Online Appendix for "When Are Stocks Less Volatile in the Long Run?"

This document provides supplementary material to the paper "When Are Stocks Less Volatile in the Long Run?" It provides additional details on the estimation method (Section 1); additional details on the interval method that we use in Section 4 of the paper (Section 2); and a brief description of the lung-run risks model of Bansal and Yaron (2004) (Section 3).

1 Estimation Method

1.1 The Bayesian Method

This section introduces the Bayesian MCMC approach for estimating the CV-DC model. Note that the estimation of the CV-DC model with the NEP constraint is a special case of this estimation procedure. Specifically, we need to drop random draws that do not satisfy the NEP condition before we form posterior beliefs. To facilitate the estimation, we rewrite the CV-DC model in a standard state-space formation. The measurement equation is

$$y_{t+1} = D + By_t + v_{t+1},$$

where

$$y_{t+1} = \begin{bmatrix} r_{t+1} - \beta_{t+1} x_t \\ x_{t+1} \end{bmatrix}, \; D = \begin{bmatrix} \alpha \\ 0 \end{bmatrix}, \; B = \begin{bmatrix} 0 & \beta_0 \\ 0 & \beta_x \end{bmatrix}, \; v_{t+1} = \begin{bmatrix} \varepsilon_{t+1} \\ \epsilon_{t+1} \end{bmatrix},$$

and $Var(v_{t+1}) = \Omega$. The state equation is

$$\beta_{t+1} = b\beta_t + \eta_{t+1},$$

with $Var(\eta_{t+1}) = \sigma_\eta^2$. The parameters we need to estimate are: $\Theta = \{\alpha, \beta_0, \beta_x, b, \sigma_\eta^2, \Omega\}$. We also need to filter out the latent state variable β_t.

There are 6 steps to estimate the CV-DC model. These steps generate the following parameters sequentially: the latent state variable β_t, the coefficient matrix B, the intercept vector D, the covariance matrix Ω, the autoregressive coefficient b, and the variance σ_η^2. We describe how to estimate these parameters using an MCMC approach. To facilitate the presentation of the Bayesian estimation procedure, let $Y_t = \{y_i\}_{i=1}^t$ denote the observable at time t; Θ_- represents the set of model parameters except the parameters that we estimate in a specific step.

Step 1: Generation of the latent state variable β_t. A multimove-Gibbs sampling method is employed to draw the unobserved state variable (Carter and Kohn, 1994). Kim and Nelson (1999) partition the joint distribution of $\beta = \{\beta_t\}_{t=1}^T$ given the data set Y_T and the parameter set Θ:

$$p(\beta|Y_T, \Theta) = p(\beta_T|Y_T, \Theta) \prod_{t=1}^{T-1} p(\beta_t|Y_T, \Theta).$$

It is known that $p(\beta_T|Y_T,\Theta) \sim N(\beta_{T|T}, P_{T|T})$, where $\beta_{T|T}$ is the conditional expectation of β_T and $P_{T|T}$ is the conditional covariance matrix of β_T. In the estimation, $\beta_{T|T}$ and $P_{T|T}$ are obtained from the last step of the Kalman filter. The Kalman filter algorithm is described as (e.g., Zhu, 2015):

$$\beta_{t-1}|\{\beta_t, Y_{t-1}\} \sim N(\beta_{t|t,\beta_{t+1}}, P_{t|t,\beta_{t+1}}),$$

with

$$\beta_{t|t,\beta_{t+1}} = \beta_{t|t} + bP_{t|t}(b^2P_{t|t} + \sigma_\eta^2)^{-1}(\beta_{t+1} - b\beta_{t|t})$$

and

$$P_{t|t,\beta_{t+1}} = P_{t|t} - b^2P_{t|t}^2(b^2P_{t|t} + \sigma_\eta^2)^{-1}.$$

Then, we apply the forward filtering and backward sampling (FFBS) approach to draw β_t.

