

# Long-Run Risk is the Worst-Case Scenario

Rhys Bidder and Ian Dew-Becker\*

December 30, 2014

## Abstract

We study an investor who is unsure of the dynamics of the economy. Not only are parameters unknown, but the investor does not even know what order model to estimate. She estimates her consumption process non-parametrically and prices assets using a pessimistic model that minimizes lifetime utility subject to a constraint on statistical plausibility. The equilibrium is exactly solvable and we show that the pricing model always includes long-run risks. With a single free parameter determining risk preferences, the model generates high and volatile risk premia, excess volatility in stock returns, return predictability, interest rates that are uncorrelated with expected consumption growth, and investor expectations that are consistent with survey evidence. Risk aversion is equal to 4.8, there is no stochastic volatility or disasters, and the pricing model is statistically nearly indistinguishable from the true data-generating process. The analysis yields a general characterization of behavior under a very broad form of model uncertainty.

---

\*Bidder: Federal Reserve Bank of San Francisco. Dew-Becker: Northwestern University. We appreciate helpful comments and discussions from Harjoat Bhamra, Stefano Giglio, Stavros Panageas, Costis Skiadas, Matt Smith, and seminar participants at the Federal Reserve Bank of San Francisco, Duke Fuqua, the Asset Pricing Retreat in Tilburg, Kellogg, UC Santa Cruz, and the NBER Asset Pricing meeting. 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. There has been a long 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 investor is also unsure of the true model driving the world. This paper studies the behavior of such an agent.

With exactly solved results, we show that a model in which investors have Epstein–Zin preferences and uncertainty about consumption dynamics generates high and volatile risk premia, excess volatility in stock returns, a large degree of predictability in stock returns, low and stable interest rates, an estimated elasticity of intertemporal substitution from interest rate regressions of zero as measured in Campbell and Mankiw (1989), and behavior of investor expectations for stock returns that is consistent with survey evidence. The results hold with both exogenous and endogenous consumption, and without needing stochastic volatility to generate excess variance in asset prices.

We argue that investors consider a set of models of the economy that is only weakly constrained. Rather than just allowing uncertainty about the parameters in a specific model, or putting positive probability on a handful of models, we treat them as considering an infinite-dimensional space of autoregressive moving average (ARMA) specifications. People face pervasive ambiguity: no one can say they know the exact specification to estimate to forecast economic activity.

Acknowledging the presence of ambiguity, investors use a model for decision making that is robust in the sense that it generates the lowest lifetime utility among statistically plausible models. That is, they make decisions that are optimal in an unfavorable world. We derive that "worst-case" model in closed form and then explore its implications for asset prices.

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

The results are derived in the frequency domain, which allows strikingly clear conceptual and analytical insights. Two factors determine the behavior of the worst-case model at each frequency:

estimation uncertainty and how utility is affected by fluctuations at that frequency. Growth under the worst-case model has larger fluctuations at frequencies about which there is more uncertainty or that are more painful. Quantitatively, we find that differences in estimation uncertainty across frequencies are relatively small compared to differences in utility. Instead, since people with Epstein–Zin preferences are highly averse to low-frequency fluctuations, those shocks play the largest role in robust decision-making.

A criticism of the long-run risk model has always been that it depends on a process for consumption growth that is difficult to test for (Beeler and Campbell (2012); Marakani (2009)). We turn that idea on its head and argue that it is the difficulty of testing for and rejecting long-run risk that actually makes it a sensible model for investors to focus on. If anything, our result is more extreme than that of Bansal and Yaron (2004): whereas they posit a consumption growth trend with shocks that have a half-life of 3 years, the endogenous worst-case model that we derive features trend shocks with a half-life of 70 years.

In a calibration of the model, we show that it explains a wide range of features of financial markets that have been previously viewed as puzzling. Similar to the intuition from Bansal and Yaron (2004), equities earn high average returns in our model because low-frequency fluctuations in consumption growth induce large movements in both marginal utility and equity prices. In our setup, though, long-run risk need not actually exist – it only needs to be plausible.<sup>1</sup>

The results that we obtain on excess volatility, forecasting, and interest rate regressions all follow from the fact that the pricing model that our investor uses always involves more persistence than her point estimate (i.e. the model used by an econometrician with access to the same data sample as the investor). Since the pricing model has excess persistence, investors overextrapolate recent news relative to what the point estimate would imply. Following positive shocks, then, stock prices are high and econometric forecasts of returns are low. We are thus able to generate predictability without any changes in risk or risk aversion over time. Moreover, investors’ expectations for returns are positively correlated with past returns and negatively correlated with econometric forecasts of future returns, exactly as in the survey evidence discussed by Greenwood and Shleifer (2014) and Kojien, Schmeling, and Vrugt (2014). Significantly, we obtain this result in a rational model where

---

<sup>1</sup>This result is similar to that of Bidder and Smith (2013), who also develop a model where the worst-case process for consumption growth features persistence not present in the true process.

people make no mistakes – they are not irrational, just uncertain.

In generating all these results we have no more free parameters than other standard models of consumption and asset prices. We link the parameter that determines how the agent penalizes deviations from her point estimate for consumption dynamics directly to the coefficient of relative risk aversion. There is thus a single free parameter that determines risk preferences, and it corresponds to a coefficient of relative risk aversion of only 4.8. We also take no extreme liberties with beliefs – the investor’s pricing model is essentially impossible to distinguish from the true model in a 100-year sample. Using a correctly specified likelihood ratio test, the null hypothesis that the pricing model is true is rejected at the 5-percent level in only 6.6 percent of samples. And finally, the results are not driven by our assumption of an endowment economy – they also hold when consumption is endogenous.

It is difficult to argue that people do not face model uncertainty, but what is difficult is identifying exactly what features of the world people are uncertain about. The types of model uncertainty most difficult to resolve empirically can be divided into three classes: tail events, regime changes, and long-run behavior. Important work has been and will be done on the first two areas.<sup>2</sup> The goal of this paper is to address the last phenomenon and provide a very general analysis of long-run uncertainty and uncertainty about autocorrelations in the economy more broadly. So while we choose to examine the asset pricing implications of our model as a way of understanding our agent’s behavior, the applications and implications for our approach are much broader.

Our analysis directly builds on a number of important areas of research. First, the focus on a single worst-case outcome is closely related to Gilboa and Schmeidler’s (1989) work on ambiguity aversion. Second, we build on the analysis of generalized recursive preferences to allow for the consideration of multiple models, especially Hansen and Sargent (2010) and Ju and Miao (2012).<sup>3</sup> The work of Hansen and Sargent (2010) is perhaps most comparable to ours, in that they study an investor who puts positive probability on both a white-noise and a long-run risk model for

---

<sup>2</sup>Among others, Liu et al. (2004) and Collin-Dufresne et al. (2013) study extreme events and Ju and Miao (2012) study regime changes.

<sup>3</sup>See, e.g., Kreps and Porteus (1978); Weil (1989); Epstein and Zin (1991); Maccheroni, Marinacci, and Rustichini (2006); and Hansen and Sargent (2005), among many others. There is also a large recent literature in finance that specializes models of ambiguity aversion to answer particularly interesting economic questions, such as Liu 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. Recent papers on asset pricing with learning and ambiguity aversion include Veronesi (2000), Brennan and Xia (2001), Epstein and Schneider (2007), Cogley and Sargent (2008), Leippold et al. (2008), Ju and Miao (2012), and Collin-Dufresne et al. (2013).

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 also show that the pessimistic model examined by Hansen and Sargent is twice as easy for an investor to reject than the one we obtain.

Finally, since the worst-case model is more persistent than the point estimate, pricing behavior is similar to the extrapolation implied by the "natural expectations" studied by Fuster et al. (2011). Our results differ from theirs, though, in that we *always* obtain excess extrapolation, whereas in their setting it results from the interaction of suboptimal estimation on the part of investors with a specific data-generating process. Nevertheless, our paper complements the literature on belief distortions and extrapolative expectations by deriving them as a natural response to model uncertainty.<sup>4</sup>

More generally, we provide a framework for linking ambiguity aversion with non-parametric estimation, which we view as a realistic description of how people might think about the models they estimate. While people must always estimate finite-order models when only finite data is available, they likely understand that those models almost certainly are misspecified. So if they want to be prepared for a worst-case scenario, they need to consider very general deviations from their point estimate. We provide a way to characterize and analyze those types of deviations.

The remainder of the paper is organized as follows. Section 2 discusses the agent's estimation method. Section 3 describes the basic structure of the agent's preferences, and section 4 then derives the worst-case model. We examine asset prices in general under the preferences in section 5. Section 6 then discusses the calibration and section 7 analyzes the quantitative implications of the model. We extend the results to account for endogenous consumption in section 8, and section 9 concludes.

## 2 Non-parametric estimation

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

---

<sup>4</sup>See Barsky and De Long (1993), Cecchetti et al. (1997), Fuster et al. (2011), and Hirshleifer and Yu (2012)

## 2.1 Economic environment

For our main analysis we study a pure endowment economy.

**Assumption 1.** *Investors form expectations for future log consumption growth,  $\Delta c$ , using models of the form*

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

$$\varepsilon_t \sim N(0, 1) \quad (2)$$

where  $\mu$  is mean consumption growth,  $a(\cdot)$  is a polynomial function,  $L$  is the lag operator, and  $\varepsilon_t$  is an innovation.

The change in log consumption on date  $t$ ,  $\Delta c_t$ , is a function of past consumption growth and a shock. We restrict our attention to models with purely linear feedback from past to current consumption growth. It seems reasonable to assume that people use linear models for forecasting, even if consumption dynamics are not truly linear, given that the economics literature focuses almost exclusively on linear models. Moreover, the Wold theorem states that any covariance stationary process can be represented in the form (1), or the associated moving average representation, with uncorrelated, though not necessarily independent innovations  $\varepsilon_t$ . For our purposes, the restriction to the class of linear processes is a description of the agent's modeling method, rather than an assumption about the true process driving consumption growth. The assumption that  $\varepsilon_t$  is normally distributed is not necessary, but it simplifies the exposition.<sup>5</sup>

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

$$\Delta c_t = \mu + b(L)\varepsilon_t \quad (3)$$

$$\text{where } b(L) \equiv (1 - La(L))^{-1} b_0 \quad (4)$$

$$= \sum_{j=1}^{\infty} b_j L^j \quad (5)$$

We can thus express the dynamics of consumption equivalently as depending on  $\{a, b_0\}$  or just on the polynomial  $b$ , with coefficients  $b_j$ . The two different representations are each more convenient

---

<sup>5</sup>The appendix solves the model when  $\varepsilon_t$  has an arbitrary distribution.

than the other in certain settings, so we will refer to both in what follows. They are directly linked to each other through equation (4), so that a particular choice of  $\{a, b_0\}$  is always associated with a distinct value of  $b$  and vice versa.

There are no latent state variables. When a model  $a(L)$  has infinite order we assume that the agent knows all the necessary lagged values of consumption growth for forecasting (or has dogmatic beliefs about them) so that no filtering is required.<sup>6</sup> We discuss necessary constraints on the models below. For now simply assume that they are sufficiently constrained that any quantities we must derive exist.

We assume that the investor knows the value of  $\mu$  with certainty but is uncertain about consumption dynamics.<sup>7</sup>

## 2.2 Estimation

For the purpose of forecasting consumption growth, the agent in our model chooses among specifications for consumption growth,  $b$ , partly based on their statistical plausibility. That plausibility is measured by a loss function  $g(b; \bar{b})$  relative to a point estimate  $\bar{b}$ . As a simple example, if a person were to estimate a parameterized model, such as an AR(1), on a sample of data, she would have a likelihood (or posterior distribution) over the autocorrelation, and she could rank different AR(1) processes by how far their autocorrelations are from her point estimate. That example, though, imposes a specific parametric specification of consumption growth and rules out all other possible models.

In the spirit of modeling the agent as looking for decision rules that are robust to a broad class of potential models, we assume that she estimates the dynamic process driving consumption non-parametrically so as to make only minimal assumptions about the driving process. Following Berk (1974) and Brockwell and Davis (1988), we assume that the investor estimates a finite-order

---

<sup>6</sup>Croce, Lettau, and Ludvigson (2014) study in detail issues surrounding filtering and the type of information available to agents in long-run risk models. Our agent has less information than in Bansal and Yaron (2004), in some sense, because she can only observe past consumption growth and no other state variables. Conditional on a model for consumption growth, though, our agent filters optimally. Croce, Lettau, and Ludvigson (2014) suggest optimal filtering may in fact be rather difficult.

<sup>7</sup>In addition to parsimony, there are two justifications for that assumption. First, for the estimation method we model the agent as using, estimates of the coefficients  $\{a, b_0\}$  converge at an asymptotically slower rate than estimates of  $\mu$ . Second, we will model the agent as having Epstein–Zin (1991) preferences, which, as shown by Barillas, Hansen, and Sargent (2009), can be viewed as appearing when an agent with power utility is unsure of the distribution of the innovations of consumption growth. So uncertainty about the mean will implicitly be accounted for through the preferences.

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

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

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

for  $i = \sqrt{-1}$ . The transfer function measures how the filter  $b(L)$  transfers power at each frequency,  $\omega$ , from the white-noise innovations,  $\varepsilon$ , to consumption growth. Berk (1974) and Brockwell and Davis (1988) show that under standard conditions, estimates of the transfer function,  $\hat{B}(\omega)$ , are asymptotically complex normal and independent across frequencies, with variance proportional to a function  $\bar{f}(\omega) \equiv |\bar{B}(\omega)|^2$ , the spectral density.<sup>9,10</sup>

Given a point estimate  $\bar{b}$  (with associated transfer function  $\bar{B}(\omega)$ ) a natural distance measure relative to the point estimate is then embodied in the following assumption

**Assumption 2.** *Investors measure the statistical plausibility of a model through the distance measure,*

$$g(b; \bar{b}) = \int \frac{|B(\omega) - \bar{B}(\omega)|^2}{\bar{f}(\omega)} d\omega \tag{7}$$

---

<sup>8</sup>We provide a more detailed treatment of this estimation approach in the appendix.

<sup>9</sup>The key condition on the dynamic process for consumption growth that is required for the asymptotic distribution theory is that the true spectral density is finite and bounded away from zero. Alternatively, one may assume that the MA coefficients are absolutely summable. See Stock (1994) for a discussion of the relationship between such conditions. The conditions eliminate some pathological processes and also fractional integration. The innovations to consumption growth must also have a finite fourth moment.

<sup>10</sup>Technically, the two papers derive results on the spectral density of consumption growth. The appendix extends their results to the transfer function.



where  $\bar{f}(\omega)$  measures the estimation uncertainty at frequency  $\omega$  and, here and below, integrals without limits denote  $(2\pi)^{-1} \int_{-\pi}^{\pi}$ .

$g(b; \bar{b})$  is a  $\chi^2$ -type test statistic for the null hypothesis that  $B = \bar{B}$ .<sup>11</sup> The appendix gives a fuller derivation of this distance measure and shows that it is essentially equivalent to a Wald test in the time domain, based on the non-parametric asymptotics of Berk (1974) and Brockwell and Davis (1988).

### 3 Preferences

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

#### 3.1 Utility given a model

**Assumption 3.** *Given a forecasting model  $\{a, b_0\}$ , the investor's utility is described by Epstein–Zin (1991) preferences. The investor's coefficient of relative risk aversion is  $\alpha$ , her time discount parameter is  $\beta$ , and her elasticity of intertemporal substitution (EIS) is equal to 1.<sup>12</sup> Lifetime utility,  $v$ , for a fixed model  $\{a, b_0\}$ , is*

$$v(\Delta c^t; a, b_0) = (1 - \beta) c_t + \frac{\beta}{1 - \alpha} \log E_t [\exp(v(\Delta c^{t+1}; a, b_0) (1 - \alpha)) | a, b_0] \quad (8)$$

$$= c_t + \sum_{k=1}^{\infty} \beta^k E_t [\Delta c_{t+k} | a, b_0] + \frac{\beta}{1 - \beta} \frac{1 - \alpha}{2} b(\beta)^2 \quad (9)$$

where  $E_t[\cdot | a, b_0]$  denotes the expectation operator conditional on the history of consumption growth up to date  $t$ ,  $\Delta c^t$ , assuming that consumption is driven by the model  $\{a, b_0\}$ .

<sup>11</sup>Hong (1996) studies a closely related distance metric in the context of testing for general deviations from a benchmark spectral density.

