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

Rhys Bidder and Ian Dew-Becker

April 20, 2016

Contents

A Full derivation of proposition 1 and corollary 2

A.1 Deriving the Kullback–Leibler divergence

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

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

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

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

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

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

Now note that

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

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

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

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

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

therefore write

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

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

Which implies that

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

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

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

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

$$\lim_{T\to\infty} T^{-1}E_{\Theta}\left[LL(T,\Theta) - LL\left(T,\bar{\Theta}\right)\right] = \frac{1}{2}\frac{1}{2\pi}\frac{\sigma^2}{\bar{\sigma}^2}\int_{-\pi}^{\pi}\frac{\left|B(\omega)-\bar{B}(\omega)\right|^2}{\left|\bar{B}(\omega)\right|^2}d\omega$$
$$-\frac{1}{2}\left(\log\left(\frac{\sigma^2}{\bar{\sigma}^2}\right) - \frac{\sigma^2-\bar{\sigma}^2}{\bar{\sigma}^2}\right) + \frac{1}{2}\frac{(\mu-\hat{\mu})^2}{f_{\bar{\Theta}}(0)} \tag{A.9}$$

A.2 Minimization

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

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

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

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

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

for a set of real coefficients c_j (Priestley (1981)). The coefficients c_j are simply the Fourier coefficients of the log of the spectrum; we only include coefficients for non-negative j because the spectrum is a real and even function. Furthermore, setting $\sigma = \exp(c_0)$, we have

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

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

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

We first calculate derivatives involved in the optimization

$$\begin{aligned}
\frac{d}{dc_j}\left[\sigma b(\beta)\right] &= \frac{d}{dc_j}\sum_{m=0}^{\infty}\beta^m \int \exp\left(\sum_{j=0}^{\infty} c_j e j\right) e^{-i\omega m} d\omega && \text{(A.14)}\\
&= \sum_{m=0}^{\infty}\beta^m \int \frac{d}{dc_j}\exp\left(\sum_{j=0}^{\infty} c_j e^{i\omega j}\right) e^{-i\omega m} d\omega && \text{(A.15)}\\
&= \sum_{m=0}^{\infty}\beta^m \int \exp\left(\sum_{j=0}^{\infty} c_j e^{i\omega j}\right) e^{-i\omega (m-j)} d\omega && \text{(A.16)}\\
&= \sigma \sum_{m=0}^{\infty}\beta^m b_{m-j} && \text{(A.17)}\\
&= \sigma \sum_{m=0}^{\infty}\beta^{m+j} b_m = \sigma b(\beta)\beta^j && \text{(A.18)}
\end{aligned}$$

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

Next, the derivative of the ratio of the spectra is

$$\begin{aligned}
\frac{d}{dc_j}\int \frac{f(\omega)}{\bar{f}(\omega)} d\omega &= \frac{d}{dc_j}\int \frac{\exp\left(2\sum_{j=0}^{\infty} c_j \cos(\omega j)\right)}{\bar{f}(\omega)} d\omega && \text{(A.19)}\\
&= 2\int \frac{f(\omega)}{\bar{f}(\omega)}\cos(\omega j)\, d\omega && \text{(A.20)}
\end{aligned}$$

And last,

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

So we have

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

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

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

For $j=0$,

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

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

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

where the third line follows by

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

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

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

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

We have

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

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

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

The FOC then becomes

$$\begin{aligned}
0 &= \frac{\beta}{1-\beta}\frac{1-\alpha}{2}\sigma_{w}^{2}b^{w}\left(\beta\right)^{2}\left(1-\beta^{2}\right)\left|Z\left(\kappa\right)\right|^{2}+\frac{\lambda}{2}\left(\frac{f^{w}\left(\omega\right)}{\bar{f}\left(\omega\right)}-1\right) && \text{(A.32)}\\
\frac{f^{w}\left(\omega\right)}{\bar{f}\left(\omega\right)}-1 &= \lambda^{-1}\frac{\beta}{1-\beta}\left(\alpha-1\right)\sigma_{w}^{2}b^{w}\left(\beta\right)^{2}\left(1-\beta^{2}\right)\left|Z\left(\kappa\right)\right|^{2} && \text{(A.33)}\\
f^{w}\left(\omega\right) &= \bar{f}\left(\kappa\right)+\lambda^{-1}\beta\left(1+\beta\right)\left(\alpha-1\right)\bar{f}\left(\kappa\right)\sigma_{w}^{2}b^{w}\left(\beta\right)^{2}\left|Z\left(\kappa\right)\right|^{2} && \text{(A.34)}
\end{aligned}$$

This is the main result in the text.

