2 + (\delta k_0 + \gamma - 1)(1-\alpha) b^w(\beta) b_0^w \right) && (86) \\ (\delta k_{j+1} - k_j) &= -(\delta k_0 + \gamma - 1) a_j^w && (87) \end{aligned}$$

And thus

$$k_0 = \frac{(\gamma - 1) a^w(\delta)}{1 - \delta a^w(\delta)}$$

$$
\begin{aligned}
(\delta k_{j+1} - k_j) &= -(\delta k_0 + \gamma - 1)\, a_j^w && (88)\\
k_j &= (\delta k_0 + \gamma - 1)\, a_j^w + \delta k_{j+1} && (89)\\
&= \sum_{k=0}^{\infty} \delta^k (\delta k_0 + \gamma - 1)\, a_{j+k}^w && (90)
\end{aligned}
$$

For the risk-free rate, we have

$$
\begin{aligned}
r_{f,t+1} &= -\log E_t \left[\beta \exp\left(-a^w(L)\,\Delta c_t - b_0^w \varepsilon_{t+1} + (1-\alpha)\, b^w(\beta)\, \varepsilon_{t+1} - \frac{(1-\alpha)^2}{2} b^w(\beta)^2 \right) \right] && (91)\\
&= -\log\beta + a^w(L)\,\Delta c_t - \frac{1}{2}(b_0^w)^2 + b_0^w (1-\alpha)\, b^w(\beta) && (92)
\end{aligned}
$$

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

$$
\begin{aligned}
E_t\left[r_{t+1} - r_{f,t+1} | \bar{b}\right] &= E_t \left[\begin{array}{c} \delta_0 + (\delta - 1)\,\bar{k} + (\delta k_0 + \gamma)\,\Delta c_{t+1} - \sum_{j=0}^{\infty} (\delta k_0 + \gamma - 1)\, a_j^w \Delta c_{t-j} \\ + \log\beta - a^w(L)\,\Delta c_t + \frac{1}{2}(b_0^w)^2 - b_0^w (1-\alpha)\, b^w(\beta) \end{array} | \bar{b} \right] && (93)\\
&= (\delta k_0 + \gamma)\,(a(L) - a^w(L))\,\Delta c_t + \frac{1}{2}(b_0^w)^2 + \delta_0 + (\delta - 1)\,\bar{k} + \log\beta - b_0^w (1-\alpha)\, b^w(\beta) && (94)\\
&= (\delta k_0 + \gamma)\,(a(L) - a^w(L))\,\Delta c_t && (95)\\
&\quad -\frac{1}{2}\left((\gamma - 1)\, b^w(\delta) + b_0^w\right)^2 - \left((\gamma - 1)\, b^w(\delta) + b_0^w\right)\left((1-\alpha)\, b^w(\beta) - b_0^w\right) && (96)\\
&= (\delta k_0 + \gamma)\,(a(L) - a^w(L))\,\Delta c_t - cov\left(m_{t+1}, r_{t+1}\right) - \frac{1}{2} var\left(r_{t+1}\right) && (97)
\end{aligned}
$$

The standard deviation of the SDF is

$$
b_w^0 + (\alpha - 1)\, b^w(\beta) \tag{98}
$$

The standard deviation of the return is

$$
(\delta k_0 + \gamma)\, b_0^w = (\gamma - 1)\, b^w(\delta) + b_0^w \tag{99}
$$

D.1 Results for the white-noise benchmark

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

$$\Delta c_t = \beta \Delta c_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1} \tag{100}$$

$$a^w(L) = (\beta - \theta) \sum_{j=0}^{\infty} \theta^j L^j \tag{101}$$

$$a^w(\delta) = \frac{\beta - \theta}{1 - \theta\delta} \tag{102}$$

$$b_j = \beta^j (\beta - \theta) \tag{103}$$

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

$$\frac{(\delta k_0 + \gamma - 1)(\beta - \theta)}{1 - \delta\theta} \frac{\sigma_{\Delta c}}{\sqrt{1 - \theta^2}} \tag{104}$$

E Test statistics and entropy constraints

E.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

$$v^{BHS}\left(\Delta c^{t};a,b_{0}\right)=c_{t}+\min_{\tilde{h}(\varepsilon_{t+1})}\left\{\begin{array}{c}\beta E_{t}\left[\frac{\tilde{h}(\varepsilon_{t+1})}{h(\varepsilon_{t+1})}\exp\left(\left(v^{BHS}\left(\Delta c^{t+1};a,b_{0}\right)\right)\right)|a,b_{0}\right]\\+\frac{\beta}{(\alpha-1)(1-\beta)}E_{t}\left[\frac{\tilde{h}(\varepsilon_{t+1})}{h(\varepsilon_{t+1})}\log\frac{\tilde{h}(\varepsilon_{t+1})}{h(\varepsilon_{t+1})}\right]\end{array}\right\} \tag{105}$$

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

$$KL\left(h,\tilde{h}\right)\equiv E_{t}\left[\frac{\tilde{h}\left(\varepsilon_{t+1}\right)}{h\left(\varepsilon_{t+1}\right)}\log\frac{\tilde{h}\left(\varepsilon_{t+1}\right)}{h\left(\varepsilon_{t+1}\right)}\right] \tag{106}$$