<sup>12</sup>We focus on the case of a unit EIS to ensure that we can easily derive analytic results. The precise behavior of interest rates is not our primary concern, so a unit EIS is not particularly restrictive. The unit EIS also allows us to retain the result that Epstein–Zin preferences are observationally equivalent to a robust control model, as in Barillas, Hansen, and Sargent (2009), which will be helpful in our calibration below

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

### 3.2 Robustness over dynamics

Equation (8) gives lifetime utility conditional on consumption dynamics. We now discuss the investor's consideration of alternative models of dynamics.

The investor entertains a set of possible values for the lag polynomial and can associate with any model a measure of its plausibility,  $g(b; \bar{b})$ . Seeking robustness, the investor makes decisions that are optimal in an unfavorable world – specifically, as though consumption growth is driven by worst-case dynamics, denoted  $b^w$ . These dynamics are not the worst in an unrestricted sense but, rather, are the worst among statistically plausible models. So the investor does not fear completely arbitrary models – she focuses on models that do not have Wald statistics (relative to her point estimate) that are too high.

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

$$b^w = \arg \min_b \{ E [v(\Delta c^t; b) | b] + \lambda g(b; \bar{b}) \} \quad (10)$$

$b^w$  is the model that gives the agent the lowest unconditional expected lifetime utility, subject to the penalty  $g(b; \bar{b})$ .<sup>13</sup>  $\lambda$  is a parameter that determines how much weight the penalty receives. As usual,  $\lambda$  can either be interpreted directly as a parameter or as a Lagrange multiplier on a constraint on the Wald-type statistic  $g(b; \bar{b})$ .

Models that deviate from the point estimate by a larger amount on average across frequencies ( $g(b; \bar{b})$  is big) are implicitly viewed as less plausible. The agent's assessment of plausibility is

---

<sup>13</sup>Since consumption can be non-stationary, this expectation does not always exist. In that case, we simply rescale lifetime utility by the level of consumption yielding  $E[v(\Delta c^t; b) - c_t | b]$ , which does exist. Scaling by consumption is a normalization that has no effect other than to ensure the expectation exists.

based on our statistical measure of distance, and controlled by  $\lambda$ . We are modeling the agent's beliefs about potential models by assuming that she compares possible models to a point estimate  $\bar{b}$ . The role of  $g(b; \bar{b})$  in our analysis is intuitively similar to that of the relative entropy penalty (or distance constraint) used in the robust control model of Hansen and Sargent (2001), in that it imposes discipline on what models the investor considers and prevents the analysis from being vacuous.

It is important to note that the penalty function can be calculated for values of  $b$  that are ruled out under the asymptotic assumptions of Berk (1974) and Brockwell and Davis (1988) that we used in connecting  $g(b; \bar{b})$  to the estimation process. Most interestingly,  $g(b; \bar{b})$  is well defined for certain fractionally integrated processes so that the agent therefore allows for fractional integration in the models she considers. So if we find the worst case does not involve fractional integration (as, in fact, will be the case below), it is a result rather than an assumption.

In the end, our ambiguity-averse investor's utility takes the form of that of an Epstein–Zin agent but using  $b^w$  to form expectations about future consumption growth,<sup>14</sup>

$$v^w(\Delta c^t) = v(\Delta c^t; b^w) = c_t + \frac{\beta}{1-\beta} \frac{1-\alpha}{2} b^w (\beta)^2 + \sum_{k=1}^{\infty} \beta^k E_t[\Delta c_{t+k} | b^w] \quad (11)$$

In modeling investors as choosing a single worst-case  $b^w$ , we obtain a setup similar to Gilboa and Schmeidler (1989), Maccheroni, Marinacci, and Rustichini (2006), and Epstein and Schneider (2007) in the limited sense that we are essentially constructing a set of models and minimizing over that set. Our worst-case model is, however, chosen once and for all and is not state- or choice-dependent. The choice of  $b^w$  is timeless – it is invariant to the time-series evolution of consumption – so what it represents is an unconditional worst-case model: if an agent had to choose a worst-case model to experience prior to being born into the world, it would be  $b^w$ . Unlike in some related recent papers, the investor in this model does not change her probability weights every day. She chooses a single pessimistic model to protect against.

A natural question is why we analyze a worst case instead of allowing the agent to average as a Bayesian across all the possible models. A simple and reasonable answer is that people may not

---

<sup>14</sup>Note that since utility is recursive, the agent's preferences are time-consistent, but under a pessimistic probability measure. Furthermore, the assumption that  $b^w$  is chosen unconditionally means that  $b^w$  is essentially unaffected by the length of a time period, so the finding in Skiadas (2013) that certain types of ambiguity aversion become irrelevant in continuous time does not apply here.

actually be Bayesians, or they may not be able to assign priors to all models.<sup>15</sup> This answer seems particularly compelling in the context of uncertainty about the long run which, as we shall now see, is an important concern of our agent.<sup>16</sup>

## 4 The worst-case scenario

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

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

$$b^w = \arg \min_b \frac{\beta}{1-\beta} \frac{1-\alpha}{2} b(\beta)^2 + \lambda g(b; \bar{b}) \quad (12)$$

*the worst-case model is*

$$B^w(\omega) = \bar{B}(\omega) + \lambda^{-1} \frac{\beta}{1-\beta} \frac{\alpha-1}{2} b^w(\beta) \times \bar{f}(\omega) \times Z^*(\omega) \quad (13)$$

where  $Z(\omega) \equiv \sum_{j=0}^{\infty} \beta^j e^{-i\omega j}$ ,  $a^*$  denotes a complex conjugate, and  $b^w(\beta)$  is given by

$$b^w(\beta) = \frac{\bar{b}(\beta)}{1 - \lambda^{-1} \frac{\beta}{1-\beta} \frac{\alpha-1}{2} \int Z(\omega)^* Z(\omega) \bar{f}(\omega) d\omega} \quad (14)$$

The time-domain model  $b^w(L)$  has coefficients  $b_j^w$  that are obtained through the inverse Fourier transform,

$$b_j^w = \int B^w(\omega) e^{-i\omega j} d\omega \quad (15)$$

<sup>15</sup>See Machina and Siniscalchi (2014) for a recent review of the experimental literature on ambiguity aversion.

<sup>16</sup>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  $\varepsilon$ , which does not take a tractable form. Kreps (1998) discusses related issues in models with learning. 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. The analysis of the worst-case model gives us a tractable view into an agent's decision problem that would otherwise not be available.

We thus have a closed form expression for the worst-case model.<sup>17</sup> The worst-case transfer function  $B^w(\omega)$  in (13) is equal to the true transfer function plus a term that depends on three factors. First,  $\lambda^{-1} \frac{\beta}{1-\beta} \frac{\alpha-1}{2} b^w(\beta)$  represents the ratio of the utility losses from a marginal increase in  $b^w(\beta)$  to the cost of deviations from the point estimate,  $\lambda$ . When risk aversion,  $\alpha$ , is higher or the cost of deviating from the point estimate,  $\lambda$ , is lower, the worst-case model is farther from the point estimate.

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

Finally,  $Z(\omega)$  determines how much weight the lifetime utility function places on frequency  $\omega$ . Figure 1 plots the real part of  $Z(\omega)$  for  $\beta = 0.99$ , a standard annual calibration.<sup>18</sup> It is strikingly peaked near frequency zero; in fact, the x-axis does not even show frequencies corresponding to cycles lasting less than 10 years because they carry essentially zero weight. Since the mass of  $Z(\omega)$  lies very close to frequency 0, the worst case shifts power to very low frequencies. In that sense, the worst-case model *always* includes long-run risk.

Equation (13) represents the completion of the solution to the model. To summarize, given a point estimate for dynamics,  $\bar{B}(\omega)$  (estimated from a finite-order model that we need not specify here), the agent selects a worst-case model  $B^w(\omega)$ , which is associated with a unique  $b^w(L)$  through the inverse Fourier transform. She then uses the worst-case model when calculating expectations and pricing assets.

#### 4.1 Long-run risk is the worst case scenario

**Corollary 2** *Suppose the agent's point estimate is that consumption growth is white noise, with*

$$\bar{b}(L) = \bar{b}_0 \tag{16}$$

---

<sup>17</sup>In the case where  $\varepsilon_t$  is not normally distributed, the solution is determined by the condition  $B^w(\omega) = \bar{B}(\omega) - \frac{\lambda^{-1}}{2} \frac{\beta}{1-\beta} f(\omega) Z^*(\omega) \Gamma'(b^w(\beta)(1-\alpha))$ , where  $\Gamma'$  is the derivative of the cumulant-generating function of  $\varepsilon$ .

<sup>18</sup>Dew-Becker and Giglio (2013) use the fact that  $B(\omega)$  is a causal filter to show that the real part of  $Z(\omega)$  is sufficient to characterize  $\int Z(\omega) B(\omega) d\omega$ .

The worst-case model is then

$$b_0^w = \bar{b}_0 + \varphi \quad (17)$$

$$b_j^w = \varphi \beta^j \text{ for } j > 0 \quad (18)$$

$$\text{where } \varphi \equiv \frac{\beta}{1-\beta} \frac{\alpha-1}{2} b^w(\beta) \bar{b}_0^2 \lambda^{-1} \quad (19)$$

The MA process in (17-19) is an ARMA(1,1) and has an equivalent state-space representation

$$\Delta c_t = \mu + x_{t-1} + \eta_t \quad (20)$$

$$x_t = \beta x_{t-1} + v_t \quad (21)$$

where  $\eta_t$  and  $v_t$  are independent and normally distributed innovations.<sup>19</sup> The state-space form in equations (20–21) is observationally equivalent to the MA process (17–19) in the sense that they have identical autocovariances, and it is exactly case I from Bansal and Yaron (2004), the homoskedastic long-run risk model. So when the agent’s point estimate is that consumption growth is white noise, her worst-case model is literally the long-run risk model.

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

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

---

<sup>19</sup>  $\eta_t \sim N\left(0, \theta \beta^{-1} (\bar{b}_0 + \varphi)^2\right)$  and  $v_t \sim N\left(0, (1 - \beta\theta)(\beta - \theta) \beta^{-1} (\bar{b}_0 + \varphi)^2\right)$ , where  $\theta \equiv \beta(1 - \varphi \bar{b}_0^{-1})$

## 5 The behavior of asset prices

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

$$1 = E_t [R_{t+1} M_{t+1} | b^w] \quad (22)$$

$$\text{where } M_{t+1} \equiv \beta \exp(-\Delta c_{t+1}) \frac{\exp(v(\Delta c^{t+1}; b^w) \times (1 - \alpha))}{E_t [\exp(v(\Delta c^{t+1}; b^w) \times (1 - \alpha)) | b^w]} \quad (23)$$

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

### 5.1 Consumption claims

It is straightforward, given that log consumption follows a linear Gaussian process, to derive approximate expressions for prices and returns on levered consumption claims. We consider an asset whose dividend is  $C_t^\gamma$  in every period, where  $\gamma$  represents leverage. Denote the return on that asset on date  $t + 1$  as  $r_{t+1}$  and the real risk-free rate as  $r_{f,t+1}$ . We will often refer to the levered consumption claim as an equity claim, and we view it as a simple way to model equity returns (Abel (1999)).

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

$$E_t [r_{t+1} - r_{f,t+1} | \bar{a}, \bar{b}_0] = \frac{\gamma - \delta a^w(\delta)}{1 - \delta a^w(\delta)} (\bar{a}(L) - a^w(L)) (\Delta c_t - \mu) + \underbrace{-\frac{\gamma - \delta a^w(\delta)}{1 - \delta a^w(\delta)} b_0^w [b_0^w - (1 - \alpha) b^w(\beta)]}_{cov^w(r_{t+1}, m_{t+1})} - \underbrace{\frac{1}{2} \left( \frac{\gamma - \delta a^w(\delta)}{1 - \delta a^w(\delta)} \right)^2 (b_0^w)^2}_{\frac{1}{2} var_t^w(r_{t+1})} \quad (24)$$

where  $\delta$  is a linearization parameter from the Campbell–Shiller approximation that depends on the

steady-state price/dividend ratio.<sup>20</sup>

The second line, which is equal to  $-cov_t^w(r_{t+1}, \log M_{t+1}) - \frac{1}{2}var_t^w(r_{t+1})$  (i.e. a conditional covariance and variance measured under the worst-case dynamics), is the standard risk premium, and it is calculated under the worst-case model. The primary way that the model increases risk premia compared to standard Epstein–Zin preferences is that the covariance of the return with the SDF is higher. That covariance, in turn, is higher for two reasons. First, since the agent believes that shocks to consumption growth are highly persistent, they have large effects on lifetime utility, thus making the SDF very volatile (the term  $b_0^w - (1 - \alpha)b^w(\beta)$ ). Second, again because of the persistence of consumption growth under the worst case, shocks to consumption have large effects on expected long-run dividend growth, so the return on the levered consumption claim is also very sensitive to shocks (through  $\frac{\gamma - \delta a^w(\delta)}{1 - \delta a^w(\delta)} b_0^w$ ). These two effects cause the consumption claim to strongly negatively covary with the SDF and generate a high risk premium.

The second difference between the risk premium in this model and a setting where the investor prices assets under the point estimate is the term  $\frac{\gamma - \delta a^w(\delta)}{1 - \delta a^w(\delta)} (\bar{a}(L) - a^w(L)) (\Delta c_t - \mu)$ , which reflects the difference in forecasts of dividend growth between the point estimate, used by the econometrician, and the worst-case model, used by investors. Since  $\Delta c_t - \mu$  is zero on average, this term is also zero on average. But it induces predictability in returns. When the worst-case implies higher future consumption growth, investors pay relatively more for equity, thus raising asset prices and lowering expected returns. This channel leads to procyclical asset prices and countercyclical expected returns when  $a^w(L)$  implies more persistent dynamics than  $\bar{a}(L)$ , similarly to Fuster et al. (2011).

We also note that since risk aversion and conditional variances are constant, the excess return on a levered consumption claim has a constant conditional expectation from the perspective of investors. That is, while returns are predictable from the perspective of an econometrician, investors believe that they are unpredictable. So if investors in this model are surveyed about their expectations of excess returns, their expectations will not vary, even if econometric evidence implies that returns are predictable.

---

<sup>20</sup>See the appendix for a full derivation.



## 5.2 Interest rates

The risk-free rate follows

$$r_{f,t+1} = -\log \beta + \mu + a^w(L) (\Delta c_t - \mu) - \frac{1}{2} (b_0^w)^2 + b_0^w b^w(\beta) (1 - \alpha) \quad (25)$$

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

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

## 5.3 Investor expectations of returns

Because both risk aversion and the quantity of risk in the model are constant, investor expectations of excess returns on consumption claims are constant. Equation (25), though, shows that interest rates vary over time. If the worst-case model  $a^w(L)$  induces persistence in consumption growth, then interest rates are high following past positive shocks. Investors therefore expect high total equity returns following past high returns. At the same time, when  $a^w(L)$  induces more persistence than  $\bar{a}(L)$ , econometricians expect low excess returns following past high returns. So we find that investor expectations for returns are positively correlated with past returns and negatively correlated with statistical expectations of future returns. This behavior is exactly the qualitative behavior that Greenwood and Shleifer (2014) observe in surveys of investor expectations. We analyze these results quantitatively below.

## 6 Calibration

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

Many of the required parameters are standard. We use a quarterly calibration of  $\beta = 0.99^{1/4}$ ,

implying a pure rate of time preference of 1 percent per year. The steady-state dividend/price ratio used in the Campbell–Shiller approximation is 5 percent per year, as in Campbell and Vuolteenaho (2004), so  $\delta = 0.95^{1/4}$ . Other parameters are calibrated to match moments reported in Bansal and Yaron (2004). The agent’s point estimate is that consumption growth is i.i.d. with a quarterly standard deviation of 1.47 percent, which we also assume is the true data-generating process. Finally, the leverage parameter for the consumption claim,  $\gamma$ , is set to 4.626 to generate mean annualized equity returns of 6.33 percent.

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

$$\alpha = 1 + \frac{1}{2\lambda(1 - \beta)} \tag{26}$$

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

We calibrate  $\lambda$  to equal 52.23 to match the observed Sharpe ratio on equities. Formula (26) then implies  $\alpha$  should equal 4.81. That level of risk aversion is extraordinarily small in the context of the consumption-based asset pricing literature with Gaussian innovations. It is only half the value used by Bansal and Yaron (2004), for example, who themselves are notable for using a relatively low value.  $\alpha$  therefore immediately seems to take on a plausible value in its own right, separate from any connection it has to  $\lambda$ .