Step 2: Generation of the intercept vector D. A standard Gibbs sampling step can be used to draw parameters D. The conjugate priors and posteriors of D follow a normal distribution. Suppose that the posterior of D conditional on the observed data Y_T, the filtered state variable β, and the other parameters Θ_- is:

$$\begin{aligned} P(D|\Theta_-, Y_T, \beta) &\propto P(Y_T|\Theta,\beta)P(\beta|\Theta)P(D) \\ &\propto P(Y_T|D,B,\beta)P(\beta|b,\sigma_\eta^2)P(D) \\ &\propto P(Y_T|D,B,\beta)P(D), \end{aligned}$$

where $P(Y_T|D,B,\beta)$ is the likelihood function of the measurement equation, which can be easily computed conditional on the observations and the filtered state variables, and $P(\beta|b,\sigma_\eta^2)$ is the likelihood function of the state equation.

Step 3: Generation of the autogressive coefficient b. Note that the posterior of b conditional on Y_T, β, and Θ_- is:

$$\begin{aligned} P(b|\Theta_-,Y_T,\beta) &\propto P(Y_T|\Theta,\beta)P(\beta|\Theta)P(b) \\ &\propto P(Y_T|D,B,\beta)P(\beta|b,\sigma_\eta^2)P(b) \\ &\propto P(\beta|b,\sigma_\eta^2)P(b). \end{aligned}$$

As is standard in the literature, the prior and the posterior of b can be drawn using a conjugate normal distribution.

Step 4: Generation of the coefficient matrix B. The posterior distribution of B conditional on Y_T, β, and Θ_- is:

$$\begin{aligned} P(B|\Theta_-,Y_T,\beta) &\propto P(Y_T|\Theta,\beta)P(X|\Theta)P(B) \\ &\propto P(Y_T|D,B,\beta)P(\beta|b,\sigma_\eta^2)P(B) \\ &\propto P(Y_T|D,B,\beta)P(B). \end{aligned}$$

Naturally, we specify the prior and the posterior of B as a conjugate normal distribution in the MCMC estimation.

Step 5: Generation of the variance parameter σ_η^2. The posterior probability of σ_η^2 conditional on Y_T, β, and Θ_- is:

$$
\begin{aligned}
P(\sigma_\eta^2|\Theta_-, Y_T, \beta) &\propto P(Y_T|\Theta, \beta)P(\beta|\Theta)P(\sigma_\eta^2) \\
&\propto P(Y_T|\Theta, \beta)P(\beta|b, \sigma_\eta^2)P(\sigma_\eta^2) \\
&\propto P(\beta|b, \sigma_\eta^2)P(\sigma_\eta^2).
\end{aligned}
$$

According to this formula, if the prior of σ_η^2 is specified as an inverted Gamma distribution, $\sigma_\eta^2 \sim IG(v_0/2, \delta_0/2)$, the posterior distribution of σ_η^2 is still an inverted Gamma distribution, $\sigma_\eta^2 \sim IG(v_1/2, \delta_1/2)$, with $v_1 = v_0 + T$ and $\delta_1 = \delta_0 + \sum_{t=1}^{T} \eta_t^2$.

Step 6: Generation of the covariance matrix Ω. The posterior of Ω conditional on Y_T, β, and Θ_- is:

$$
\begin{aligned}
P(\Omega|\Theta_-, Y_T, \beta) &\propto P(Y_T|\Theta, \beta)P(\beta|\Theta)P(\Omega) \\
&\propto P(Y_T|D, B)P(\beta|b, \sigma_\eta^2)P(\Omega) \\
&\propto P(Y_T|D, B)P(\Omega).
\end{aligned}
$$

The covariance matrix can be drawn from the inverted Wishart (IW) distribution. With an informative prior, the posterior distribution of covariance matrix follows:

$$
\Omega|Y, \Theta_- = IW\left(T, \sum\nolimits_{t=1}^{T} v_t' v_t\right).
$$

As demonstrated by several studies, Bayesian estimation depends on prior distributions. It is particularly so for predictive models (see, for example, Pástor and Stambaugh, 2009). To make our results comparable to Johannes et al. (2014), we set priors in our Bayesian estimation as described in their paper. Specifically, we train the priors from 1952 to 1955 by regressing excess returns on a constant and the predictor. In so doing, we assume noninformative priors. In the estimation, we drop the first 100,000 draws and use the subsequent 200,000 draws to calculate the posteriors.