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 (6). 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

$$\varepsilon_{t} \sim {}_{h}N\left(0,\sigma^{2}\right) \tag{107}$$

$$\varepsilon_{t} \sim {}_{\tilde{h}}N\left(\mu,\sigma^{2}\right) \tag{108}$$

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

$$\varepsilon_{i} \sim {}_{h}N\left(0,\frac{T}{m}\sigma^{2}\right) \tag{109}$$

$$\varepsilon_{i} \sim {}_{\tilde{h}}N\left(\frac{T}{m}\mu,\frac{T}{m}\sigma^{2}\right) \tag{110}$$

So then

$$\sum_{i=1}^{m}\left(\frac{\varepsilon_{i}}{\left(\frac{T}{m}\right)^{1/2}\sigma}\right)^{2} \sim {}_{h}\chi_{m}^{2}\left(0\right) \tag{111}$$

$$\sum_{i=1}^{m}\left(\frac{\varepsilon_{i}}{\left(\frac{T}{m}\right)^{1/2}\sigma}\right)^{2} \sim {}_{\tilde{h}}\chi_{m}^{2}\left(m\left(\frac{\frac{T}{m}\mu}{\left(\frac{T}{m}\right)^{1/2}\sigma}\right)^{2}\right)=\chi_{m}^{2}\left(T\frac{\mu^{2}}{\sigma^{2}}\right) \tag{112}$$

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^2_m with non-centrality parameter $T\frac{\mu^2}{\sigma^2}$. Note that in both cases, 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

$$\begin{aligned}\theta KL\left(h,\tilde{h}\right) &= \theta\frac{1}{2}\frac{\mu^2}{\sigma^2} &(113)\\ &= \theta\frac{1}{T}\frac{1}{2}\left(T\frac{\mu^2}{\sigma^2}\right) &(114)\end{aligned}$$

So the multiplier on the non-centrality parameter is $\theta\frac{1}{T}\frac{1}{2}$.

E.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 $b(L)$,

$$x_t = b(L)\,\varepsilon_t \tag{115}$$

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

$$\frac{1}{\bar{b}(L)}x_t = \frac{b(L)}{\bar{b}(L)}\varepsilon_t \tag{116}$$

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

$$\frac{1}{m}\sum_{j=1}^{m}\hat{d}_j^2 \tag{117}$$

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),

$$T^{1/2}\hat{d}_j \Rightarrow N(0,1) \tag{118}$$

and thus

$$\frac{1}{m}\sum_{j=1}^{m}T\hat{d}_j^2 \Rightarrow \chi_m^2 \tag{119}$$

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),

$$T^{1/2}\left(\hat{\mathbf{d}}^T - \bar{\mathbf{d}}^T\right) \Rightarrow N(0, \Sigma_T) \tag{120}$$

$$T^{1/2}\hat{\mathbf{d}}^T \Rightarrow N\left(T^{1/2}\bar{\mathbf{d}}^T, \Sigma_T\right) = N(\gamma, \Sigma_T) \tag{121}$$

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

$$\begin{aligned} cov\left(x_t, x_{t-j}\right) &= \sum_{k=0}^{\infty} \bar{d}_k^T \bar{d}_{k+j}^T & (122) \\ &= \gamma_j T^{-1/2} + T^{-1} \sum_{k=1}^{\infty} \gamma_k \gamma_{k+j} & (123) \end{aligned}$$

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

$$\begin{aligned} cov\left(x_t, x_t\right) &= 1 + \sum_{k=1}^{\infty} \left(\bar{d}_k^T\right)^2 & (124) \\ &= 1 + T^{-1} \sum_{k=1}^{\infty} \gamma_k^2 & (125) \end{aligned}$$

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

$$\Sigma_T = I + T^{-1}\Omega_1 + T^{-1/2}\Omega_2 \quad (126)$$

We then have

$$\begin{aligned} T^{1/2}\hat{\mathbf{d}}^T &\Rightarrow N\left(T^{1/2}\bar{\mathbf{d}}^T, I + \Omega_1 T^{-1} + \Omega_2 T^{-1/2}\right) & (127) \\ T^{1/2}\left(\hat{\mathbf{d}}^T - \bar{\mathbf{d}}^T\right) &\Rightarrow \varepsilon_I + \varepsilon_{\Omega_1} T^{-1/2} + \varepsilon_{\Omega_2} T^{-1/4} & (128) \end{aligned}$$

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

$$\begin{aligned} T\left(\hat{\mathbf{d}}^T - \bar{\mathbf{d}}^T\right)\left(\hat{\mathbf{d}}^T - \bar{\mathbf{d}}^T\right)' &= \varepsilon_I \varepsilon_I' + \varepsilon_{\Omega_1}\varepsilon_{\Omega_1}' T^{-1} + \varepsilon_{\Omega_2}\varepsilon_{\Omega_2}' T^{-1/2} & (129) \\ &\quad + 2\varepsilon_I \varepsilon_{\Omega_1}' T^{-1/2} + 2\varepsilon_I \varepsilon_{\Omega_2}' T^{-1/4} + 2\varepsilon_{\Omega_1}\varepsilon_{\Omega_2}' T^{-3/4} & (130) \end{aligned}$$