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

## 7 Quantitative implications

### 7.1 The white noise case

We report the values of the parameters in the worst-case consumption process in table 1. As noted above, the autocorrelation of the predictable part of consumption growth under the worst-case model is  $\beta$ , implying that trend shocks have a half-life of 70 years, as opposed to the three-year half-life in the original calibration in Bansal and Yaron (2004). However,  $b^w(\beta)$ , the relevant measure of the total risk in the economy, is 0.044 at the quarterly frequency in both our model and theirs. The two models thus both have the same quantity of long-run risk, but in our case the long-run risk comes from a smaller but more persistent shock.

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

#### 7.1.1 Unconditional moments

Table 1 reports key asset pricing moments. The first column shows that the model can generate a high standard deviation for the pricing kernel (and hence a high maximal Sharpe ratio), high and volatile equity returns, and low and stable real interest rates, as in the data. The equity premium and its volatility are 6.33 and 19.42 percent respectively, identical to the data. The real risk-free rate has a mean of 1.89 percent and a standard deviation of 0.33 percent.

The second column in the bottom section of table 1 shows what would happen if we set  $\lambda = \infty$  but held  $\alpha$  fixed at 4.81, so that we would be back in the standard Epstein–Zin setting where there is no uncertainty about dynamics. The equity premium then falls from 6.3 to 1.9 percent, since the agent exhibits no concern for long-run risk. Furthermore, because the agent no longer behaves as if consumption growth is persistent, a shock to consumption has far smaller effects on asset prices. The standard deviation of returns falls from 19.4 to 13.6 percent and the standard deviation of the price/dividend ratio falls from 20 percent to exactly zero. The agent’s fear of a model with long-run risk thus raises the mean of returns by a factor of more than 3 and the volatility by a factor of 1.4.

Going back to the first column, we see that there are large and persistent movements in the

price/dividend ratio in our model. The one-year autocorrelation of the price/dividend ratio at 0.96 is somewhat higher than the empirical autocorrelation, while the standard deviation is 0.20, similar to the empirical value of 0.29. These results are particularly notable given that there is no free parameter that allows us to directly match the behavior of prices.

Volatility in the price/dividend ratio has the same source as the predictability in equity returns discussed above: the agent prices assets under a model where consumption growth has a persistent component. So following positive shocks, she is willing to pay relatively more, 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 investor (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.

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

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

### **7.1.2 Return predictability**

To quantify the degree of predictability in returns, figure 3 plots percentiles of sample  $R^2$ s from regressions of returns on price/dividend ratios in 240-quarter samples (the approximate length of the post-war period). The gray line is the set of corresponding values from the empirical post-war

(1950–2010) sample. We report  $R^2$ s for horizons of 1 quarter to 10 years. At both short and long horizons the model matches well. The median  $R^2$  from the predictive regressions at the ten-year horizon is 37 percent, while in the data it is 29 percent.

### 7.1.3 Expectations of returns and expected returns

Greenwood and Shleifer (2014) summarize evidence from a wide range of surveys on investor expectations for equity returns. They obtain four key facts about investor expectations: they are positively correlated with price/dividend ratios, positively correlated with past returns, negatively correlated with future returns, and negatively correlated with statistical expectations of future returns. Our model replicates all four of those facts.

Table 2 summarizes regression coefficients and correlations analyzed by Greenwood and Shleifer (2014). All the regression coefficients and correlations in our model have the same sign as they report. Specifically, a regression of investor expectations (calculated under the pricing measure) on the log price/dividend ratio for the market yields a coefficient of 1.98, versus 1.08 in Greenwood and Shleifer (the difference is within their confidence intervals). A regression of expected returns on past returns yields a positive coefficient, though with a value much smaller than they observe.

In terms of forecasting, we replicate their finding that future excess returns are negatively predicted by the price/dividend ratio and also negatively predicted by investor expectations of future returns. The coefficients again differ somewhat from their reported values, but the results are qualitatively consistent: investor forecasts run the *wrong* direction.

### 7.1.4 Probability of rejecting the worst-case dynamics

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

We consider two tests of the fit of the worst-case model to the true white-noise consumption process: Ljung and Box’s (1978) portmanteau test and the likelihood-based test of an ARMA(1,1) suggested by Andrews and Ploberger (1996).<sup>21</sup> The likelihood-based test is in fact a correctly

---

<sup>21</sup>The intuition behind this approach is similar to that underpinning the detection error probability (DEP) calculations of Barillas, Hansen, and Sargent (2009), which are widely used to calibrate robustness models. Although we do not report them here, the DEPs in our case also indicate that the worst-case and benchmark models are extremely

specified likelihood-ratio test and thus should be asymptotically most powerful. To test that the worst-case model is the correct specification, we take a simulated sample of consumption growth,  $\Delta c_t$ , and construct artificial residuals,

$$\varepsilon_t^{\{a^w, b_0^w\}} \equiv (\Delta c_t - \mu - a^w(L)(\Delta c_{t-1} - \mu))(b_0^w)^{-1} \quad (27)$$

Under the null that the worst-case model is the correct specification,  $\varepsilon_t^{\{a^w, b_0^w\}}$  should be white noise. The Ljung–Box and Andrews–Ploberger tests both ask whether that null can be rejected. Since consumption growth is generated as white noise,  $\varepsilon_t^{\{a^w, b_0^w\}}$  is in fact not i.i.d.. In a sufficiently large sample, an investor will be able to reject the hypothesis that consumption was driven by the worst-case model by observing that  $\varepsilon_t^{\{a^w, b_0^w\}}$  is serially correlated.<sup>22</sup>

The top section of table 3 reports the probability that the agent would reject the hypothesis that consumption growth was driven by the worst-case model after observing a sample of white-noise consumption growth. We simulate the tests in both 50- and 100-year samples. In all four cases, the rejection probabilities are only marginally higher than they would be if the null hypothesis were actually true. The Ljung–Box test is the weaker of the two, with rejection rates of 4.7 and 4.8 percent in the 50- and 100-year samples, respectively, while the ARMA(1,1) likelihood ratio test performs only slightly better, with rates of 5.6 and 6.6 percent.

Table 3 thus shows that the worst-case model, while having economically large differences from the point estimate in terms of its asset pricing implications, can barely be distinguished from the point estimate in long samples of consumption growth. From a statistical perspective, it is entirely plausible that an investor would be concerned that the worst-case model could be what drives the data. Thus both  $\lambda$  and  $\alpha$  (which were calibrated jointly with only a single degree of freedom) take on independently reasonable values.

### 7.1.5 Alternative calibrations of the pricing model

We derive the worst-case model endogenously, but similar models have also been assumed for investor expectations. Bansal and Yaron (2004) argue that a model with trend shocks with a quarterly persistence 0.94 fits the data well. Hansen and Sargent (2010) consider a setting where

---

difficult to distinguish.

<sup>22</sup>We obtain small-sample critical values for the two test statistics by simulating their distributions under the null.

investors believe that consumption may be driven by one of two models, the more persistent of which has a trend component with an autocorrelation of 0.99. Due to ambiguity aversion in their model, asset prices are primarily driven by the more persistent model.

The bottom section of table 3 examines how rejection probabilities change if we modify the pricing model to use a less persistent trend. In all rows we hold the price of risk (proportional to  $b^w(\beta)$ ) fixed, and we simply modify the persistence of consumption growth under the pricing (null) model. In other words, we ask how easy different models are to distinguish from white noise, holding the price of risk under them fixed and varying their persistence.

The top row is the calibration from the main analysis, where persistence is equal to the time discount factor. As the degree of persistence falls, the investor's ability to reject the pricing model in a century-long sample rapidly improves. When the persistence is 0.99, as in Hansen and Sargent (2010), the pricing model is rejected 12.5 percent of the time – twice as often as our endogenous worst-case model. When the persistence falls to 0.94 as in Bansal and Yaron (2004), the pricing model is rejected 86.9 percent of the time.<sup>23</sup> The result that the persistence of the worst-case model should be equal to  $\beta$  is clearly key to ensuring that the model is difficult to reject in simulated data.

## 7.2 Historical aggregate price/dividend ratios

To try to compare the model more directly to historical data, we now ask how the price/dividend ratio implied by the model compares to what we observe empirically. A natural benchmark is to treat investors' 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 given observed consumption data. 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 by assuming dividends are 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

---

<sup>23</sup>To be clear, the statistics reported in this table do not constitute a rejection of Bansal and Yaron's (2004) calibration. Rather, they just quantify the statistical difference between their calibration and white noise.

the model matches the data.

Figure 4 plots the historical price/dividend ratio on the S&P 500 from Robert Shiller against the price/dividend ratios implied by consumption and dividend growth from 1882 to 2012 derived from the model (see the appendix). The consumption growth data is from Barro and Ursua (2010) (and extended by us to 2012), while the dividends also come from Shiller.<sup>24</sup>

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 data 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 model performs well in matching historical price/dividend ratios, at least up to the 1990s.

### 7.3 An AR(2) point estimate

To emphasize the robustness of the result that the ambiguity aversion we study here only affects the very low-frequency features of consumption growth, we now analyze a case where the point estimate for consumption dynamics is an AR(2) model. The AR(2) process we examine in this section implies a hump in the spectral density at an intermediate frequency. This value of  $\bar{b}$  generates power at business-cycle frequencies as in many economic time series of interest (see Baxter and King (1999), for example), as illustrated by the spectral density shown in figure 5. This alternative calibration implies that the term  $\bar{f}(\omega)$  in (13) varies across frequencies, which allows us to ask whether variation in estimation uncertainty across frequencies is quantitatively relevant in determining the worst-case model.

---

<sup>24</sup>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 the non-linearities they included.



We first consider what worst-case model the agent would derive if she were constrained to minimize utility with respect to a transfer function implied by an AR(2). That is, utility is minimized by choosing a worst-case  $\{a_1, a_2, b_0\}$  in the model

$$\Delta c_t = \mu + a_1 (\Delta c_{t-1} - \mu) + a_2 (\Delta c_{t-2} - \mu) + b_0 \varepsilon_t \quad (28)$$

This is the usual approach taken in the literature; it assumes that investors know the model driving consumption growth and they need only estimate its parameters.<sup>25</sup>

We hold the value of  $\lambda$  fixed at its calibration from the previous section, to ensure that the results are comparable. We also assume that when the agent chooses a worst-case model, she still uses the penalty function  $g(b; \bar{b})$ . The only difference is that  $b^w$  must be an AR(2), so the optimization problem is highly restricted.

The worst case that emerges implies the real transfer function,  $B_r^w(\omega)$ , plotted in figure 6; we refer to it as the parametric worst case. The parametric worst case is essentially indistinguishable from the point estimate – it is the gray line that lines up nearly perfectly with the black line representing the point estimate 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.<sup>26</sup>

When we allow the agent to choose an unrestricted worst-case model, the outcome is very similar to what we obtained for the white-noise benchmark, as shown in figure 6. The figure is dominated by the non-parametric worst-case mainly deviating from the benchmark at very low frequencies. Again this reflects  $Z(\omega)$  being small at all but the lowest frequencies, which implies that the worst-case leaves  $B_r(\omega)$  essentially unchanged at all but very low frequencies. The worst-

<sup>25</sup>See Collin-Dufresne, Johannes, and Lochstoer (2013) for a deep analysis of the case of parameter uncertainty with Epstein–Zin preferences. Andrei, Carlin, and Hasler (2013) study a model in which investors disagree about the persistence of trend growth rates.

<sup>26</sup>The specific parameters in the benchmark and parametric worst-case models are (dropping  $\mu$  for legibility)

$$\text{Benchmark:} \quad \Delta c_t = (0.6974) \times \Delta c_{t-1} + (-0.34) \times \Delta c_{t-2} + (0.0118) \times \varepsilon_t \quad (29)$$

$$\text{Parametric worst case:} \quad \Delta c_t = (0.6996) \times \Delta c_{t-1} + (-0.3478) \Delta c_{t-2} + (0.0118) \times \varepsilon_t \quad (30)$$

The parametric worst-case model is thus nearly identical to the benchmark model. Under the benchmark,  $\bar{b}(\beta) = 0.01807$ , while under the worst case,  $b^w(\beta) = 0.01825$ . 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. In contrast, under the unconstrained non-parametric worst case  $b^w(\beta) = 0.04049$ .

case thus inherits the local peak in power at middle frequencies that we observe in the benchmark AR(2).

## 8 Endogenous consumption

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

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

$$W_{t+1} = \exp(r_t) W_t - \exp(c_t) \tag{31}$$

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

$$r_t = \mu_{ret} + b_{ret}(L) \varepsilon_t \tag{32}$$

$$\varepsilon_t \sim N(0, 1) \tag{33}$$

As is well known, with a unit EIS, an agent will always consume a constant fraction of current wealth, regardless of expectations for future returns. In our case, optimal consumption is,

$$\exp(c_t) = (1 - \beta) \exp(r_t) W_t \tag{34}$$

The income effect from higher future expected returns is exactly offset by the substitution effect caused by the increased price of consumption in the current period compared to future periods, so consumption is invariant to expected future returns.

Using the budget constraint, the consumption function (34) implies

$$\Delta c_t = r_t + \log \beta \tag{35}$$

Consumption growth itself then directly inherits the dynamics of returns: if investors believe that real returns to investment are persistent, then they also believe that consumption growth is persistent.

A common metaphor in the ambiguity aversion literature is that people play a game with an evil agent who chooses the worst-case model for returns conditional on the consumption policy the agent chooses. If that game has a Nash equilibrium, then our investor's consumption policy must be optimal taking the worst-case model chosen by the evil agent as given. So the investor understands that the evil agent will choose a process  $b_{ret}^w(L)$ , and she chooses an optimal consumption policy, taking the return process as given. That optimal consumption policy is always to consume a constant fraction of wealth, regardless of  $b_{ret}^w(L)$ . Under her forecasting model, consumption growth then inherits the dynamics of  $b_{ret}^w(L)$  through equation (35).

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

$$v_t = c_t + \log(1 - \beta) + \frac{\beta}{1 - \beta} \log \beta + \frac{\beta}{1 - \beta} \frac{(1 - \alpha)}{2} b_{ret}^w(\beta)^2 \tag{36}$$

If the worst-case process is again chosen as a minimization over lifetime utility, the worst-case model  $b_{ret}^w(L)$  takes exactly the same form as in the endowment economy since the utility cost of uncertainty is again determined by  $b_{ret}^w(\beta)^2$ . Our analysis of worst-case persistence in consumption growth can thus also be interpreted as worst-case persistence in the returns to financial or real investment with endogenous consumption.

## 9 Conclusion

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

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

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

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

## References

- Andrew B. Abel. Risk premia and term premia in general equilibrium. *Journal of Monetary Economics*, 43:3–33, 1999.
- Daniel Andrei, Bruce Carlin, and Michael Hasler. Model disagreement, volatility, and trading volume. Working paper, 2013.
- Donald W. K. Andrews and Werner Ploberger. Testing for serial correlation against an arma(1, 1,) process. *Journal of the American Statistical Association*, 91(435):pp. 1331–1342, 1996. ISSN 01621459. URL <http://www.jstor.org/stable/2291751>.
- Ravi Bansal and Amir Yaron. Risks for the long-run: A potential resolution of asset pricing puzzles. *Journal of Finance*, 59(4):1481–1509, August 2004.
- Ravi Bansal, Dana Kiku, and Amir Yaron. An empirical evaluation of the long-run risks model for asset prices. *Critical Finance Review*, 1(1):183–221, January 2012. URL <http://ideas.repec.org/a/now/jnlcfr/104.00000005.html>.
- Francisco Barillas, Lars Peter Hansen, and Thomas J. Sargent. Doubts or variability? *Journal of Economic Theory*, 144(6):2388–2418, November 2009.
- Robert J. Barro and Jose F. Ursua. Barro–ursua macroeconomic dataset. <http://scholar.harvard.edu/barro/publications/barro-ursua-macroeconomic-data>, 2010.
- Robert B. Barsky and Bradford De Long. Why does the stock market fluctuate? *Quarterly Journal of Economics*, 108(2):291–311, May 1993.
- Marianne Baxter and Robert G. King. Measuring business cycles: Approximate band-pass filters for economic time series. *The Review of Economics and Statistics*, 81(4):575–593, November 1999. URL <http://ideas.repec.org/a/tpr/restat/v81y1999i4p575-593.html>.
- Jason Beeler and John Y. Campbell. The long-run risks model and aggregate asset prices: An empirical assessment. *Critical Finance Review*, 1(1):141–182, January 2012. URL <http://ideas.repec.org/a/now/jnlcfr/104.00000004.html>.