A.3 The white-noise benchmark

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

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

For the dynamics,

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

Denote the autocovariances under the worst-case model as γ_{j}^{w}. Then

$$\begin{aligned}
\gamma_{0}^{w}+2\sum_{j=1}^{\infty}\gamma_{j}^{w}\cos\left(\omega j\right) &= \bar{\sigma}^{2}\left(1+\varphi\left(1+2\sum_{j=1}^{\infty}\cos\left(\omega j\right)\beta^{j}\right)\right) && \text{(A.37)}\\
\text{where }\varphi &\equiv \lambda^{-1}\frac{\beta}{1-\beta}\left(\alpha-1\right)\sigma_{w}^{2}b^{w}\left(\beta\right)^{2} && \text{(A.38)}
\end{aligned}$$

Matching coefficients on each side yields

$$\begin{aligned}
\gamma_{0}^{w} &= \bar{\sigma}^{2}\left(1+\varphi\right) && \text{(A.39)}\\
\gamma_{j}^{w} &= \bar{\sigma}^{2}\varphi\beta^{j}\text{ for }\left|j\right|>0 && \text{(A.40)}
\end{aligned}$$

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

$$\Delta c_t = x_t + v_t \tag{A.41}$$
$$x_t = \beta x_{t-1} + \mu_t \tag{A.42}$$
$$\sigma_v^2 = \bar{\sigma}^2 \tag{A.43}$$
$$\sigma_\mu^2 = \bar{\sigma}^2 \varphi \left(1 - \beta^2\right) \tag{A.44}$$

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

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

$$\Delta c_t - \beta \Delta c_{t-1} = x_t - \beta x_{t-1} + v_t - \beta v_{t-1} \tag{A.45}$$
$$= \mu_t + v_t - \beta v_{t-1} \tag{A.46}$$

The second line is an MA(1), with

$$m_t \equiv \mu_t + v_t - \beta v_{t-1} \tag{A.47}$$
$$var\left(m_t\right) = \sigma_\mu^2 + \left(1 + \beta^2\right) \sigma_v^2 = \left(1 + \theta^2\right) \sigma_w^2 \tag{A.48}$$
$$cov\left(m_t, m_{t-1}\right) = -\beta \sigma_v^2 = -\theta \sigma_w^2 \tag{A.49}$$

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

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

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

B Testing the worst-case model

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

B.1 Test statistics

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

For the AP and LB tests, as discussed in the text, we assume that the agent takes an observed consumption history and creates a series of residuals,

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

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

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

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

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

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

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

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

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

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

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

Second, define

$$\begin{aligned} \varepsilon_t^* &\equiv \varepsilon_t - T^{-1}\sum_{t=1}^{T}\varepsilon_t && \text{(B.5)} \\ \hat{\omega}^2(\theta) &\equiv T^{-1}\sum_{t=1}^{T}\left(\varepsilon_t^*\right)^2 - \left[T^{-1}\frac{\left(\sum_{t=2}^{T}\varepsilon_t^*\sum_{i=0}^{t-2}\theta^i\varepsilon_{t-i-1}^*\right)^2}{\sum_{t=2}^{T}\left(\sum_{i=0}^{t-2}\theta^i\varepsilon_{t-i-1}^*\right)}\right] && \text{(B.6)} \end{aligned}$$

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

The likelihood ratio statistic is then

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

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

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

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

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

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

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

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

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

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

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

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

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

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

B.2 Extended results

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

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

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

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

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

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

C Interpretation of the distance measure as a Wald test

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

Brockwell and Davis (1988b) show that for an MA model of order m, the coefficients are asymptotically normal with a covariance matrix denoted Σ_m. As $m \to \infty$, Σ_m converges to a product,[2]

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

D Lifetime utility (assumption 3)

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

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

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

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

with subjective distribution

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

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

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

Inserting into the recursion for lifetime utility yields

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

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

Matching the coefficients on each side of the equality yields

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

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

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

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

E Multiplier preference interpretation

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

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

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

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

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

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

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

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

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

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

Which immediately yields a connection between λ and α,

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

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

F Asset prices and expected returns

F.1 Pricing a levered consumption claim

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

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

where δ is a linearization parameter slightly less than 1.