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

$$\left(T^{1/2}\hat{\mathbf{d}}^T - T^{1/2}\bar{\mathbf{d}}^T\right)\left(T^{1/2}\hat{\mathbf{d}}^T - T^{1/2}\bar{\mathbf{d}}^T\right)' \approx \varepsilon_I\varepsilon_I' = \chi^2_m(0) \tag{131}$$

$$\left(T^{1/2}\hat{\mathbf{d}}^T\right)\left(T^{1/2}\hat{\mathbf{d}}^T\right)' \approx \chi^2_m\left(T\bar{\mathbf{d}}^T\bar{\mathbf{d}}^{T\prime}\right) \tag{132}$$

That is, the test statistic is a non-central χ^2_m, 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

$$\int \left|\bar{D}^T(\omega) - 1\right|^2 d\omega = \int \left|\bar{D}^T(\omega)\right|^2 - \bar{D}^T(\omega) - \bar{D}^T(\omega)^* + 1 d\omega \tag{133}$$

$$= \int \left|\bar{D}^T(\omega)\right|^2 - 1 d\omega \tag{134}$$

$$= \sum_{j=1}^{\infty}\left(\bar{d}_j^T\right)^2 = \bar{\mathbf{d}}^T\bar{\mathbf{d}}^{T\prime} \tag{135}$$

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^2_m 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,

$$\int \left|\bar{D}^T(\omega) - 1\right|^2 d\omega = \int \left|\frac{b(L)}{\bar{b}(L)} - 1\right|^2 d\omega \tag{136}$$

$$= \int \left|\frac{b(L) - \bar{b}(L)}{\bar{b}(L)}\right|^2 d\omega \tag{137}$$

$$= \int \frac{\left|b(L) - \bar{b}(L)\right|^2}{\bar{f}(L)} d\omega \tag{138}$$

So we have that

$$g(b) = \sum_{j=1}^{\infty}\left(\bar{d}_j^T\right)^2 = \bar{\mathbf{d}}^T\bar{\mathbf{d}}^{T\prime} \tag{139}$$

$$\lambda g(b) = \lambda\bar{\mathbf{d}}^T\bar{\mathbf{d}}^{T\prime} \tag{140}$$

$$= \frac{\lambda}{T}\left(T\bar{\mathbf{d}}^T\bar{\mathbf{d}}^{T\prime}\right) \tag{141}$$

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

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

where the $1/(1-\beta)$ term on the left-hand side accounts for the fact that the penalty θ is paid every period while λ is paid only once. Solving for α yields

$$1 + \frac{1}{2\lambda} = \alpha \tag{142}$$

E.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^2_m tests. We can then think of the agent's distance penalty for both being in terms of the same units. To equate them, the appendix shows that we must set

$$\alpha = 1 + \frac{1}{2\lambda} \tag{143}$$

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.[29]

Equation 143 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

[29] 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 $g(b)$ 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.

(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.