- Rhys Bidder and Matthew E. Smith. Doubts and variability: A robust perspective on exotic consumption series. Federal Reserve Bank of San Francisco, Working Paper No. 2013-28, 2013.
- Michael J. Brennan and Yihong Xia. Stock price volatility and the equity premium. *Journal of Monetary Economics*, 47:249–283, 2001.
- John Y. Campbell and John H. Cochrane. By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy*, 107:205–251, 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 too 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.
- Pierre Collin-Dufresne, Michael Johannes, and Lars A. Lochstoer. Parameter learning in general equilibrium: The asset pricing implications. Working paper, 2013.
- Massimiliano Croce, Martin Lettau, and Sydney C. Ludvigson. Investor information, long-run risk, and the term structure of equity. *Review of Financial Studies*, 2014.
- 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>.
- Martin Evans. Real rates, expected inflation and inflation risk premia. *Journal of Finance*, 53(1): 187–218, 1998.
- 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>.
- Robin Greenwood and Andrei Shleifer. Expectations of returns and expected returns. *Review of Financial Studies*, 27(3):714–746, 2014.
- 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>.
- David Hirshleifer and Jianfeng Yu. Asset pricing in production economies with extrapolative expectations. Working Paper, 2012.
- Yongmiao Hong. Consistent testing for serial correlation of unknown form. *Econometrica*, 64(4): 837–864, 1996.
- Nengjiu Ju and Jianjun Miao. Ambiguity, learning, and asset returns. *Econometrica*, 80(2):559–591, 2012. ISSN 1468-0262. doi: 10.3982/ECTA7618. URL <http://dx.doi.org/10.3982/ECTA7618>.

- Ralph Koijen, Maik Schmeling, and Evert B. Vrugt. Survey expectations of returns and asset pricing puzzles. Working paper, 2014.
- 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, 1983–1997*. Cambridge University Press, 1998.
- David M. Kreps and Evan L. Porteus. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica*, 46(1):185–200, 1978.
- Markus Leippold, Fabio Trojani, and Paolo Vanini. Learning and asset prices under ambiguous information. *Review of Financial Studies*, 21(6):2656–2597, 2008.
- Jun Liu, Jun Pan, and Tan Wang. An equilibrium model of rare-event premia and its implication for option smirks. *Review of Financial Studies*, 18(1):131–164, 2004.
- G. M. Ljung and G. E. P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):pp. 297–303, 1978. ISSN 00063444. URL <http://www.jstor.org/stable/2335207>.
- Fabio Maccheroni, Massimo Marinacci, and Aldo Rustichini. Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 74(6):1447–1498, 2006.
- Mark J. Machina and Marciano Siniscalchi. Ambiguity and ambiguity aversion. In *Handbook of the Economics of Risk and Uncertainty*, pages 729–807. Elsevier, 2014.
- Pascal J. Maenhout. Robuuncertainty rules and asset pricing. *Review of Financial Studies*, 17(4): 951–983, 2004.
- Ulrich K. Müller and Mark W. Watson. Measuring uncertainty about long-run predictions. Working paper, 2013.
- Bryan R. Routledge and Stanley E. Zin. Model uncertainty and liquidity. *Review of Economic Dynamics*, 12(4):543–566, 2009.
- Alessandro Sbuelz and Fabio Trojani. Asset prices with locally constrained-entropy recursive multiple-priors utility. *Journal of Economic Dynamics & Control*, 32:3695–3717, 2008.



- Costis Skiadas. Smooth ambiguity aversion toward small risks and continuous-time recursive utility. *Journal of Political Economy*, 121(4):775–792, 2013.
- James H. Stock. Unit roots, structural breaks, and trends. In *Handbook of Econometrics*. Elsevier, 1994.
- Raman Uppal and Tan Wang. Model misspecification and underdiversification. *Journal of Finance*, 58(6):2465–2486, 2003.
- Pietro Veronesi. How does information quality affect stock returns? *Journal of Finance*, 55(2): 807–837, 2000.
- Philippe Weil. The equity premium puzzle and the risk-free rate puzzle. *Journal of Monetary Economics*, 24(3):401–421, 1989.

# Contents

<b>A Asymptotic distribution of the transfer function for non-parametric estimators (assumption 1)</b>	<b>34</b>
<b>B Alternative time-domain derivation of <math>g(b; \bar{b})</math> (assumption 1)</b>	<b>36</b>
<b>C Lifetime utility (assumption 2)</b>	<b>37</b>
<b>D Proposition 1</b>	<b>38</b>
D.1 Time-domain version . . . . .	39
<b>E Solution for non-normal innovations (proposition 1)</b>	<b>39</b>
E.1 Lifetime utility . . . . .	39
E.2 Worst-case transfer function . . . . .	40
<b>F Long-run risk is the worst-case (corollary 1)</b>	<b>41</b>
<b>G Asset prices and expected returns</b>	<b>43</b>
G.1 The risk-free rate . . . . .	44
G.2 Expected excess returns . . . . .	45
G.3 The behavior of interest rates . . . . .	45
G.4 Expectations of returns and expected returns . . . . .	46
G.5 Results for the white-noise case for the calibration . . . . .	47
<b>H Linking <math>\alpha</math> and <math>\lambda</math> (equation 26)</b>	<b>48</b>
H.1 KL distance . . . . .	48
H.2 Spectral distance . . . . .	49
<b>I Endogenous consumption</b>	<b>52</b>
 <b>A Asymptotic distribution of the transfer function for non-parametric estimators (assumption 1)</b>	

This section derives an asymptotic distribution for estimates of the transfer function,  $B(\omega)$ . Berk provides a distribution theory for the spectral density, and we show that it can be modified to provide results for the transfer function.

Berk (1974) studies estimates of the spectral density based on AR( $p$ ) processes. He shows (theorem 5) that if  $p$  grows with the sample size  $T$ , such that  $p \rightarrow \infty$  as  $T \rightarrow \infty$  and  $p^3/T \rightarrow 0$ , then the inverse of  $B(\omega)$  will be normally distributed around its true value. Specifically, define  $A(\omega) \equiv b_0 B(\omega)^{-1}$ , where  $B(\omega)$  is  $b(e^{i\omega})$  as in the main text.  $A(\omega)$  is the Fourier transform of the AR coefficients,  $A(\omega) = a(e^{i\omega})$ .

We consider the real and complex parts of  $A(\omega)$  separately,  $A(\omega) = a_r(\omega) + ia_i(\omega)$ . Berk shows

$$(T/p)^{1/2} \begin{bmatrix} a_r(\omega) - a_r^{True}(\omega) \\ a_i(\omega) - a_i^{True}(\omega) \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \frac{1}{4\pi} I_2 b_0^2 f^{True}(\omega)^{-1} \right) \quad (\text{A.1})$$

where,  $I_2$  is the  $2 \times 2$  identity matrix,  $a_r(\omega)$  and  $a_i(\omega)$  are the real and complex components of estimates of  $A$  and  $f(\omega)$  is the spectral density, and  $f(\omega) = B(\omega) B(\omega)^* = \frac{b_0^2}{A(\omega)A(\omega)^*}$ , with an asterisk denoting the complex conjugate. From here on we ignore the scaling factors  $(T/p)^{1/2}$  and  $\frac{1}{4\pi}$ .<sup>1</sup>

Now we have

$$b_0^{-1} B(\omega) = A^{-1}(\omega) \quad (\text{A.2})$$

$$= \frac{A(\omega)^*}{A(\omega) A(\omega)^*} \quad (\text{A.3})$$

$$= \frac{a_r(\omega) - ia_i(\omega)}{a_r(\omega)^2 + a_i(\omega)^2} \quad (\text{A.4})$$

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

$$B = b_r + ib_i \quad (\text{A.5})$$

$$b_r = \frac{a_r}{a_r^2 + a_i^2} b_0 \quad \text{and} \quad b_i = \frac{-a_i}{a_r^2 + a_i^2} b_0$$

The derivatives with respect to  $a_i$  and  $a_r$  are

$$\frac{db_r}{da_r} = \frac{a_i^2 - a_r^2}{(a_r^2 + a_i^2)^2} b_0 \quad (\text{A.6})$$

$$\frac{db_r}{da_i} = \frac{-2a_r a_i}{(a_r^2 + a_i^2)^2} b_0 \quad (\text{A.7})$$

$$\frac{db_i}{da_r} = \frac{2a_i a_r}{(a_r^2 + a_i^2)^2} b_0 \quad (\text{A.8})$$

$$\frac{db_i}{da_i} = \frac{a_i^2 - a_r^2}{(a_r^2 + a_i^2)^2} b_0 \quad (\text{A.9})$$

---

<sup>1</sup>In Berk's notation,  $\sigma^2 = b_0^2$ .

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

$$\begin{aligned} & \begin{bmatrix} \left( \frac{a_i^2 - a_r^2}{(a_r^2 + a_i^2)^2} \right)^2 + \left( \frac{-2a_r a_i}{(a_r^2 + a_i^2)^2} \right)^2 & \frac{-2a_r a_i}{(a_r^2 + a_i^2)^2} \frac{a_i^2 - a_r^2}{(a_r^2 + a_i^2)^2} + \frac{2a_i a_r}{(a_r^2 + a_i^2)^2} \frac{a_i^2 - a_r^2}{(a_r^2 + a_i^2)^2} \\ \frac{-2a_r a_i}{(a_r^2 + a_i^2)^2} \frac{a_i^2 - a_r^2}{(a_r^2 + a_i^2)^2} + \frac{2a_i a_r}{(a_r^2 + a_i^2)^2} \frac{a_i^2 - a_r^2}{(a_r^2 + a_i^2)^2} & \left( \frac{a_i^2 - a_r^2}{(a_r^2 + a_i^2)^2} \right)^2 + \left( \frac{2a_r a_i}{(a_r^2 + a_i^2)^2} \right)^2 \end{bmatrix} b_0^4 f(\omega)^{-1} \quad (\text{A.10}) \\ & = \begin{bmatrix} \frac{1}{(a_r^2 + a_i^2)} b_0^2 & 0 \\ 0 & \frac{1}{(a_r^2 + a_i^2)} b_0^2 \end{bmatrix} = I_2 f(\omega) \quad (\text{A.11}) \end{aligned}$$

where  $\frac{b_0^2}{(a_r^2 + a_i^2)} = \frac{b_0^2}{A(\omega)A(\omega)^*} = B(\omega)B(\omega)^* = f(\omega)$ . The two components of  $B(\omega)$  are thus independent with variances  $f^{True}(\omega)$ . Finally, then, a Wald statistic (again, ignoring scaling factors) for a particular  $B(\omega)$  is

$$\frac{(B(\omega) - B^{True}(\omega))(B(\omega) - B^{True}(\omega))^*}{f^{True}(\omega)} \quad (\text{A.12})$$

We construct  $g(b; \bar{b})$  by integrating the Wald statistic across frequencies.

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  $b_0^w = \bar{b}_0$  (which is a restriction on the integral of the transfer function  $B^w(\omega)$ ). The results are essentially unaffected by this constraint (which we know from the fact that in the calibration  $b_0^w$  is nearly identical to  $\bar{b}_0$ ).

## B Alternative time-domain derivation of $g(b; \bar{b})$ (assumption 1)

Brockwell and Davis (1988) show that for an MA model of order  $m$ , the coefficients are asymptotically normal with a covariance matrix denoted  $\Sigma_m$ . As  $m \rightarrow \infty$ ,  $\Sigma_m$  converges to a product,<sup>2</sup>

$$\Sigma_m \rightarrow J_m^{True} J_m^{True'} \quad (\text{B.1})$$

$$\text{where } J_m^{True} \equiv \begin{bmatrix} b_0^{True} & b_1^{True} & \dots & b_m^{True} \\ 0 & b_0^{True} & \dots & b_{m-1}^{True} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & b_0^{True} \end{bmatrix} \quad (\text{B.2})$$

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

$$m^{-1} (\mathbf{b}_{1:m} - \bar{\mathbf{b}}_{1:m}) (\bar{J}_m \bar{J}_m')^{-1} (\mathbf{b}_{1:m} - \bar{\mathbf{b}}_{1:m})' \quad (\text{B.3})$$

---

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

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

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

$$\bar{\Sigma}_m^{-1} \approx \Lambda_m \bar{F}_m^{-1} \Lambda_m^* \quad (\text{B.4})$$

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

$$\begin{aligned} m^{-1} (\mathbf{b}_{1:m} - \bar{\mathbf{b}}_{1:m}) \bar{\Sigma}_m^{-1} (\mathbf{b}_{1:m} - \bar{\mathbf{b}}_{1:m})' &\approx m^{-1} (\mathbf{B}_m \Lambda_m^* - \bar{\mathbf{B}}_m \Lambda_m^*) \Lambda_m \bar{F}_m^{-1} \Lambda_m^* (\mathbf{B}_m^* \Lambda_m' - \bar{\mathbf{B}}_m^* \Lambda_m') \\ &= m^{-1} (\mathbf{B}_m - \bar{\mathbf{B}}_m) \bar{F}_m^{-1} (\mathbf{B}_m - \bar{\mathbf{B}}_m)^* \end{aligned} \quad (\text{B.5}) \quad (\text{B.6})$$

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

$$m^{-1} (\mathbf{B}_m - \bar{\mathbf{B}}_m) \bar{F}_m^{-1} (\mathbf{B}_m - \bar{\mathbf{B}}_m)^* \rightarrow \int \frac{|B(\omega) - \bar{B}(\omega)|^2}{\bar{f}(\omega)} d\omega \quad (\text{B.7})$$

This section thus provides an alternative derivation of the quadratic distance measure based on a specific Wald statistic. An alternative and equivalent definition of the distance measure is  $g(b; \bar{b}) = (\mathbf{b} - \bar{\mathbf{b}}) \bar{\Sigma}^{-1} (\mathbf{b} - \bar{\mathbf{b}})'$ .