1.2 Metropolis-Hasting Algorithm

Let Λ denote the parameters that determine the expected return, and Σ_Λ denote the other parameters of the predictive model. Notably, Σ_Λ does not include the time series of $\{\mu_t\}$ when the model is the predictive system. We have $\Lambda = \{E_r, \beta\}$ for the restricted predictive system and $\Lambda = \{\alpha, \beta_0\}$ for the restricted CV-DC model. In the present context, the Metropolis-Hastings algorithm proceeds as follows:

Step 1: Initialize Λ, Σ_Λ, and $\{\mu_t\}$. Initial values of $\{\mu_t\}_{t=1}^T$ are given by the estimates obtained with the unrestricted predictive system, but we set $\mu_t = 0$ if the posterior mean of μ_t is negative. Set $i = 1$.

Step 2: Conditional on the data, Λ^i, and $\{\mu_t\}^i$, draw Σ_Λ^i following the method presented in Pástor and Stambaugh (2009) (see Section B5.1 in their Online Appendix for details) for the estimation of the predictive system or the Gibbs sampler presented in Appendix 1.1 for the estimation of the CV-DC model. Note that this step is a combination step since it involves multiple steps to draw the elements of Σ_Λ.

Step 3: Given the data and the current value Λ^i, use a symmetric transition density $q(\Lambda^i, \Lambda^c)$ to generate a candidate Λ^c for the next value in the MCMC sequence.[1] In our context, the independent Metropolis-Hastings algorithm is applied to draw a candidate from a normal proposal distribution. The mean of E_r (or α for the CV-DC

[1] If $i = 1$, we need to keep drawing Λ^c until Λ^c does not violate the NEP condition. This involves repeatedly conducting Steps 3 and 4. We then set $\Lambda^1 = \Lambda^c$.

model) is equal to the mean of the posterior distribution implied by the unrestricted model, and the covariance matrix of the normal distribution is equal to that of the posterior distribution implied by the unrestricted model tuned by a scalar ψ.

Step 4: Given the data, Σ_Λ^i, and Λ^c, employ the forward filtering and backward sampling method to draw the time series of $\{\mu_t\}^c$ for the predictive system.

Step 5: Evaluate the NEP condition at each point in time. If any NEP condition is violated, set the acceptance probability $\alpha(\Lambda^i, \Lambda^c) = 0$ and go to Step 7. For the predictive system, the NEP condition is binding if $\{\mu_t\}^i < 0$ for $t = 1, ..., T$. For the CV-DC model, the NEP condition is binding if $\{\alpha_0 + \beta_0 x_t + \beta_{t+1} x_t\} < 0$ for $t = 1, ..., T$.

Step 6: Calculate $\alpha(\Lambda^i, \Lambda^c) = \min(g(\Lambda^c)/g(\Lambda^i), 1)$, where $g(\Lambda)$ is the kernel of the density function of Λ conditional on the data, $f(\Lambda|Y)$.

Step 7: Generate an independent uniform random variable u from the interval $[0, 1]$.

Step 8: Set $\Lambda^{i+1} = \begin{cases} \Lambda^c \text{ if } u \leq \alpha(\Lambda^i, \Lambda^c) \\ \Lambda^i \text{ if } u > \alpha(\Lambda^i, \Lambda^c). \end{cases}$
Similarly, set $\{\mu_t\}^{i+1} = \{\mu_t\}^c$ or $\{\mu_t\}^{i+1} = \{\mu_t\}^i$.

Step 9: Set $i = i + 1$ and go to Step 2.