Figure 1. Weighting function Z for varying values of β

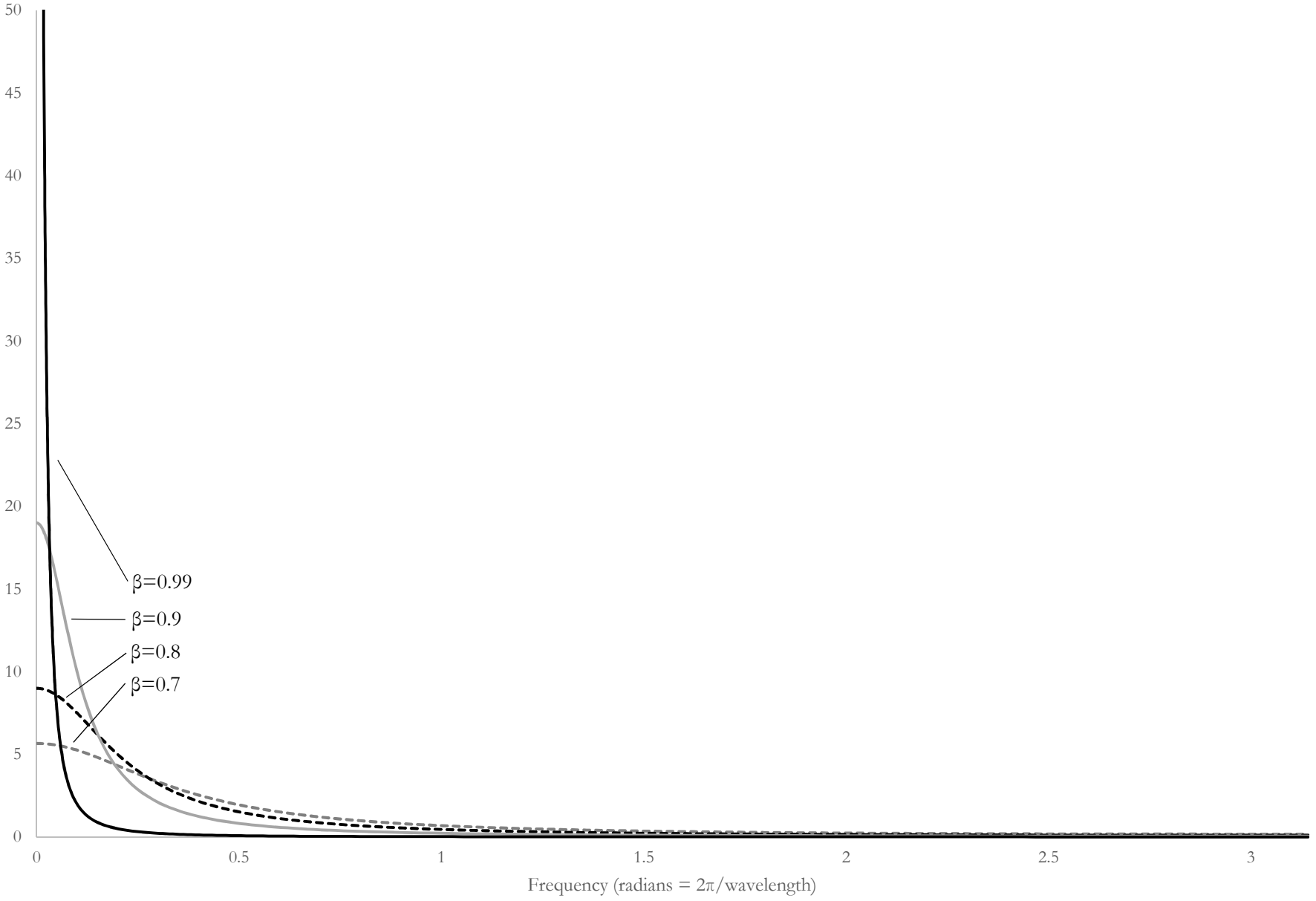


Figure 2. Transfer function under benchmark and worst case for white-noise benchmark