## C Lifetime utility (assumption 2)

As discussed in the text, the agent's expectation of future consumption growth,  $E_t [\Delta c_{t+j} | a, b_0]$  is equal to expected consumption growth at date  $t+j$  given the past observed history of consumption growth and the assumption that  $\varepsilon_t$  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

$$\varepsilon_t^{\{a, b_0\}} = (\Delta c_t - \mu - a(L)(\Delta c_{t-1} - \mu)) / b_0 \quad (\text{C.1})$$

That is,  $\varepsilon_t^{\{a, b_0\}}$  is the innovation that the agent would believe occurred given the observed history of consumption growth and the model  $\{a, b_0\}$ . The agent's subjective expectations for future consumption growth are then

$$E_t [\Delta c_{t+j} | a, b_0] = \mu + \sum_{j=0}^{\infty} b_{k+j} \varepsilon_{t-j}^{\{a, b_0\}} \quad (\text{C.2})$$

<sup>3</sup>Specifically,  $\bar{J}_m \approx \Lambda_m \bar{B}_m \Lambda_m^* = \Lambda_m^* \bar{B}_m^* \Lambda_m'$ , and thus  $\bar{J}_m \bar{J}_m' \approx \Lambda_m \bar{B}_m \Lambda_m^* \Lambda_m \bar{B}_m^* \Lambda_m' = \Lambda_m (\bar{B}_m \bar{B}_m^*) \Lambda_m^* = \Lambda_m \bar{F}_m \Lambda_m^*$ , where  $\bar{B}_m$  is the diagonal matrix of the discrete Fourier transform of  $[\bar{b}_0, \bar{b}_1, \dots, \bar{b}_m]$ . Again, the approximations become exact as  $m \rightarrow \infty$ .

with subjective distribution

$$\frac{\Delta c_{t+1} - E_t[\Delta c_{t+1}|a, b_0]}{b_0} \sim N(0, 1) \quad (\text{C.3})$$

We guess that  $v(\Delta c^t; a, b_0)$  takes the form

$$v(\Delta c^t; a, b_0) = c_t + \bar{k} + \sum_{j=0}^{\infty} k_j \varepsilon_{t-j}^{\{a, b_0\}} \quad (\text{C.4})$$

Inserting into the recursion for lifetime utility yields

$$\bar{k} + \sum_{j=0}^{\infty} k_j \varepsilon_{t-j}^{\{a, b_0\}} = \frac{\beta}{1-\alpha} \log E_t \left[ \exp \left( \left( \bar{k} + \mu + \sum_{j=0}^{\infty} (k_j + b_j) \varepsilon_{t-j+1}^{\{a, b_0\}} \right) (1-\alpha) \right) | a, b_0 \right] \quad (\text{C.5})$$

$$= \beta (\bar{k} + \mu) + \beta \sum_{j=0}^{\infty} (k_{j+1} + b_{j+1}) \varepsilon_{t-j}^{\{a, b_0\}} + \beta \frac{1-\alpha}{2} (k_0 + b_0)^2 \quad (\text{C.6})$$

Matching the coefficients on each side of the equality yields

$$v(\Delta c^t; b) = c_t + \frac{\beta}{1-\beta} \frac{1-\alpha}{2} b (\beta)^2 + \frac{\beta}{1-\beta} \mu + \sum_{k=1}^{\infty} \beta^k \sum_{j=0}^{\infty} b_{j+k} \varepsilon_{t-j}^{\{a, b_0\}} \quad (\text{C.7})$$

$$= c_t + \frac{\beta}{1-\beta} \frac{1-\alpha}{2} b (\beta)^2 + \frac{\beta}{1-\beta} \mu + \sum_{j=0}^{\infty} \left( \sum_{k=1}^{\infty} \beta^k b_{j+k} \right) \varepsilon_{t-j}^{\{a, b_0\}} \quad (\text{C.8})$$

$$= c_t + \frac{\beta}{1-\beta} \frac{1-\alpha}{2} b (\beta)^2 + \sum_{k=1}^{\infty} \beta^k E_t[\Delta c_{t+k}|a, b_0] \quad (\text{C.9})$$

## D Proposition 1

The optimization problem is

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

We guess that

$$B^w(\omega) = \bar{B}(\omega) + k \bar{f}(\omega) Z(\omega)^* \quad (\text{D.3})$$

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

$$0 = 2\frac{\beta}{1-\beta}\frac{1-\alpha}{2}b^w(\beta)\beta^j + \lambda \int \frac{\exp(i\omega j)(B(\omega) - \bar{B}(\omega))^* + \exp(-i\omega j)(B(\omega) - \bar{B}(\omega))}{\bar{f}(\omega)} d\omega \quad (\text{D.4})$$

$$= 2\frac{\beta}{1-\beta}\frac{1-\alpha}{2}b^w(\beta)\beta^j + \lambda \int \frac{\exp(i\omega j)k\bar{f}(\omega)Z(\omega) + \exp(-i\omega j)k\bar{f}(\omega)Z(\omega)^*}{\bar{f}(\omega)} d\omega \quad (\text{D.5})$$

$$= 2\frac{\beta}{1-\beta}\frac{1-\alpha}{2}b^w(\beta)\beta^j + 2\lambda k\beta^j \quad (\text{D.6})$$

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 the correct solution if

$$k = -\lambda^{-1}\frac{\beta}{1-\beta}\frac{1-\alpha}{2}b^w(\beta) \quad (\text{D.7})$$

which is the result from the text,

$$B^w(\omega) = \bar{B}(\omega) + \lambda^{-1}\frac{\beta}{1-\beta}\frac{\alpha-1}{2}b^w(\beta)\bar{f}(\omega)Z(\omega)^* \quad (\text{D.8})$$

## D.1 Time-domain version

The above result can also be derived using the time-domain Wald statistic. We have

$$\mathbf{b}^w = \arg \min_b \left\{ -\beta(1-\alpha)\mathbf{b}'zz'\mathbf{b} + \lambda(\mathbf{b} - \bar{\mathbf{b}})\bar{\Sigma}^{-1}(\mathbf{b} - \bar{\mathbf{b}})' \right\} \quad (\text{D.9})$$

where  $z \equiv [1, \beta, \beta^2, \dots]'$  and  $\bar{\Sigma}$  is the asymptotic covariance from Brockwell and Davis (1988). The solution is

$$(\mathbf{b}^w - \bar{\mathbf{b}})' = \lambda^{-1}\bar{\Sigma}\beta(1-\alpha)zb^w(\beta) \quad (\text{D.10})$$

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

## E Solution for non-normal innovations (proposition 1)

### E.1 Lifetime utility

We now assume that  $\varepsilon_t$  has an arbitrary distribution (with mean zero) that is characterized by a cumulant generating function  $\Gamma$ . We assume that  $\varepsilon_t$  is serially independent.

The recursion remains,

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

$$= \beta (\bar{k} + \mu) + \beta \sum_{j=0}^{\infty} (k_{j+1} + b_{j+1}) \varepsilon_{t-j}^{\{a, b_0\}} + \frac{\beta}{1-\alpha} \log E_t \left[ \exp \left( (k_0 + b_0) \varepsilon_{t+1}^{\{a, b_0\}} (1-\alpha) \right) \right] \quad (\text{E.2})$$

$$= \beta (\bar{k} + \mu) + \beta \sum_{j=0}^{\infty} (k_{j+1} + b_{j+1}) \varepsilon_{t-j}^{\{a, b_0\}} + \frac{\beta}{1-\alpha} \Gamma((k_0 + b_0)(1-\alpha)) \quad (\text{E.3})$$

Again, by matching coefficients, we obtain

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

## E.2 Worst-case transfer function

The optimization problem is now

$$\begin{aligned} B^w(\omega) &= \arg \min_{b(L)} \frac{\beta}{1-\beta} \frac{1}{1-\alpha} \Gamma(b^w(\beta)(1-\alpha)) + \lambda \int \frac{|B(\omega) - \bar{B}(\omega)|^2}{\bar{f}(\omega)} d\omega \quad (\text{E.4}) \\ &= \arg \min_{b(L)} \frac{\beta}{1-\beta} \frac{1}{1-\alpha} \Gamma(b^w(\beta)(1-\alpha)) + \lambda \int \frac{\left( \sum_{j=0}^{\infty} \exp(i\omega j) (b_j - \bar{b}_j) \right) \left( \sum_{j=0}^{\infty} \exp(-i\omega j) (b_j - \bar{b}_j) \right)}{\bar{f}(\omega)} d\omega \quad (\text{E.5}) \end{aligned}$$

We again guess that

$$B^w(\omega) = \bar{B}(\omega) + k \bar{f}(\omega) Z(\omega)^* \quad (\text{E.6})$$

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

$$0 = \frac{\beta}{1-\beta} \Gamma'(b^w(\beta)(1-\alpha)) \beta^j + \lambda \int \frac{\exp(i\omega j) (B(\omega) - \bar{B}(\omega))^* + \exp(-i\omega j) (B(\omega) - \bar{B}(\omega))}{\bar{f}(\omega)} d\omega \quad (\text{E.7})$$

$$= \frac{\beta}{1-\beta} \Gamma'(b^w(\beta)(1-\alpha)) \beta^j + \lambda \int \frac{\exp(i\omega j) k \bar{f}(\omega) Z(\omega) + \exp(-i\omega j) k \bar{f}(\omega) Z(\omega)^*}{\bar{f}(\omega)} d\omega \quad (\text{E.8})$$

$$= \frac{\beta}{1-\beta} \Gamma'(b^w(\beta)(1-\alpha)) \beta^j + 2\lambda k \beta^j \quad (\text{E.9})$$

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

$$k = \frac{-\lambda^{-1}}{2} \frac{\beta}{1-\beta} \Gamma'(b^w(\beta)(1-\alpha)) \quad (\text{E.10})$$

$$B^w(\omega) = \bar{B}(\omega) - \frac{\lambda^{-1}}{2} \frac{\beta}{1-\beta} \Gamma'(b^w(\beta)(1-\alpha)) \bar{f}(\omega) Z(\omega)^* \quad (\text{E.11})$$



Note also that the CGF for the standard normal distribution is  $\Gamma(x) = x^2/2$ , so  $\Gamma'(x) = x$ , so (E.10) reduces to (D.7) when  $\varepsilon$  is a standard normal.

## F Long-run risk is the worst-case (corollary 1)

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

For a white-noise point estimate,  $\bar{B}(\omega) = \bar{b}_0$  and  $\bar{f}(\omega) = \bar{b}_0^2$ . So then

$$\sum_{j=0}^{\infty} b_j^w e^{i\omega j} = \bar{b}_0 + \varphi \sum_{j=0}^{\infty} \beta^j e^{i\omega j} \quad (\text{F.2})$$

$$\varphi \equiv \lambda^{-1} \frac{\beta}{1-\beta} \frac{\alpha-1}{2} b^w(\beta) \times \bar{b}_0^2 \quad (\text{F.3})$$

This immediately yields the result in the text, repeated here:

$$b_0^w = \bar{b}_0 + \varphi \quad (\text{F.4})$$

$$b_j^w = \varphi \beta^j \text{ for } j > 0 \quad (\text{F.5})$$

To see more clearly the equivalence (in terms of sharing the same autocovariance structure) with the homoskedastic long-run risk system we here provide mappings between a long-run risk system to an *ARMA*(1, 1) reduced form and then from the MA coefficients derived above to a particular *ARMA*(1, 1).

A long run risk system takes the form of an unobserved components model where the series of interest,  $y_t$ , is given by an *AR*(1) plus ‘noise’

$$y_t = x_{t-1} + \eta_t \quad (\text{F.6})$$

$$x_t = \rho x_{t-1} + \nu_t \quad (\text{F.7})$$

$$\eta_t \sim N(0, \sigma_\eta^2)$$

$$\nu_t \sim N(0, \sigma_\nu^2)$$

We shall show that the autocovariances of  $y_t$  are identical to those of an *ARMA*(1, 1)

$$y_t = \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1} \quad (\text{F.8})$$

$$\varepsilon_t \sim N(0, \sigma^2)$$

for an appropriate mapping between  $\chi_{LRR} \equiv \{\rho, \sigma_\eta, \sigma_\nu\}$  and  $\chi_{11} \equiv \{\phi, \theta, \sigma\}$ .

We begin by noting that the LRR representation implies

$$y_t - \rho y_{t-1} = \nu_{t-1} + \eta_t - \rho \eta_{t-1} \quad (\text{F.9})$$

The term on the right hand side is a *MA*(1) plus ‘noise’, and therefore has a *MA*(1) reduced form.

Thus, the LRR model implies an  $ARMA(1,1)$  with an autoregressive root,  $\rho = \phi$ . We consider two representations for the right hand side

$$z_t = \nu_{t-1} + \eta_t - \rho\eta_{t-1} \quad (\text{F.10})$$

and

$$z_t = \varepsilon_t + \theta\varepsilon_{t-1} \quad (\text{F.11})$$

To obtain a mapping between  $\{\sigma_\eta, \sigma_\nu\}$  and  $\{\theta, \sigma\}$  we match the first two autocovariances, shown in the table below (note that  $z_t$  has zero mean).

	<i>Repr.1</i>	<i>Repr.2</i>
$E[z_t^2]$	$(1 + \rho^2) \sigma_\eta^2 + \sigma_\nu^2$	$(1 + \theta^2) \sigma^2$
$E[z_t z_{t-1}]$	$-\rho \sigma_\eta^2$	$\theta \sigma^2$

Matching the first order autocorrelation we have

$$\frac{\theta}{1 + \theta^2} = \kappa \equiv \frac{-\rho \sigma_\eta^2}{(1 + \rho^2) \sigma_\eta^2 + \sigma_\nu^2}$$

so that  $\theta$  solves the quadratic (assuming invertibility)

$$\begin{aligned} 0 &= a\theta^2 + b\theta + c \\ a &= \kappa \\ b &= -1 \\ c &= \kappa \end{aligned}$$

We can then obtain  $\sigma$  as

$$\sigma = \sigma_\eta \sqrt{-\frac{\rho}{\theta}} \quad (\text{F.12})$$

which completes our mapping between  $\chi_{LRR}$  and  $\chi_{11}$ .

Now, we can obtain the  $ARMA(1,1)$  parameters implied by  $b^w$  by matching  $MA$  coefficients:

	<i>Repr.1</i>	<i>Repr.2</i>
$b_0^w$	$\bar{b}_0 + \varphi$	$\sigma$
$b_1^w$	$\beta\varphi$	$\sigma(\theta + \phi)$
$b_j^w$	$\beta b_{j-1}^w$	$\phi b_{j-1}^w$

Consequently we have that

$$\phi = \beta \tag{F.13}$$

$$\theta = -\frac{\bar{b}\beta}{\bar{b}_0 + \varphi} \tag{F.14}$$

$$\sigma = \bar{b}_0 + \varphi \tag{F.15}$$

With these mappings between  $b^w$  and the  $ARMA(1,1)$  representation and between the  $ARMA(1,1)$  and LRR representations we can obtain the particular LRR system described in the text.

## G Asset prices and expected returns

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

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

where  $\delta$  is a linearization parameter slightly less than 1.

We guess that

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

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

The innovation to lifetime utility is

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

$$= b^w(\beta) \varepsilon_{t+1}^{\{a,b_0\}} \tag{G.4}$$

where the investor prices assets as though  $\varepsilon_{t+1}^{\{a,b_0\}}$  is a standard normal.

The pricing kernel can therefore be written as

$$M_{t+1} = \beta \exp \left( -\Delta c_{t+1} + (1 - \alpha) b^w(\beta) \varepsilon_{t+1}^{\{a,b_0\}} - \frac{(1 - \alpha)^2}{2} b^w(\beta)^2 \right) \tag{G.5}$$

The pricing equation for the levered consumption claim is

$$0 = \log E_t \left[ \beta \exp \left( \begin{aligned} & \delta_0 + (\delta - 1) \bar{h} + (\delta h_0 + \gamma - 1) \Delta c_{t+1} + \sum_{j=0}^{\infty} (\delta h_{j+1} - h_j) \Delta c_{t-j} \\ & + (1 - \alpha) b^w (\beta) \varepsilon_{t+1}^{\{a, b_0\}} - \frac{(1-\alpha)^2}{2} b^w (\beta)^2 \end{aligned} \right) | b^w \right] \quad (\text{G.6})$$

$$= (\delta h_0 + \gamma - 1) a^w (L) \Delta c_t + \sum_{j=0}^{\infty} (\delta h_{j+1} - h_j) \Delta c_{t-j} \quad (\text{G.7})$$

$$+ \delta_0 + \left( \frac{1}{2} (\delta h_0 + \gamma - 1)^2 (b_0^w)^2 + (\delta h_0 + \gamma - 1) (1 - \alpha) b^w (\beta) b_0^w \right) + (\delta - 1) \bar{h} + \log \beta \quad (\text{G.8})$$

Matching coefficients on  $\Delta c_{t-j}$  and on the constant yields two equations,

$$(\delta - 1) \bar{h} + \log \beta = -\delta_0 - \left( \frac{1}{2} (\delta h_0 + \gamma - 1)^2 (b_0^w)^2 + (\delta h_0 + \gamma - 1) (1 - \alpha) b^w (\beta) b_0^w \right) \quad (\text{G.9})$$

$$(\delta h_{j+1} - h_j) = -(\delta h_0 + \gamma - 1) a_j^w \quad (\text{G.10})$$

And thus

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

and

$$\bar{h} = \frac{1}{1 - \delta} \left[ \log \beta + \delta_0 + \left( \frac{1}{2} (\delta h_0 + \gamma - 1)^2 (b_0^w)^2 + (\delta h_0 + \gamma - 1) (1 - \alpha) b^w (\beta) b_0^w \right) \right] \quad (\text{G.11})$$

### G.1 The risk-free rate

For the risk-free rate, we have

$$r_{f,t+1} = -\log E_t \left[ \beta \exp \left( -\Delta c_{t+1} + (1 - \alpha) b^w (\beta) \varepsilon_{t+1}^{\{a, b_0\}} - \frac{(1 - \alpha)^2}{2} b^w (\beta)^2 \right) | b^w \right] \quad (\text{G.12})$$

$$= -\log \beta + a^w (L) \Delta c_t - \frac{1}{2} (b_0^w)^2 + b_0^w (1 - \alpha) b^w (\beta) \quad (\text{G.13})$$