We guess that

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

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

The innovation to lifetime utility is

$$\begin{aligned} v_{t+1} - E_t\left[v_{t+1}|b^w\right] &= \sum_{k=0}^{\infty} \beta^k \Delta E_{t+1}\left[\Delta c_{t+k+1}|\Theta^w\right] && \text{(F.3)} \\ &= b^w(\beta)\,\varepsilon_{t+1}^{\Theta^w} && \text{(F.4)} \end{aligned}$$

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

The pricing kernel can therefore be written as

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

The pricing equation for the levered consumption claim is

$$\begin{aligned} 0 &= \log E_t\left[\beta \exp\left(\begin{array}{c} \delta_0 + (\delta-1)\,\bar{h} + (\delta h_0 + \gamma - 1)\,\Delta c_{t+1} + \sum_{j=0}^{\infty} (\delta h_{j+1} - h_j)\,\Delta c_{t-j} \\ + (1-\alpha)\, b^w(\beta)\, \varepsilon_{t+1}^{\Theta^w} - \frac{(1-\alpha)^2}{2} b^w(\beta)^2 \sigma_w^2 \end{array}\right) |\Theta^w\right] && \text{(F.6)} \\ &= (\delta h_0 + \gamma - 1)\left((1 - a^w(1))\,\mu^w + a^w(L)\,\Delta c_t\right) + \sum_{j=0}^{\infty} (\delta h_{j+1} - h_j)\,\Delta c_{t-j} \\ &\quad + \delta_0 + \left(\frac{1}{2}(\delta h_0 + \gamma - 1)^2 + (\delta h_0 + \gamma - 1)(1-\alpha)\, b^w(\beta)\right)\sigma_w^2 + (\delta - 1)\,\bar{h} + \log \beta && \text{(F.7)} \end{aligned}$$

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

$$\begin{aligned} (\delta - 1)\,\bar{h} + \log\beta + \delta_0 &= -\left(\frac{1}{2}(\delta h_0 + \gamma - 1)^2 + (\delta h_0 + \gamma - 1)(1-\alpha)\, b^w(\beta)\right)\sigma_w^2 \\ &\quad - (\delta h_0 + \gamma - 1)(1 - a^w(1))\,\mu^w && \text{(F.8)} \\ (\delta h_{j+1} - h_j) &= -(\delta h_0 + \gamma - 1)\, a_j^w && \text{(F.9)} \end{aligned}$$

And thus

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

and

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

Note then that

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

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

F.2 The risk-free rate

For the risk-free rate, we have

$$\begin{aligned} r_{f,t+1} &= -\log E_t\left[\beta \exp\left(-\Delta c_{t+1} + (1-\alpha)\, b^w(\beta)\, \varepsilon_{t+1}^{\Theta^w} - \frac{(1-\alpha)^2}{2} b^w(\beta)^2 \sigma_w^2\right) |\Theta^w\right] && \text{(F.14)} \\ &= -\log\beta + (1 - a^w(1))\, \mu^w + a^w(L)\, \Delta c_t - \frac{1}{2}\sigma_w^2 + (1-\alpha)\, b^w(\beta)\, \sigma_w^2 && \text{(F.15)} \\ &= -\log\beta + \mu^w + a^w(L)\,(\Delta c_t - \mu^w) - \frac{1}{2}\sigma_w^2 + (1-\alpha)\, b^w(\beta)\, \sigma_w^2 && \text{(F.16)} \end{aligned}$$

F.3 Expected excess returns

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

$$\begin{aligned} E_t\left[r_{t+1}|\bar{\Theta}\right] &= E_t\left[\delta_0 + (\delta-1)\,\bar{h} + (\delta h_0 + \gamma)\,\Delta c_{t+1} + \sum_{j=0}^{\infty}(\delta h_{j+1} - h_j)\,\Delta c_{t-j}|\bar{\Theta}\right] && \text{(F.17)} \\ &= \delta_0 + (\delta-1)\,\bar{h} - (\delta h_0 + \gamma - 1)\, a^w(L)\,\Delta c_t + E_t\left[(\delta h_0 + \gamma)\,\Delta c_{t+1}|\bar{\Theta}\right] && \text{(F.18)} \end{aligned}$$

$$\begin{aligned} &= \delta_0 + (\delta-1)\,\bar{h} + (-(\delta h_0 + \gamma - 1)\, a^w(L) + (\delta h_0 + \gamma)\, a(L))\,\Delta c_t \\ &\quad + (\delta h_0 + \gamma)(1 - a(1))\,\mu && \text{(F.19)} \end{aligned}$$

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

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

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

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

where

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

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

Substituting in

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

yields the result from the text.