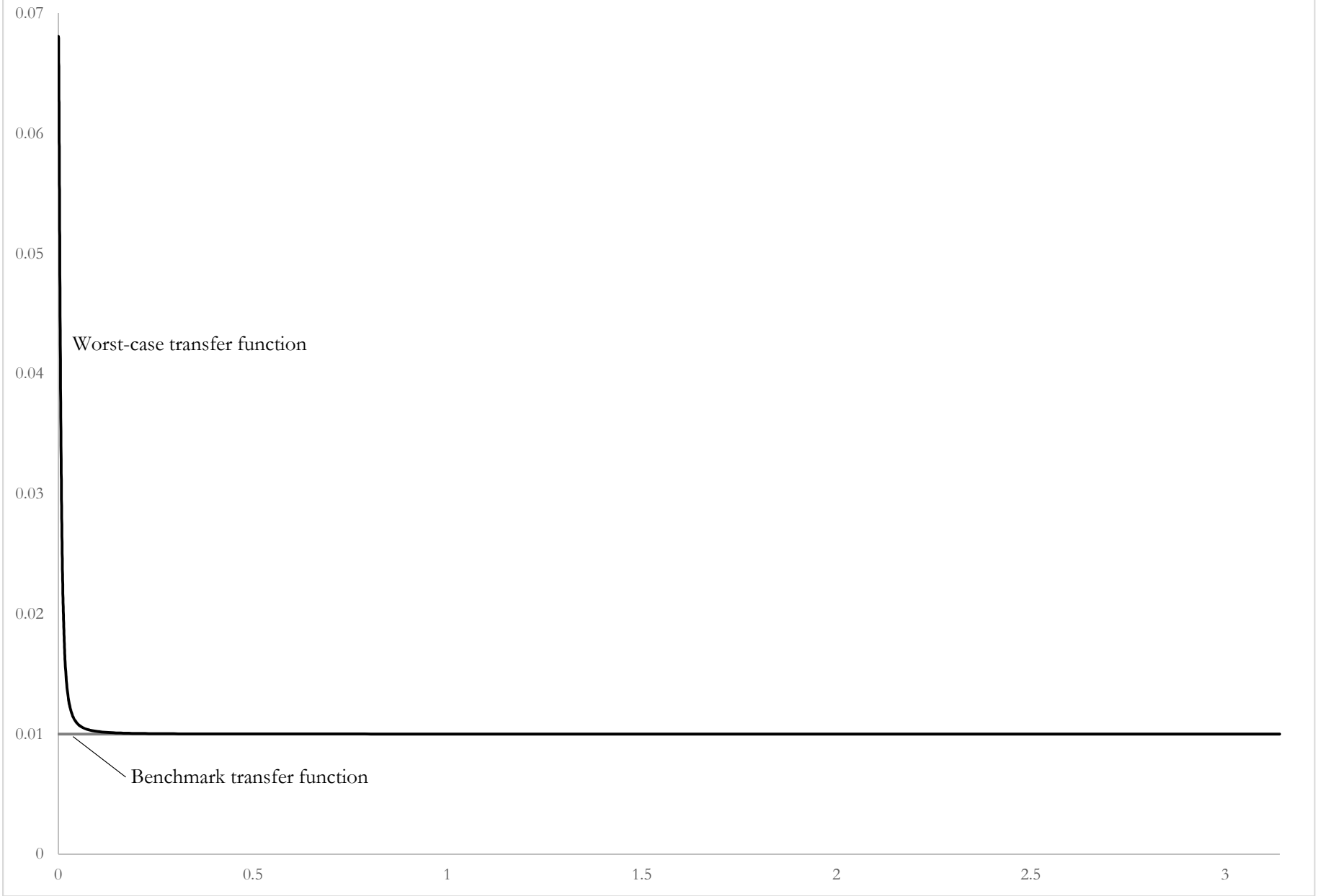
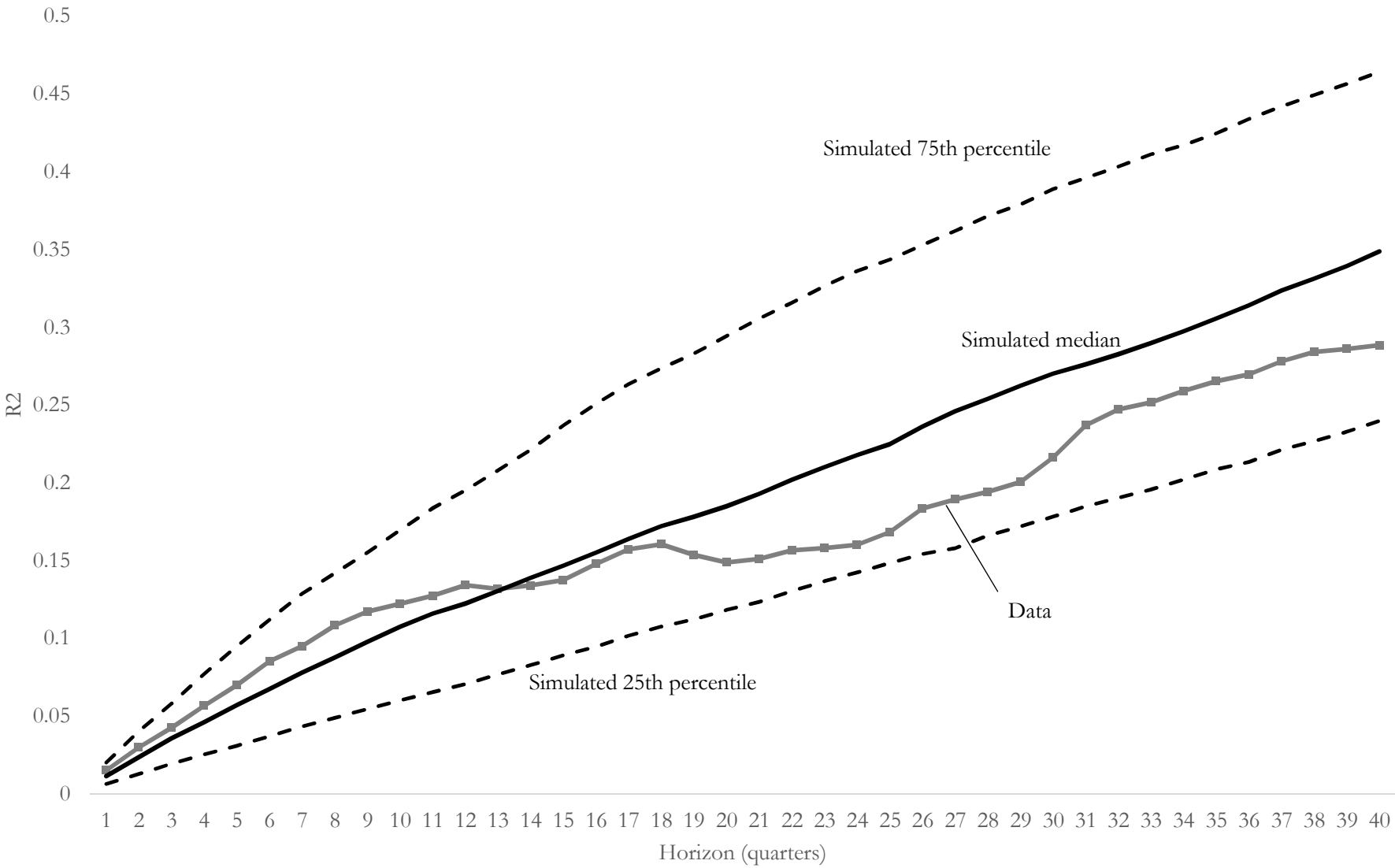