## G.2 Expected excess returns

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

$$E_t [r_{t+1} - r_{f,t+1} | \bar{b}] = E_t \left[ \begin{array}{l} \delta_0 + (\delta - 1) \bar{h} + (\delta h_0 + \gamma) \Delta c_{t+1} - \sum_{j=0}^{\infty} (\delta h_0 + \gamma - 1) a_j^w \Delta c_{t-j} \\ + \log \beta - a^w(L) \Delta c_t + \frac{1}{2} (b_0^w)^2 - b_0^w (1 - \alpha) b^w(\beta) \end{array} \middle| \bar{b} \right] \quad (\text{G.14})$$

$$= (\delta h_0 + \gamma) (a(L) - a^w(L)) \Delta c_t + \frac{1}{2} (b_0^w)^2 + \delta_0 + (\delta - 1) \bar{h} + \log \beta - b_0^w (1 - \alpha) b^w(\beta) \quad (\text{G.15})$$

$$= (\delta h_0 + \gamma) (a(L) - a^w(L)) \Delta c_t + \frac{1}{2} (b_0^w)^2 + \delta_0 \quad (\text{G.16})$$

$$- \delta_0 - \left( \frac{1}{2} (\delta h_0 + \gamma - 1)^2 (b_0^w)^2 + (\delta h_0 + \gamma - 1) (1 - \alpha) b^w(\beta) b_0^w \right) - b_0^w (1 - \alpha) b^w(\beta) \quad (\text{G.17})$$

$$= (\delta h_0 + \gamma) (a(L) - a^w(L)) \Delta c_t + \frac{1}{2} (b_0^w)^2 \quad (\text{G.18})$$

$$- \left( \frac{1}{2} (\delta h_0 + \gamma)^2 (b_0^w)^2 + \frac{1}{2} (b_0^w)^2 - (\delta h_0 + \gamma) (b_0^w)^2 \right) - b_0^w (1 - \alpha) b^w(\beta) + (\delta h_0 + \gamma - 1) (1 - \alpha) b^w(\beta) b_0^w \quad (\text{G.19})$$

$$= (\delta h_0 + \gamma) (a(L) - a^w(L)) \Delta c_t \quad (\text{G.20})$$

$$- \left( \frac{1}{2} (\delta h_0 + \gamma)^2 (b_0^w)^2 - (\delta h_0 + \gamma) (b_0^w)^2 + (\delta h_0 + \gamma) (1 - \alpha) b^w(\beta) b_0^w \right) \quad (\text{G.21})$$

$$= (\delta h_0 + \gamma) (a(L) - a^w(L)) \Delta c_t - \text{cov}(m_{t+1}, r_{t+1}) - \frac{1}{2} \text{var}(r_{t+1}) \quad (\text{G.22})$$

Where

$$\text{var}_t^w(r_{t+1}) = (\delta h_0 + \gamma)^2 (b_0^w)^2 \quad (\text{G.23})$$

$$\text{cov}(r_{t+1}, m_{t+1}) = (\delta h_0 + \gamma) b_0^w (-b_0^w + (1 - \alpha) b^w(\beta)) \quad (\text{G.24})$$

$$= -(\delta h_0 + \gamma) (b_0^w)^2 + (\delta h_0 + \gamma) (1 - \alpha) b^w(\beta) b_0^w \quad (\text{G.25})$$

Substituting in

$$\delta h_0 + \gamma = \delta \frac{(\gamma - 1) a^w(\delta)}{1 - \delta a^w(\delta)} + \gamma = \frac{\gamma - \delta a^w(\delta)}{1 - \delta a^w(\delta)} \quad (\text{G.26})$$

yields the result from the text.

## G.3 The behavior of interest rates

The mean of the risk-free rate is

$$- \log \beta - \frac{1}{2} (b_0^w)^2 + b_0^w (1 - \alpha) b^w(\beta) \quad (\text{G.27})$$

And its standard deviation is

$$\text{std}(a^w(L) \Delta c_t) \quad (\text{G.28})$$

When consumption growth is white noise, this is

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

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

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

$$p_{j,t} = \phi^{(j)}(L) \Delta c_t + n_j \quad (\text{G.31})$$

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

The pricing condition for a bond is

$$M_{t+1} = \beta \exp \left( -\Delta c_{t+1} + (1 - \alpha) b^w(\beta) \varepsilon_{t+1}^{\{a, b_0\}} - \frac{(1 - \alpha)^2}{2} b^w(\beta)^2 \right) \quad (\text{G.32})$$

$$m^{(j)}(L) \Delta c_t + n_j = \log E_t \left[ \exp \left( \frac{\log \beta - \Delta c_{t+1} + (1 - \alpha) b^w(\beta) \varepsilon_{t+1}^{\{a, b_0\}}}{-\frac{(1 - \alpha)^2}{2} b^w(\beta)^2 + \phi^{(j-1)}(L) \Delta c_{t+1} + n_{j-1}} \right) | b^w \right] \quad (\text{G.33})$$

$$= \log \beta - a^w(L) \Delta c_t - \frac{(1 - \alpha)^2}{2} b^w(\beta)^2 + \sum_{k=0}^{\infty} \phi_{k+1}^{(j-1)} \Delta c_{t-k} + n_{j-1} \quad (\text{G.34})$$

$$+ \frac{1}{2} \left( \left( \phi_0^{(j-1)} - 1 \right) b_0^w + (1 - \alpha) b^w(\beta) \right)^2 \quad (\text{G.35})$$

Matching coefficients yields,

$$\phi^{(j)}(L) = -a^w(L) + \sum_{k=0}^{\infty} \phi_{k+1}^{(j-1)} L^k \quad (\text{G.36})$$

$$n_j = \log \beta - \frac{(1 - \alpha)^2}{2} b^w(\beta)^2 + n_{j-1} + \frac{1}{2} \left( \left( \phi_0^{(j-1)} - 1 \right) b_0^w + (1 - \alpha) b^w(\beta) \right)^2 \quad (\text{G.37})$$

We also have the boundary condition that the price of a unit of consumption today is 1, so that  $n_0 = 0$  and  $m^{(0)}(L) = 0$ .

#### G.4 Expectations of returns and expected returns

The risk-free rate is

$$r_{f,t+1} = -\log \beta + \mu + a^w(L) (\Delta c_t - \mu) - \frac{1}{2} (b_0^w)^2 + b_0^w b^w(\beta) (1 - \alpha) \quad (\text{G.38})$$

Expected returns are

$$E_t [r_{t+1} | \bar{a}, \bar{b}_0] = \frac{\gamma - \delta a^w(\delta)}{1 - \delta a^w(\delta)} (\bar{a}(L) - a^w(L)) (\Delta c_t - \mu) +$$

$$-\frac{\gamma - \delta a^w(\delta)}{1 - \delta a^w(\delta)} b_0^w [b_0^w - (1 - \alpha) b^w(\beta)] - \frac{1}{2} \left( \frac{\gamma - \delta a^w(\delta)}{1 - \delta a^w(\delta)} \right)^2 (b_0^w)^2 \quad (\text{G.39})$$

$$-\log \beta + \mu + a^w(L) (\Delta c_t - \mu) - \frac{1}{2} (b_0^w)^2 + b_0^w b^w(\beta) (1 - \alpha) \quad (\text{G.40})$$

Expected returns from the perspective of the investor are

$$E_t [r_{t+1} | b^w] = \frac{\gamma - \delta a^w(\delta)}{1 - \delta a^w(\delta)} b_0^w [b_0^w - (1 - \alpha) b^w(\beta)] - \frac{1}{2} \left( \frac{\gamma - \delta a^w(\delta)}{1 - \delta a^w(\delta)} \right)^2 (b_0^w)^2 \quad (\text{G.41})$$

$$-\log \beta + \mu + a^w(L) (\Delta c_t - \mu) - \frac{1}{2} (b_0^w)^2 + b_0^w b^w(\beta) (1 - \alpha) \quad (\text{G.42})$$

The realized return is

$$r_{t+1} = \delta_0 + (\delta - 1) \bar{h} + (\delta h_0 + \gamma) \Delta c_{t+1} - \sum_{j=0}^{\infty} (\delta h_0 + \gamma - 1) a_j^w \Delta c_{t-j} \quad (\text{G.43})$$

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

and

$$\bar{h} = \frac{1}{1 - \delta} \left[ \log \beta + \delta_0 + \left( \frac{1}{2} (\delta h_0 + \gamma - 1)^2 (b_0^w)^2 + (\delta h_0 + \gamma - 1) (1 - \alpha) b^w(\beta) b_0^w \right) \right] \quad (\text{G.44})$$

## G.5 Results for the white-noise case for the calibration

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

$$\Delta c_t = \beta \Delta c_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1} \quad (\text{G.45})$$

$$a^w(L) = (\beta - \theta) \sum_{j=0}^{\infty} \theta^j L^j \quad (\text{G.46})$$

where  $\theta \equiv \beta (1 - \varphi \bar{b}_0^{-1})$  and  $\varphi \equiv \frac{\beta}{1 - \beta} \frac{\alpha - 1}{2} b^w(\beta) \bar{b}_0^2 \lambda^{-1}$ .

$$a^w(\delta) = \frac{\beta - \theta}{1 - \theta \delta} \quad (\text{G.47})$$

$$b_j = \beta^j (\beta - \theta) \quad (\text{G.48})$$

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

$$\frac{(\delta k_0 + \gamma - 1)(\beta - \theta)}{1 - \delta\theta} \frac{\sigma_{\Delta c}}{\sqrt{1 - \theta^2}} \quad (\text{G.49})$$

## H Linking $\alpha$ and $\lambda$ (equation 26)

### H.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  $\varepsilon_t$  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  $\alpha$ . The coefficient of relative risk aversion in Epstein–Zin preferences,  $\alpha$ , 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

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

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

$$KL(h, \tilde{h}) \equiv E_t \left[ \frac{\tilde{h}(\varepsilon_{t+1})}{h(\varepsilon_{t+1})} \log \frac{\tilde{h}(\varepsilon_{t+1})}{h(\varepsilon_{t+1})} \right] \quad (\text{H.2})$$

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

Consider the case where

$$\varepsilon_t \sim {}_h N(0, \sigma^2) \quad (\text{H.3})$$

$$\varepsilon_t \sim_{\tilde{h}} N(\mu, \sigma^2) \quad (\text{H.4})$$

The KL distance is then  $\frac{1}{2} \frac{\mu^2}{\sigma^2}$ . The KL distance here is related 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  $\varepsilon_t$  in each of those  $m$  pieces as  $\varepsilon_i$ . We have

$$\varepsilon_i \sim {}_h N \left( 0, \frac{T}{m} \sigma^2 \right) \quad (\text{H.5})$$

$$\varepsilon_i \sim {}_{\tilde{h}} N \left( \frac{T}{m} \mu, \frac{T}{m} \sigma^2 \right) \quad (\text{H.6})$$

So then

$$\sum_{i=1}^m \left( \frac{\varepsilon_i}{\left(\frac{T}{m}\right)^{1/2} \sigma} \right)^2 \sim {}_h \chi_m^2(0) \quad (\text{H.7})$$

$$\sum_{i=1}^m \left( \frac{\varepsilon_i}{\left(\frac{T}{m}\right)^{1/2} \sigma} \right)^2 \sim {}_{\tilde{h}} \chi_m^2 \left( m \left( \frac{\frac{T}{m} \mu}{\left(\frac{T}{m}\right)^{1/2} \sigma} \right)^2 \right) = \chi_m^2 \left( T \frac{\mu^2}{\sigma^2} \right) \quad (\text{H.8})$$

where  $\chi_m^2(k)$  denotes a non-central  $\chi^2$  variable with  $m$  degrees of freedom and non-centrality parameter  $k$ . Therefore, the standard  $\chi^2$  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  $\chi_m^2$  with non-centrality parameter  $T \frac{\mu^2}{\sigma^2}$ . Note that the non-centrality parameter is proportional to the sample size, showing how larger samples make it easier to reject the null given a fixed alternative. Moreover, the non-centrality parameter does not depend on  $m$ .

Finally, we have

$$\frac{1}{(\alpha-1)} KL(h, \tilde{h}) = \frac{1}{(\alpha-1)} \frac{1}{2} \frac{\mu^2}{\sigma^2} \quad (\text{H.9})$$

$$= \frac{1}{(\alpha-1)} \frac{1}{T} \frac{1}{2} \left( T \frac{\mu^2}{\sigma^2} \right) \quad (\text{H.10})$$

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

## H.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  $b(L)$ ,

$$x_t = b(L) \varepsilon_t \quad (\text{H.11})$$

where  $\varepsilon_t$  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

$$\frac{1}{\bar{b}(L)} x_t = \frac{b(L)}{\bar{b}(L)} \varepsilon_t \quad (\text{H.12})$$

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

$$\frac{1}{m} \sum_{j=1}^m \hat{d}_j^2 \quad (\text{H.13})$$

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  $m, T \rightarrow \infty$  (at the appropriate joint rate),

$$T^{1/2} \hat{d}_j \Rightarrow N(0, 1) \quad (\text{H.14})$$

and thus

$$\frac{1}{m} \sum_{j=1}^m T \hat{d}_j^2 \Rightarrow \chi_m^2 \quad (\text{H.15})$$

under the null.

More generally, Brockwell and Davis (1988) show that  $\hat{d}_j \Rightarrow N(\bar{d}_j, \text{var}(\bar{d}(L) \varepsilon_t))$ . So  $\frac{1}{m} \sum_{j=1}^m T \hat{d}_j^2$  is not exactly a non-central  $\chi^2$  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,  $\hat{\mathbf{d}} \equiv [\hat{d}_1, \hat{d}_2, \dots, \hat{d}_m]$ ,  $\bar{\mathbf{d}} \equiv [\bar{d}_1, \bar{d}_2, \dots]$ . 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  $\gamma$  (with elements  $\gamma_j$ ). 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),

$$T^{1/2} (\hat{\mathbf{d}}^T - \bar{\mathbf{d}}^T) \Rightarrow N(0, \Sigma_T) \quad (\text{H.16})$$

$$T^{1/2} \hat{\mathbf{d}}^T \Rightarrow N(T^{1/2} \bar{\mathbf{d}}^T, \Sigma_T) = N(\gamma, \Sigma_T) \quad (\text{H.17})$$

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

$$\text{cov}(x_t, x_{t-j}) = \sum_{k=0}^{\infty} \bar{d}_k^T \bar{d}_{k+j}^T \quad (\text{H.18})$$

$$= \gamma_j T^{-1/2} + T^{-1} \sum_{k=1}^{\infty} \gamma_k \gamma_{k+j} \quad (\text{H.19})$$

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

$$\text{cov}(x_t, x_t) = 1 + \sum_{k=1}^{\infty} (\bar{d}_k^T)^2 \quad (\text{H.20})$$

$$= 1 + T^{-1} \sum_{k=1}^{\infty} \gamma_k^2 \quad (\text{H.21})$$

We then define two matrices.  $\Omega_1$  has element  $(h, j)$  equal to  $\sum_{k=1}^{\infty} \gamma_k \gamma_{k+|h-j|}$  and  $\Omega_2$  has element  $(h, j)$  equal to  $\gamma_{|h-j|}$  if  $h - j \neq 0$ , and 0 if  $h = j$ . Finally, then

$$\Sigma_T = I + T^{-1}\Omega_1 + T^{-1/2}\Omega_2 \quad (\text{H.22})$$

We then have

$$T^{1/2}\hat{\mathbf{d}}^T \Rightarrow N\left(T^{1/2}\bar{\mathbf{d}}^T, I + \Omega_1 T^{-1} + \Omega_2 T^{-1/2}\right) \quad (\text{H.23})$$

$$T^{1/2}(\hat{\mathbf{d}}^T - \bar{\mathbf{d}}^T) \Rightarrow \varepsilon_I + \varepsilon_{\Omega_1} T^{-1/2} + \varepsilon_{\Omega_2} T^{-1/4} \quad (\text{H.24})$$

where  $\varepsilon_x$  is a mean-zero normally distributed vector of innovations with covariance matrix  $x$ , for  $x \in \{I, \Omega_1, \Omega_2\}$ . Therefore

$$T(\hat{\mathbf{d}}^T - \bar{\mathbf{d}}^T)(\hat{\mathbf{d}}^T - \bar{\mathbf{d}}^T)' = \varepsilon_I \varepsilon_I' + \varepsilon_{\Omega_1} \varepsilon_{\Omega_1}' T^{-1} + \varepsilon_{\Omega_2} \varepsilon_{\Omega_2}' T^{-1/2} \quad (\text{H.25})$$

$$+ 2\varepsilon_I \varepsilon_{\Omega_1}' T^{-1/2} + 2\varepsilon_I \varepsilon_{\Omega_2}' T^{-1/4} + 2\varepsilon_{\Omega_1} \varepsilon_{\Omega_2}' T^{-3/4} \quad (\text{H.26})$$

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

$$(T^{1/2}\hat{\mathbf{d}}^T - T^{1/2}\bar{\mathbf{d}}^T)(T^{1/2}\hat{\mathbf{d}}^T - T^{1/2}\bar{\mathbf{d}}^T)' \approx \varepsilon_I \varepsilon_I' = \chi_m^2(0) \quad (\text{H.27})$$

$$(T^{1/2}\hat{\mathbf{d}}^T)(T^{1/2}\hat{\mathbf{d}}^T)' \approx \chi_m^2(T\bar{\mathbf{d}}^T\bar{\mathbf{d}}^{T'}) \quad (\text{H.28})$$

That is, the test statistic is a non-central  $\chi_m^2$ , with non-centrality parameter  $T\bar{\mathbf{d}}^T\bar{\mathbf{d}}^{T'}$ .