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

F.4 The behavior of interest rates

The mean of the risk-free rate is

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

And its standard deviation is

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

When consumption growth is white noise, this is

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

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

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

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

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

The pricing condition for a bond is

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

$$\begin{aligned}
\phi^{(j)}(L)\Delta c_t + n_j &= \log E_t\left[\exp\left(\begin{array}{c}\log\beta - \Delta c_{t+1} + (1-\alpha) b^w(\beta)\varepsilon_{t+1}^{\Theta^w} \\ -\frac{(1-\alpha)^2}{2} b^w(\beta)^2\sigma_w^2 + \phi^{(j-1)}(L)(\Delta c_{t+1} - \mu^w) + n_{j-1}\end{array}\right)|\Theta^w\right] && \text{(F.33)} \\
&= \log\beta + \left(\phi_0^{(j-1)} - 1\right)(\mu^w + a^w(L)(\Delta c_t - \mu^w)) - \phi_0^{(j-1)}\mu^w + \sum_{k=0}^{\infty}\phi_{k+1}^{(j-1)}(\Delta c_{t-k} - \mu^w) \\
&\quad -\frac{(1-\alpha)^2}{2} b^w(\beta)^2\sigma_w^2 + n_{j-1} + \frac{1}{2}\left((1-\alpha) b^w(\beta) - 1 + \phi_0^{(j-1)}\right)^2\sigma_w^2 && \text{(F.34)}
\end{aligned}$$

Matching coefficients yields,

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

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

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

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

F.5 Results used in table 1

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

$$\begin{aligned}
\Delta c_t &= \beta\Delta c_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1} && \text{(F.38)} \\
a^w(L) &= (\beta - \theta)\sum_{j=0}^{\infty}\theta^j L^j && \text{(F.39)}
\end{aligned}$$

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

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

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

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

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

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

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

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

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

And thus

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

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

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

F.6 Returns in the absence of model uncertainty

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

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

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

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

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

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

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

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

The standard deviation of the log pricing kernel is

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

G Dividends cointegrated with consumption

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

We assume the following model holds

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

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

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

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

In terms of growth rates we have

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

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

We then recapitulate the analysis from above. Specifically, we add a superscript C to the coefficients in the price/dividend function to yield the guess

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

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

The pricing equation for the dividend claim is

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

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

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

$$\begin{aligned} \delta h_{c,j+1}^C - h_{c,j}^C &= - \left(\delta h_{c,0}^C + \gamma g_{c,0} - 1 \right) a_j^w - \gamma g_{c,j+1} \qquad \text{(G.9)} \\ \delta h_{\zeta,j+1}^C - h_{\zeta,j}^C &= -\tilde{g}_{\zeta,j+1} \qquad \text{(G.10)} \end{aligned}$$

And thus

$$\begin{aligned} h_{c,j}^C &= \delta h_{c,j+1}^C + \left(\delta h_{c,0}^C + \gamma g_{c,0} - 1 \right) a_j^w + \gamma g_{c,j+1} \qquad \text{(G.11)} \\ h_{c,0}^C &= \sum_{j=0}^{\infty} \left(\delta h_{c,0}^C + \gamma g_{c,0} - 1 \right) a_j^w \delta^j + \delta^{-1} \sum_{j=1}^{\infty} \gamma g_{c,j} \delta^j \qquad \text{(G.12)} \\ \delta h_{c,0}^C + \gamma g_{c,0} - 1 &= \left(\delta h_0^C + \gamma g_{c,0} - 1 \right) \delta a^w(\delta) + \gamma g_c(\delta) - 1 \qquad \text{(G.13)} \\ \delta h_{c,0}^C + \gamma g_{c,0} - 1 &= \frac{\gamma g_c(\delta) - 1}{1 - \delta a^w(\delta)} \qquad \text{(G.14)} \end{aligned}$$