Figure 3. Empirical and model-implied R2's from return forecasting regressions



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

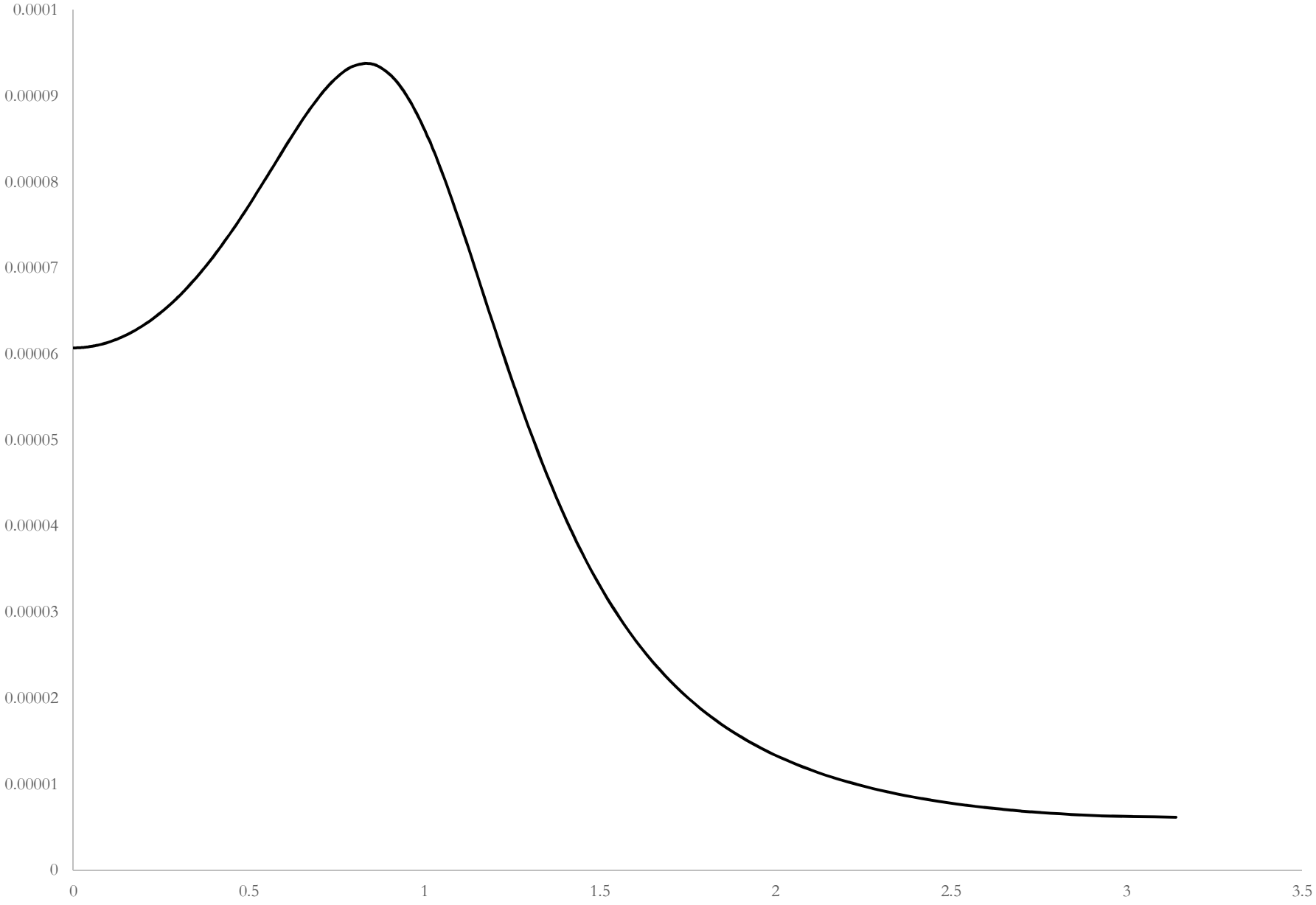


Figure 5. Benchmark and worst-case transfer function

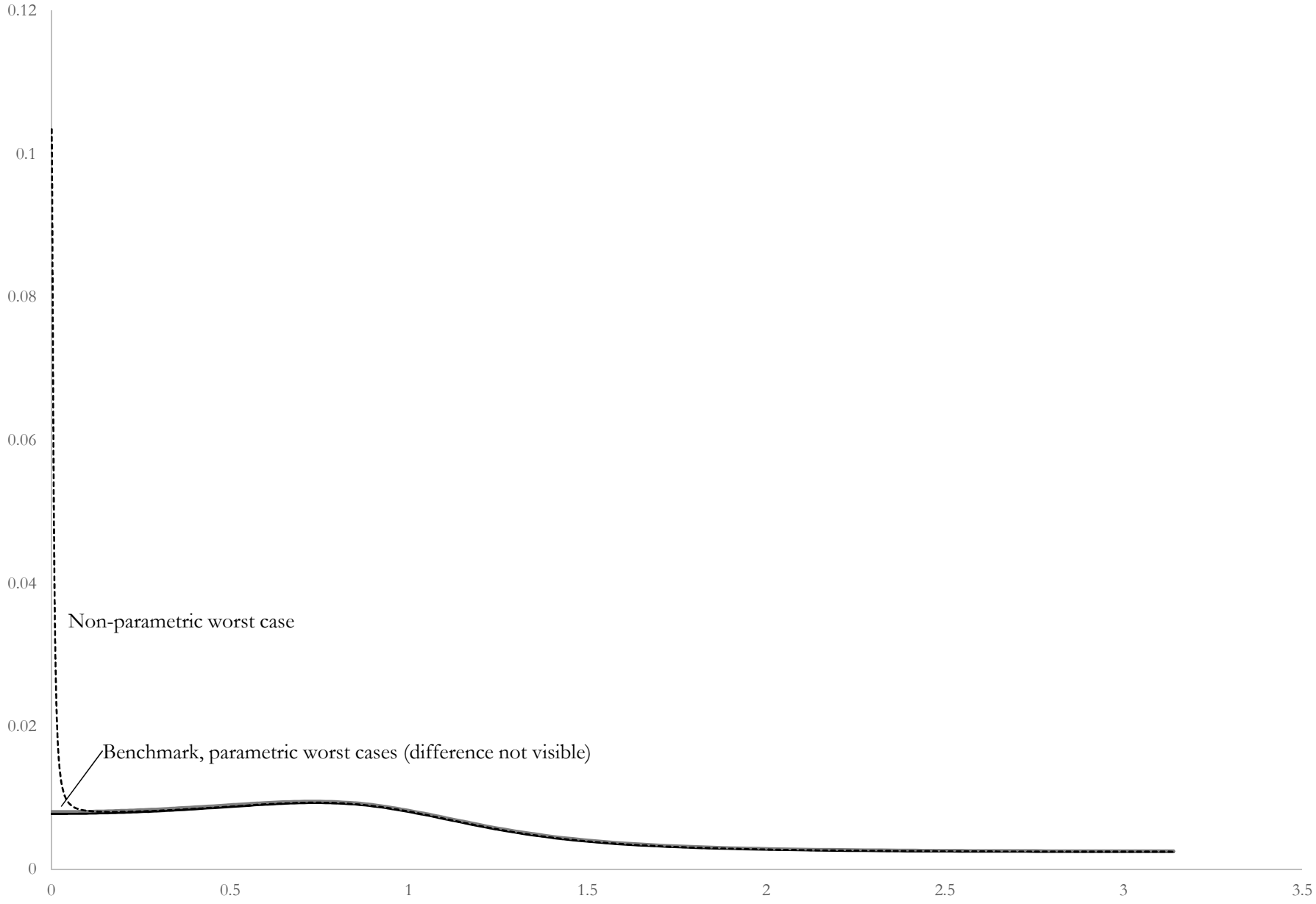


Figure 6. MA coefficients in benchmark and worst-case models for AR(2)