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

$$\int |\bar{D}^T(\omega) - 1|^2 d\omega = \int |\bar{D}^T(\omega)|^2 - \bar{D}^T(\omega) - \bar{D}^T(\omega)^* + 1 d\omega \quad (\text{H.29})$$

$$= \int |\bar{D}^T(\omega)|^2 - 1 d\omega \quad (\text{H.30})$$

$$= \sum_{j=1}^{\infty} (\bar{d}_j^T)^2 = \bar{\mathbf{d}}^T\bar{\mathbf{d}}^{T'} \quad (\text{H.31})$$

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  $m$ ,  $\int |\bar{D}^T(\omega) - 1|^2 d\omega$  is equal to the non-centrality parameter in the

$\chi_m^2$  when you generate data under the model  $B^w$  and test the null that the data was driven by  $\bar{B}$ .

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

$$\int |\bar{D}^T(\omega) - 1|^2 d\omega = \int \left| \frac{b(L)}{\bar{b}(L)} - 1 \right|^2 d\omega \quad (\text{H.32})$$

$$= \int \left| \frac{b(L) - \bar{b}(L)}{\bar{b}(L)} \right|^2 d\omega \quad (\text{H.33})$$

$$= \int \frac{|b(L) - \bar{b}(L)|^2}{\bar{f}(L)} d\omega \quad (\text{H.34})$$

So we have that

$$g(b) = \sum_{j=1}^{\infty} (\bar{d}_j^T)^2 = \bar{\mathbf{d}}^T \bar{\mathbf{d}}^{T'} \quad (\text{H.35})$$

$$\lambda g(b) = \lambda \bar{\mathbf{d}}^T \bar{\mathbf{d}}^{T'} \quad (\text{H.36})$$

$$= \frac{\lambda}{T} (T \bar{\mathbf{d}}^T \bar{\mathbf{d}}^{T'}) \quad (\text{H.37})$$

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

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

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

Solving for  $\alpha$  yields

$$\alpha = 1 + \frac{1}{2\lambda(1-\beta)} \quad (\text{H.38})$$

## I Endogenous consumption

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

$$v_t = \max_{c_t} (1-\beta) c_t + \frac{\beta}{1-\alpha} \log E_t \exp((1-\alpha) v_{t+1}) \quad (\text{I.1})$$

Wealth follows

$$W_{t+1} = R_t W_t - C_t \quad (\text{I.2})$$

where  $C = \exp(c)$  and  $R = \exp(r)$ . Think of  $W_t$  measuring investment in some technology that shifts consumption across dates. It might be a financial asset or it might be a real investment project with payoff  $R_t$ . It might also represent storage.

Now suppose

$$r_t \equiv \log R_t = b(L) \varepsilon_t \quad (\text{I.3})$$

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

$$\begin{aligned} & \max_{c_t} (1 - \beta) c_t + \frac{\beta}{1 - \alpha} \log E_t \exp \left( (1 - \alpha) \left( \bar{v} + v_w (\log (R_t W_t - C_t) + r_{t+1}) + \sum_{j=0}^{\infty} v_j \varepsilon_{t+1-j} \right) \right) \quad (\text{I.4}) \\ & = \max_{c_t} (1 - \beta) c_t + \beta \left( \bar{v} + v_w \log (R_t W_t - C_t) + \sum_{j=1}^{\infty} (v_w b_j + v_j) \varepsilon_{t+1-j} \right) + \frac{1}{2} \beta (1 - \alpha) (v_w + v_0)^2 \quad (\text{I.5}) \end{aligned}$$

The first-order condition for consumption is

$$\frac{1 - \beta}{C_t} = \frac{\beta v_w}{R_t W_t - C_t} \quad (\text{I.6})$$

$$v_w \beta C_t = (1 - \beta) (R_t W_t - C_t) \quad (\text{I.7})$$

$$C_t = \frac{(1 - \beta) R_t W_t}{v_w \beta + (1 - \beta)} \quad (\text{I.8})$$

So then,

$$R_t W_t - C_t = R_t W_t - \frac{(1 - \beta) R_t W_t}{v_w \beta + (1 - \beta)} \quad (\text{I.9})$$

$$= R_t W_t \frac{v_w \beta}{v_w \beta + (1 - \beta)} \quad (\text{I.10})$$

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

$$\bar{v} + v_w (w_t + r_t) + \sum_{j=0}^{\infty} v_j \varepsilon_{t-j} = (1 - \beta) \log \frac{(1 - \beta) R_t W_t}{v_w \beta + (1 - \beta)} + \frac{\beta (1 - \alpha)}{2} (v_w b_0 + v_0)^2 + \quad (\text{I.11})$$

$$\beta \left( \bar{v} + v_w \log \left( R_t W_t \frac{v_w \beta}{v_w \beta + (1 - \beta)} \right) + \sum_{j=1}^{\infty} (v_w b_j + v_j) \varepsilon_{t+1-j} \right) \quad (\text{I.12})$$

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

$$v_w = (1 - \beta) + \beta v_w = 1 \quad (\text{I.13})$$

$$v_j = \beta (v_w b_{j+1} + v_{j+1}) \quad (\text{I.14})$$

So for  $v_0$ ,

$$v_0 = \beta b_1 + \beta v_1 \quad (\text{I.15})$$

$$= \sum_{j=1}^{\infty} \beta^j b_j \quad (\text{I.16})$$

We then have

$$\bar{v} = (1 - \beta) \log(1 - \beta) + \beta (\bar{v} + \log \beta) + \frac{\beta(1 - \alpha)}{2} b(\beta)^2 \quad (\text{I.17})$$

$$= \log(1 - \beta) + \frac{\beta}{1 - \beta} \log \beta + \frac{\beta(1 - \alpha)}{(1 - \beta)2} b(\beta)^2 \quad (\text{I.18})$$

Figure 1. Weighting function Z

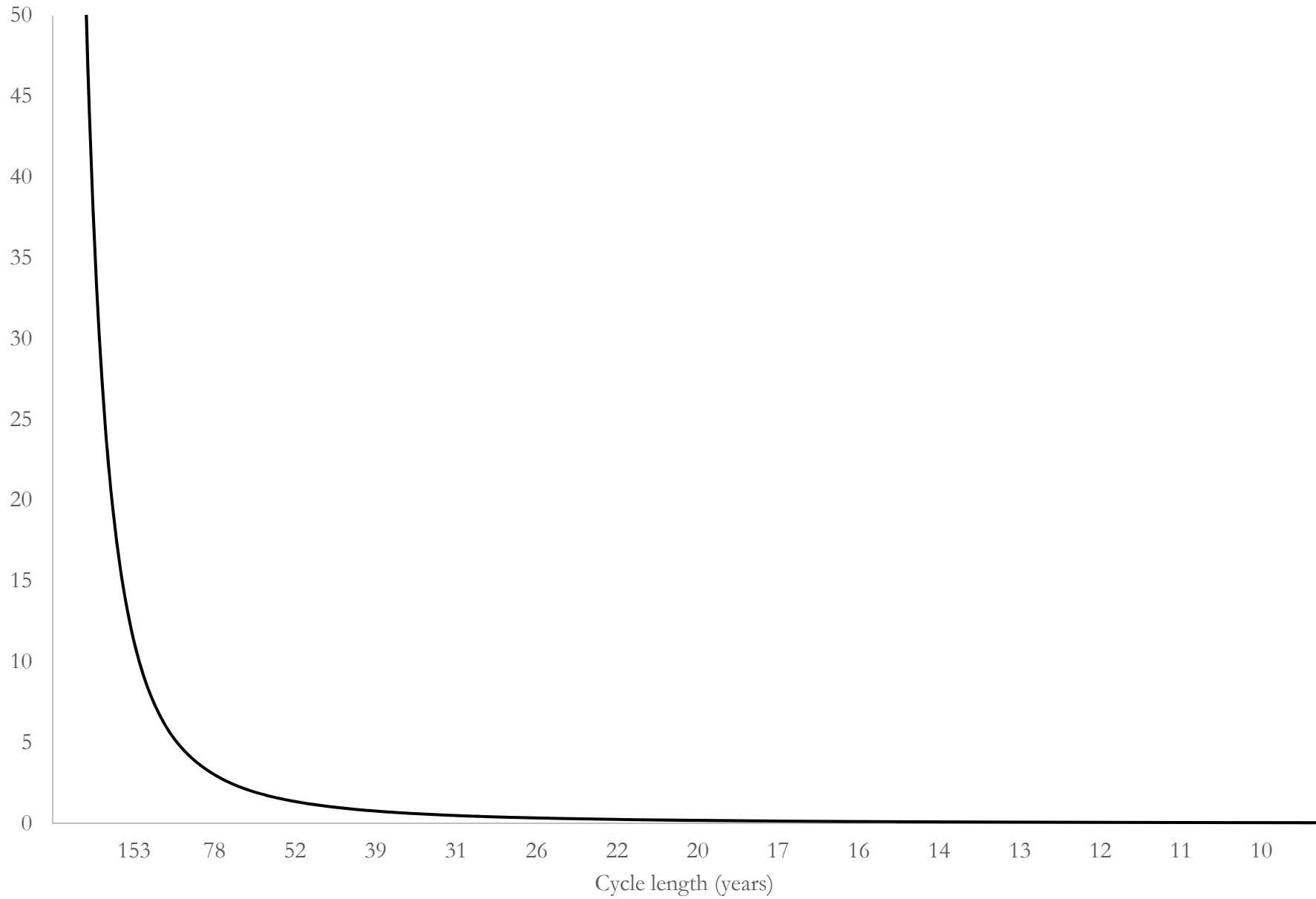


Figure 2. Transfer function under benchmark and worst case for white-noise benchmark

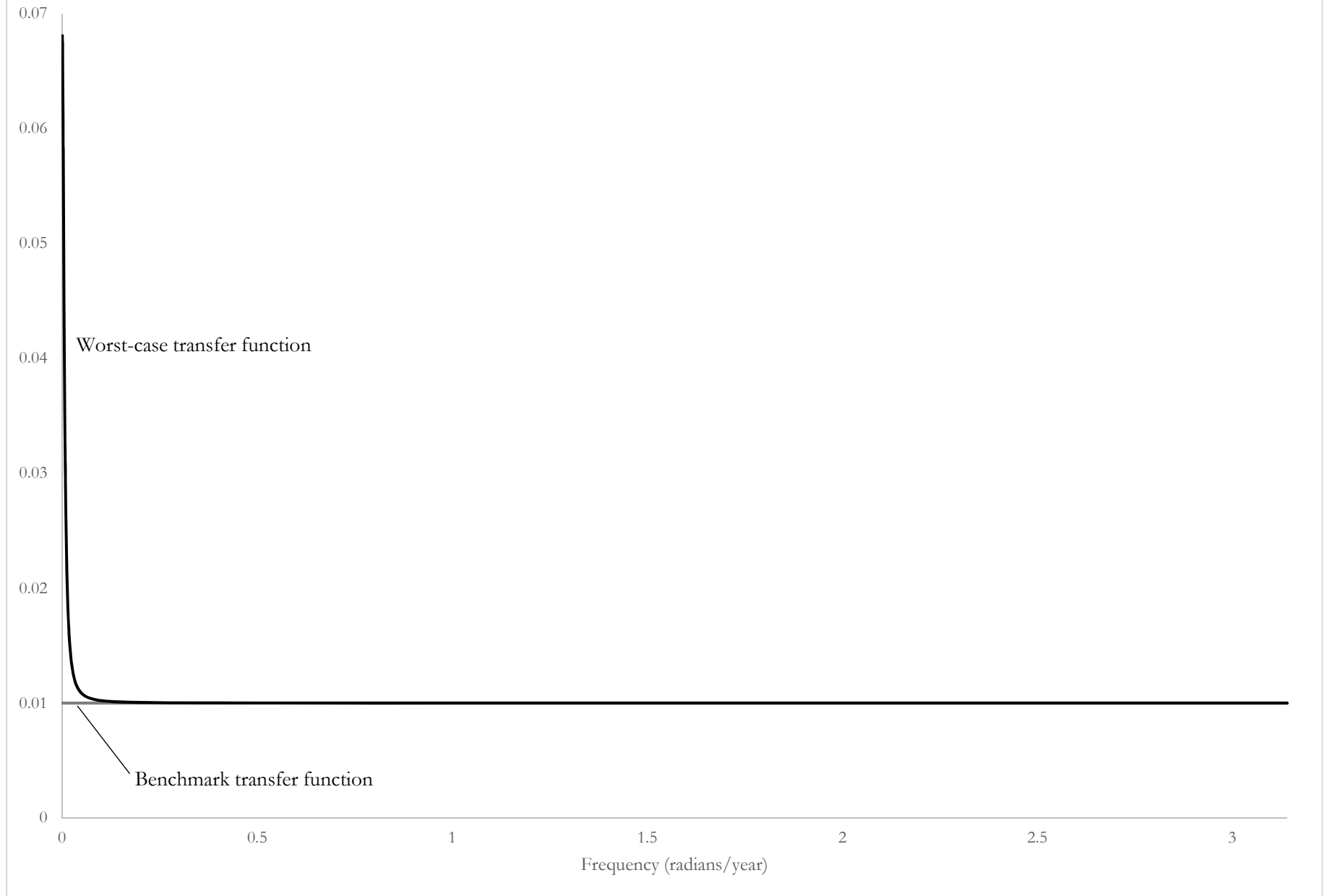
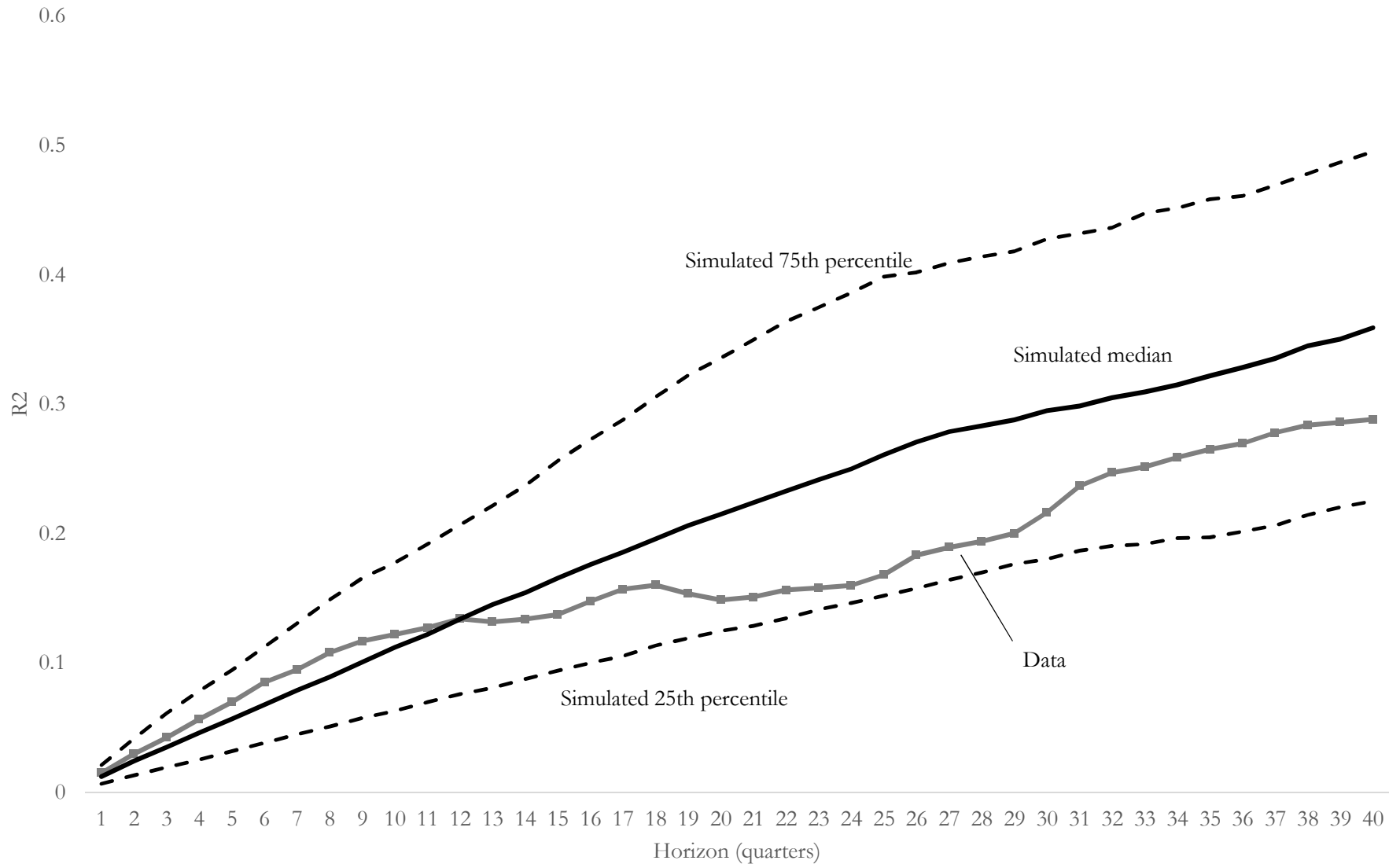