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

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

For the coefficients on ζ, we have

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

Note then that

$$\begin{aligned} var_w(r_{m,t+1}) &= \left(\delta h^C_{c,0} + \gamma g_{c,0}\right)^2 \sigma_w^2 + \tilde{g}_\zeta(\delta)^2 && \text{(G.17)} \\ cov_w(r_{m,t+1}, m_{t+1}) &= \left(\delta h^C_{c,0} + \gamma g_{c,0}\right)\left(-1 + (1-\alpha) b^w(\beta)\right)\sigma_w^2 && \text{(G.18)} \end{aligned}$$

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

G.1 Calibration

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

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

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

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

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

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

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

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

$$\begin{aligned} g^2_{\zeta,0} &= \frac{1}{2}\left(var(\Delta d) - \gamma^2\left(g^2_{c,0} + g^2_{c,1}\right) var(\Delta c)\right) && \text{(G.22)} \\ g_{\zeta,0} &= 0.019 && \text{(G.23)} \end{aligned}$$

That is, the final model of dividends is

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

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

G.2 Expected excess returns

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

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

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

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

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

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

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

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

where, from above,

$$
\begin{aligned}
var_w\left(r_{m,t+1}\right) &= \left(\frac{\gamma g_c\left(\delta\right)-\delta a^w\left(\delta\right)}{1-\delta a^w\left(\delta\right)}\right)^2 \sigma_w^2+\tilde{g}_\zeta\left(\delta\right)^2 && \text{(G.32)} \\
cov_w\left(r_{m,t+1}, m_{t+1}\right) &= \frac{\gamma g_c\left(\delta\right)-\delta a^w\left(\delta\right)}{1-\delta a^w\left(\delta\right)}\left(-1+\left(1-\alpha\right) b^w\left(\beta\right)\right) \sigma_w^2 && \text{(G.33)}
\end{aligned}
$$

G.3 Price/dividend ratio

$$
\begin{aligned}
\delta h_{c,j+1}^C-h_{c,j}^C &= -\left(\delta h_{c,0}^C+\gamma g_{c,0}-1\right) a_j^w-\gamma g_{c,j+1} && \text{(G.34)} \\
\delta h_{\zeta,j+1}^C-h_{\zeta,j}^C &= -\tilde{g}_{\zeta,j+1} && \text{(G.35)}
\end{aligned}
$$

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

$$
\begin{aligned}
h_{c,0} &= \frac{\gamma g_c\left(\delta\right)-1}{1-\delta a^w\left(\delta\right)}\left(\beta-\theta\right)\frac{1}{1-\delta\theta}+\gamma g_{c,1} && \text{(G.37)} \\
h_{c,j} &= \frac{\gamma g_c\left(\delta\right)-1}{1-\delta a^w\left(\delta\right)}\left(\beta-\theta\right)\frac{\theta^j}{1-\delta\theta} && \text{(G.38)}
\end{aligned}
$$

So the standard deviation of the pricing kernel is now

$$
\begin{aligned}
pd_t^C &= \sum_{j=0}^{\infty} h_{c,j}^C \Delta c_{t-j}+g_{\zeta,1}\zeta_t && \text{(G.39)} \\
var\left(pd_t^C\right) &= \left(\frac{\gamma g_c\left(\delta\right)-1}{1-\delta a^w\left(\delta\right)}\frac{\left(\beta-\theta\right)}{1-\delta\theta}\right)^2 \frac{\bar{\sigma}^2}{1-\theta^2}+g_{\zeta,1}^2 && \text{(G.40)}
\end{aligned}
$$

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

G.4 Results

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

The mean and standard deviation of returns are both slightly reduced – the mean is lower by 5 basis points and the standard deviation by 14 basis points. The small reduction is due to the fact that $g_c\left(\delta\right)=0.993$. The difference between the returns under the two models of dividends depends purely on that term being different from 1. The fact that it is not (which is a consequence of cointegration) is why the returns are essentially unchanged. The autocorrelation and standard deviation of the price/dividend ratio are also numerically nearly identical to what is obtained in table 1. Finally, the bottom two rows of table A1 confirm that the model is calibrated here so that the standard deviation of dividend growth and the correlation between dividend growth and consumption growth is identical to the data (the data moments are drawn from Bansal and Yaron (2004), as is the case with our other empirical targets).