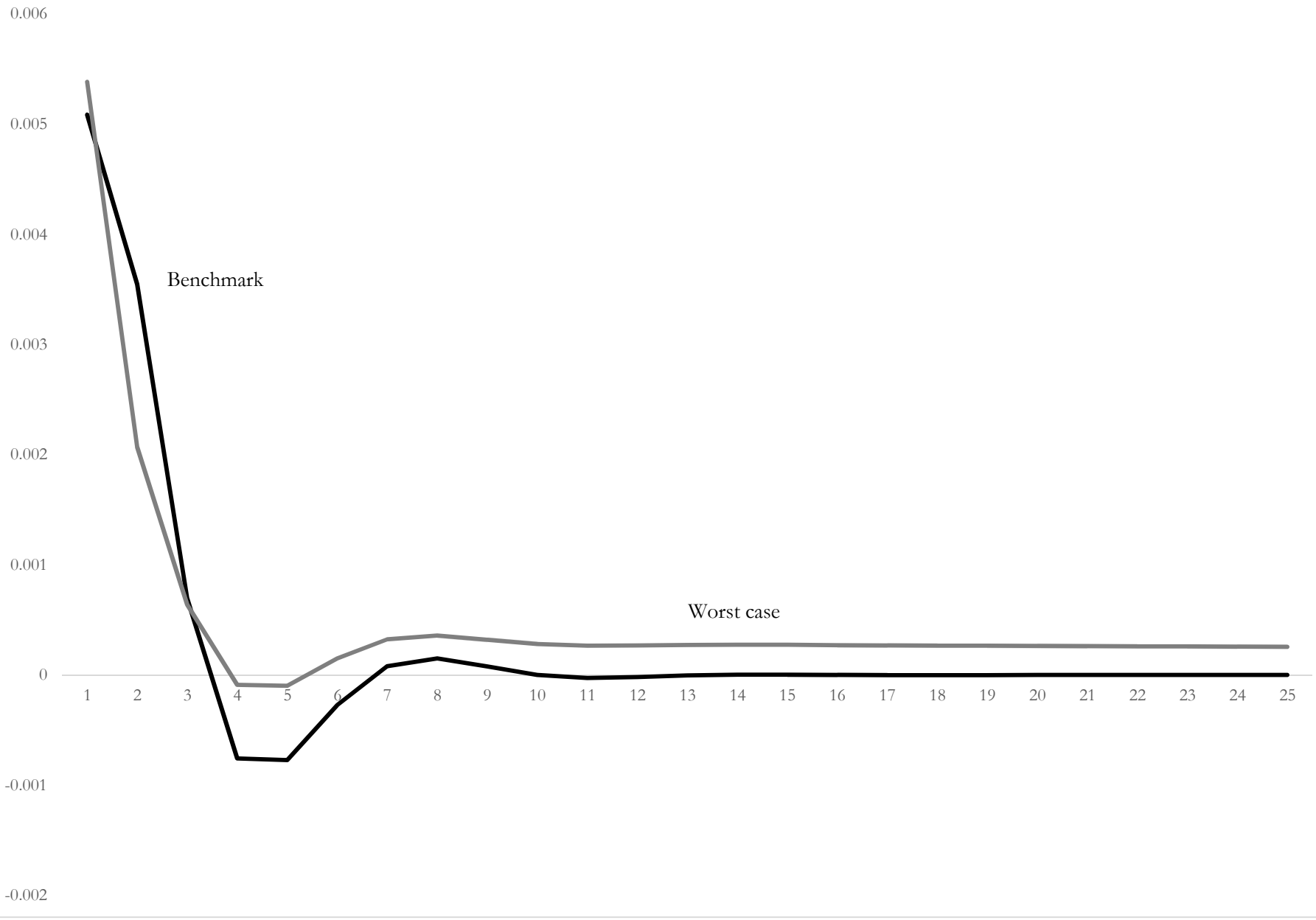
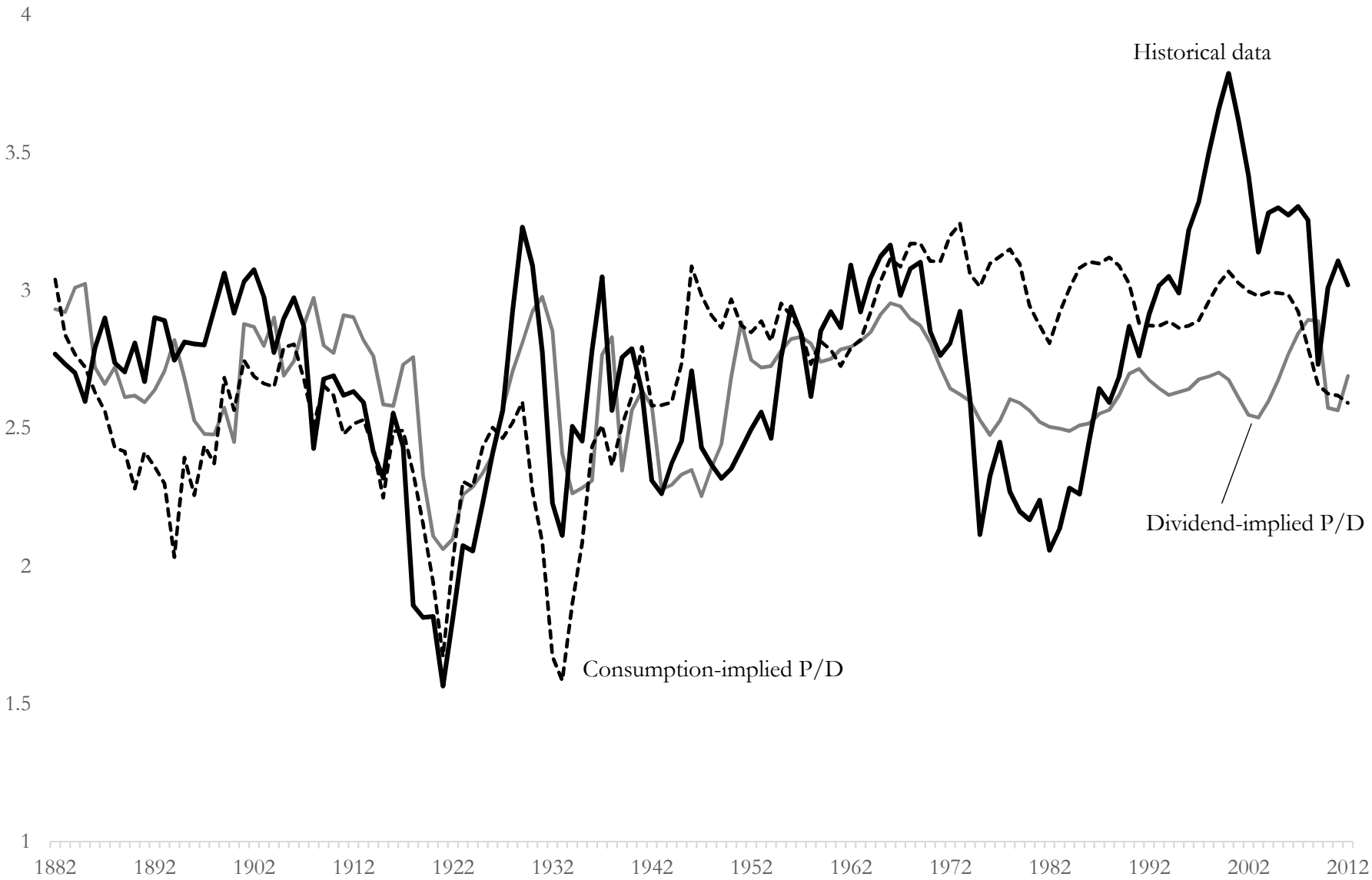


Figure 7. Historical and model-implied log price/dividend ratios



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.

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

Fundamental parameters		*Implied worst-case*		*Standard Epstein–Zin/robust-control*	
$\overline{b_0}$	0.01	b_0^w	0.010	b_0	0.01
$\overline{b}(\beta)$	0.01	$b^w(\beta)$	0.039	$b(\beta)$	0.01
β	0.997	θ	0.983		
λ	0.116	μ^w	-0.167	μ	-0.043
α (implied by λ)	5.31				
γ	5.62				

Asset pricing moments (annualized)

	Model	Standard RC	Data
std(M)	0.35	0.09	
E[r-rf]	7.10	2.09	7
std(r)	20.08	11.24	21
E[rf]	0.31	0.81	0.91
std(rf)	0.31	0	1.16
AC1(PD)	0.94	N/A	0.91
std(P/D)	0.24	0	0.40

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

	50-year sample		100-year sample	
	Model	Standard RC	Model	Standard RC
Ljung–Box	4.6%	5%	4.9%	5%
ARMA(1,1)	5.4%	5%	6.3%	5%

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.