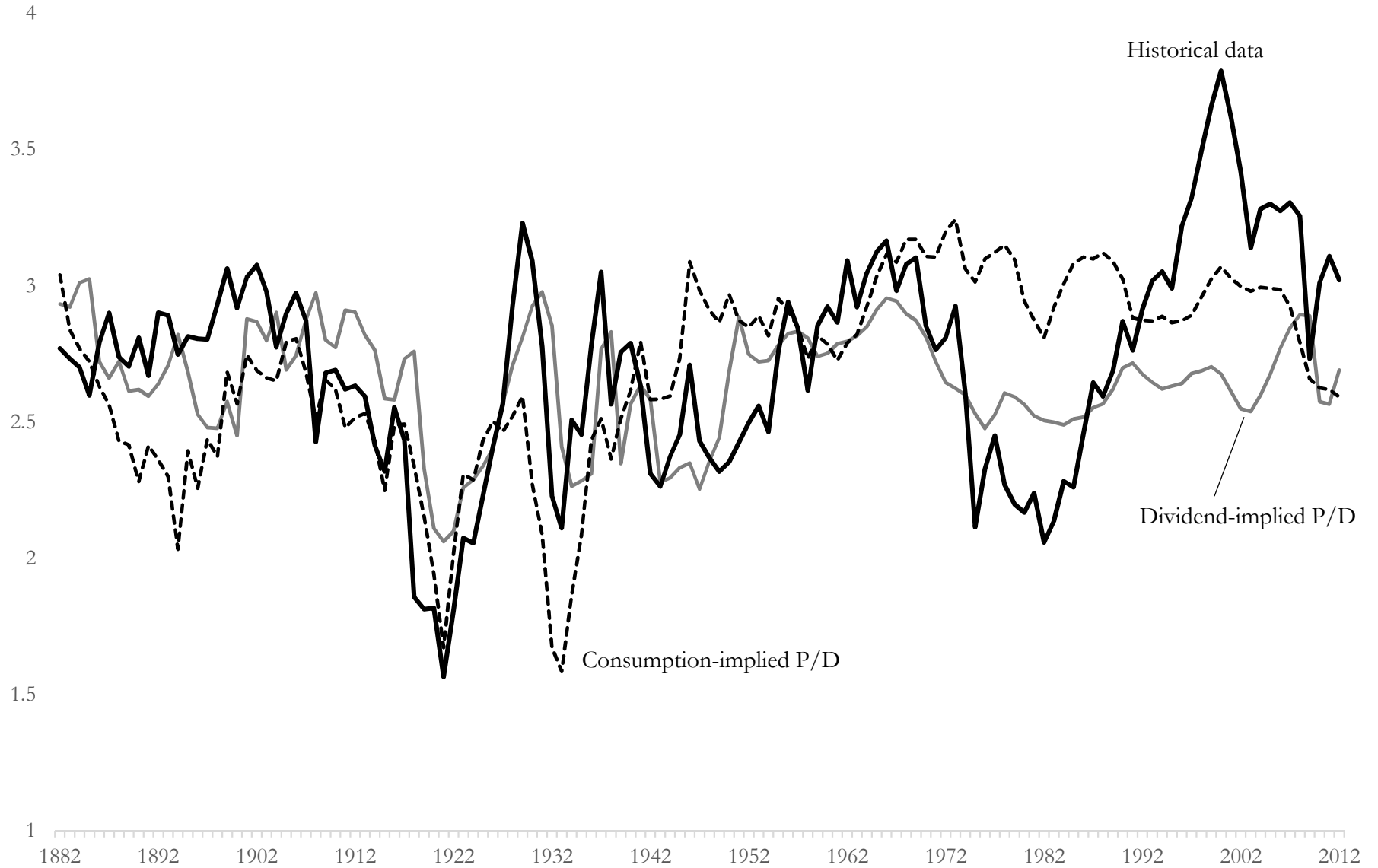


Figure 3. Empirical and model-implied R2's from return forecasting regressions



Notes: Black lines give results from simulated regressions on 60-year samples. The grey line plots R2s from regressions of aggregate equity returns on the price/dividend ratio in the post-war sample.

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



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

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

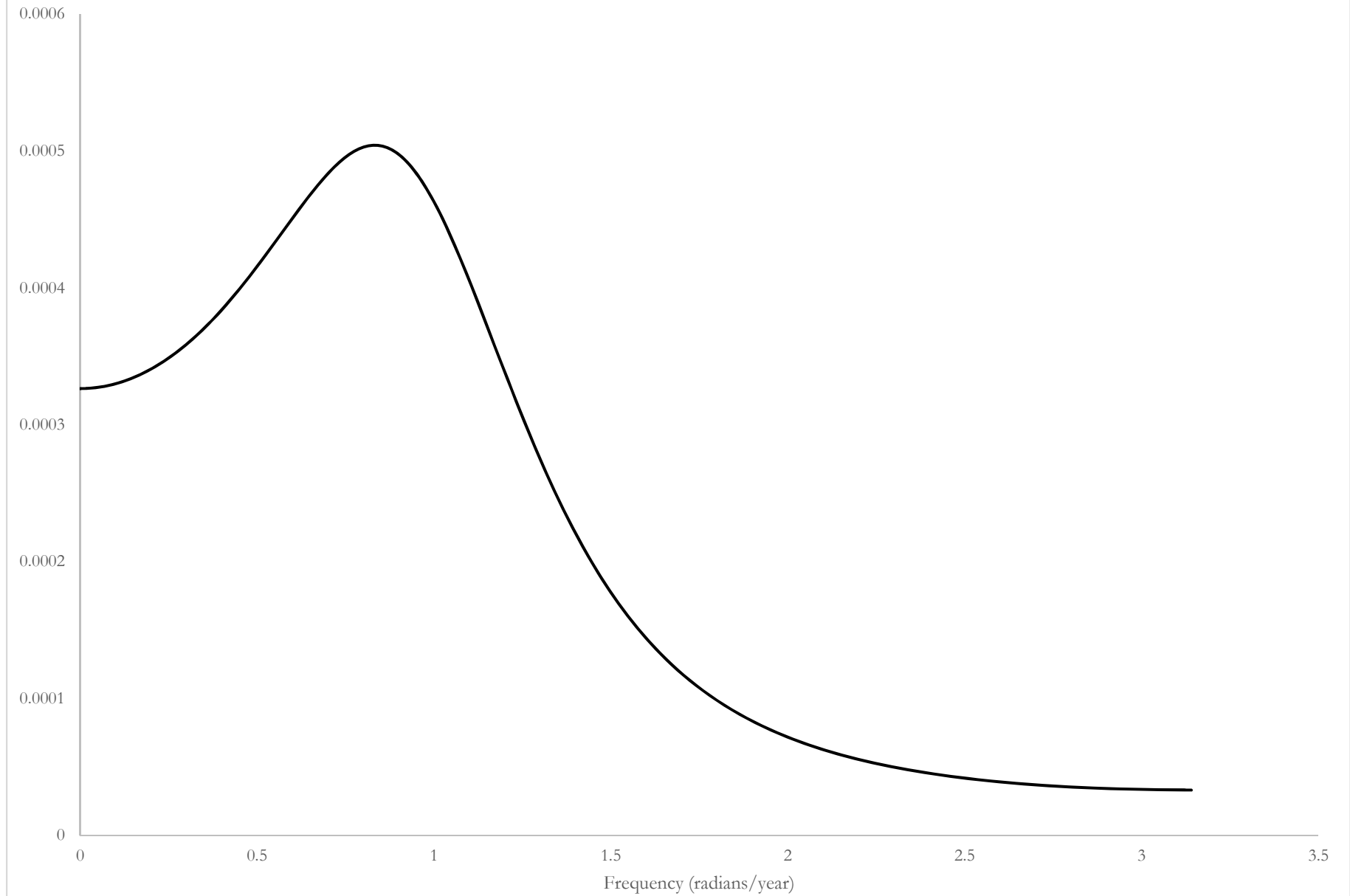
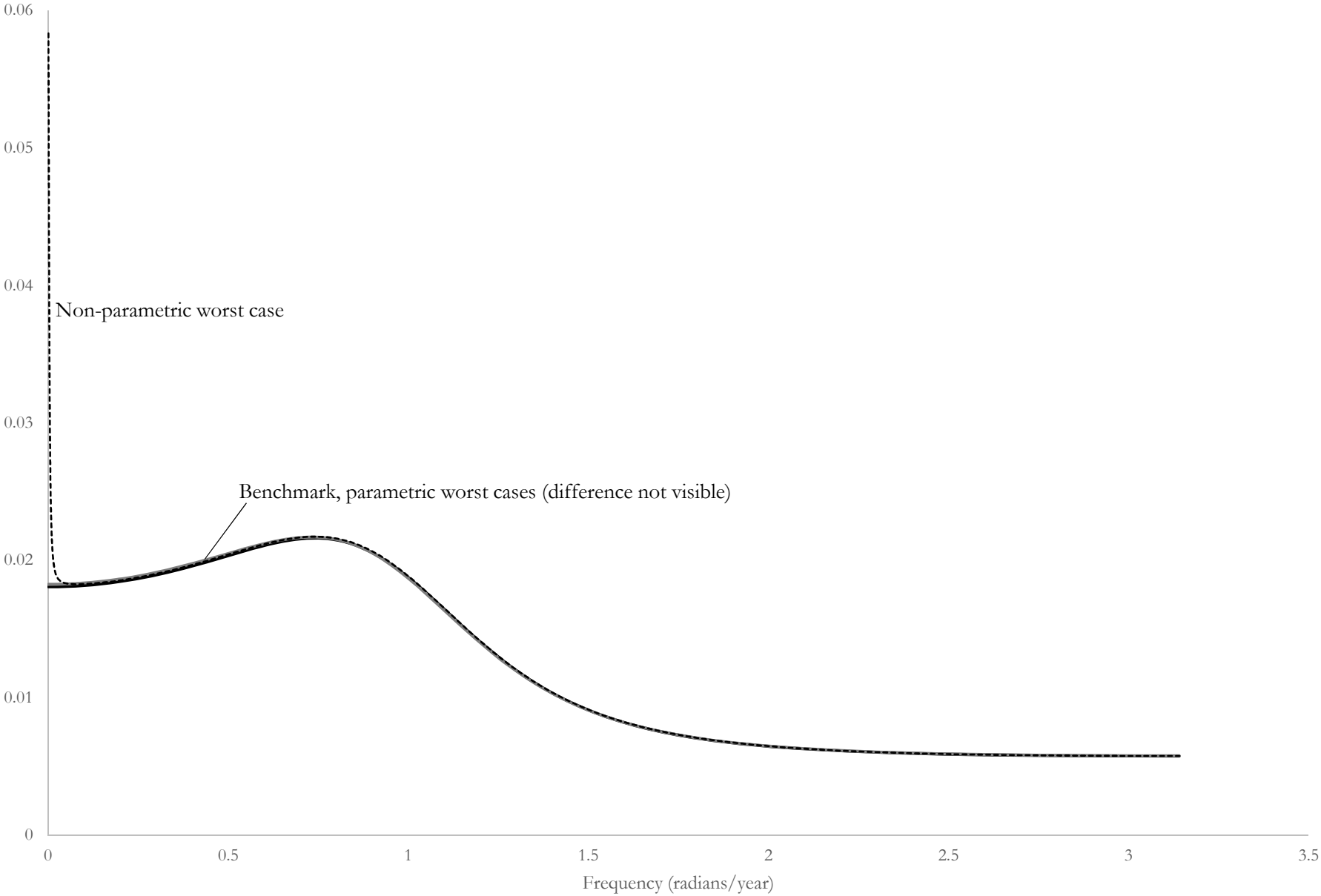


Figure 6. Benchmark and worst-case transfer function



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

<i>Fundamental parameters</i>			<i>Implied worst-case</i>	
$\bar{b}_0$	Cons. vol. point est.	0.0147	$b_0^w$	0.015
$\bar{b}(\beta)$	Long-run vol. point est.	0.0147	$b^w(\beta)$	0.039
$\mu$	Mean cons. growth	0.0045	$\theta$	0.989312
$\beta$	Time discount	0.997		
$\lambda$	Ambiguity aversion	52.23	<i>Standard Epstein–Zin / robust-control</i>	
$\alpha$	RRA (implied by $\lambda$ )	4.81	$b_0$	0.0147
$\gamma$	Leverage	4.626	$b(\beta)$	0.0147

*Asset pricing moments (annualized)*

	Model	Standard EZ	Data
std(M)	0.33	0.14	N/A
E[r-rf]	6.33	1.91	6.33
std(r)	19.42	13.55	19.42
E[rf]	1.89	2.43	0.86
std(rf)	0.33	0	0.97
AC1(PD)	0.96	N/A	0.81
std(P/D)	0.20	0	0.29
E[y <sub>10</sub> -rf]	-15bp	0	N/A
EIS estimate	0	N/A	0.14

Notes: Moments from the model with a white-noise benchmark process for consumption growth. The "standard Epstein–Zin" results are for where the agent is sure of the consumption process. 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. E[y<sub>10</sub>-rf] is the average spread between annualized yields between a one-quarter and a ten-year real riskless zero-coupon bond in basis points. The EIS estimate is based on a regression of consumption growth on interest rates. In the second column interest rates are constant, so the regression is degenerate.

**Table 2. Expectations of returns and expected returns***Regressions*

Dependent var.	Independent var.	Coefficient	Value in Greenwood and Shleifer (2014)
$E_t[r_{t+1}   b^w]$	$\log p_t/d_t$	1.68	1.08
$E_t[r_{t+1}   b^w]$	$r_t$	0.0034	0.0334
$r_{t+1}-r_{f,t+1}$	$\log p_t/d_t$	-10.57	-0.07
$r_{t+1}-r_{f,t+1}$	$E_t[r_{t+1}   b^w]$	-6.3	-1.61

*Correlations*

$\text{Corr}(E_t[r_{t+1}   b^w], \log p_t/d_t)$	-1	-0.3
$\text{Corr}(E_t[r_{t+1}   b^w], E_t[r_{t+1}   b])$	-1	N/A

Notes: Regressions and correlations involving expectations for returns calculated under investors' pricing model. The statistics are calculated for the case with the white-noise point estimate. The values from Greenwood and Shleifer (2014) are for their index of survey-based expectations.

**Table 3. Probability of rejecting the pricing model***Rejection probs. (5% critical value, H0=worst-case model)*

	50-year sample	100-year sample
Ljung–Box	4.7%	4.8%
ARMA(1,1)	5.6%	6.6%

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

Persistence	100-year sample	
0.9975	6.6%	(Our worst case)
0.995	7.7%	
0.99	12.9%	(Hansen and Sargent (2010))
0.98	27.1%	
0.94	86.6%	(Bansal and Yaron (2004))

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