Note that Step 4 is a step only for estimating the restricted predictive system, so the estimation of the restricted CV-DC model consists of Steps (1)–(3) and Steps (5)–(9).

This iteration scheme generates a sequence of draws with the property that for large i, Λ^{i+1} is effectively a sample draw from $f(\Lambda|Y)$. As a result, the last elements in the sequence can be regarded as draws from $f(\Lambda|Y)$. Importantly, this chain of draws is consistent with the NEP condition. In the empirical analysis, we simulate

$k = 300,000$ draws with an initial burn-in period of $m = 100,000$ observations for both the predictive system and the CV-DC model under the NEP condition. These numbers of k and m are set using the convergence diagnostics of Geweke (1992) and Raftery and Lewis (1992).

The implementation of the Metropolis-Hastings algorithm involves choosing an arbitrary proposal density to produce candidates for inclusion in the MCMC sequence. In our empirical analysis, we follow prior research and use a normal proposal density, with a covariance matrix equal to a tuning scalar multiplied by the estimated covariance matrix from the estimation of the unrestricted predictive model. The tuning scalar, denoted by ψ, is used to control the acceptance rate (i.e., the rate at which draws are included in the MCMC sequence). According to Roberts et al. (1997), the optimal acceptance rate for a normal transition density should be between 0.23 and 0.45. In our empirical application, we set $\psi = 0.70$ for the restricted predictive system and achieve an acceptance rate of 0.30. We set $\psi = 0.65$ for the restricted CV-DC model and obtain an acceptance rate of 0.33.

1.3 Prior Distributions

1.3.1 Predictive system

In the spirit of Pástor and Stambaugh (2009, 2012), we impose informative prior distributions on three key parameters of the predictive system that affect multiperiod predictive variance – β, R^2, and ρ_{uw}, where ρ_{uw} is the correlation between unexpected returns, u_t, and innovations in expected returns, w_t, and R^2 is the fraction of the variation in r_{t+1} that can be explained by expected returns, μ_t. Informative priors are economically motivated by a large amount of empirical evidence and economic intuition. Indeed, these prior distributions are an important feature of the predictive system. Furthermore, as stressed by Pástor and Stambaugh, these priors are necessary for model identification. Our priors roughly follow the benchmark priors of Pástor and Stambaugh (2012), but we slightly adjust the prior distributions to reflect the different data frequency: we shift the priors for R^2 and ρ_{uw} to the left and those for β to the right. The shift of prior distributions attempts to capture the fact that stock returns are generally less predictable in monthly frequency than in annual data and the correlation between expected and unexpected returns is likely to be less negative at lower frequency. Figure A1 plots the prior distributions of the three key parameter of the predictive system.

Following the method described in Pástor and Stambaugh (2009), we use a mildly informative normal distribution prior on E_r, $E_r \sim N(0.058, 0.08^2)$, centered at the sample mean return with a large prior standard deviation. The prior distribution for E_x is noninformative, $E_x \sim N(0, 0.10^2)$. To set priors for Ω, we divide

$$\Omega = \begin{bmatrix} \sigma_u^2 & \sigma_{uv} & \sigma_{uw} \\ \sigma_{vu} & \sigma_v^2 & \sigma_{vw} \\ \sigma_{wu} & \sigma_{wv} & \sigma_w^2 \end{bmatrix}$$

into two subsets: the (2×2) submatrix Ω_{11}, where

$$\Omega_{11} = \begin{bmatrix} \sigma_u^2 & \sigma_{uw} \\ \sigma_{wu} & \sigma_w^2 \end{bmatrix},$$

and the vector $\Omega_v = (\sigma_v^2, \sigma_{vu}, \sigma_{vw})$. The prior on Ω_{11} is an inverted Wishart distribution

$$\Omega_{11} \sim IW\left(\begin{bmatrix} 0.40^2 & \sigma_{uw} \\ \sigma_{wu} & 0.01^2 \end{bmatrix}, 10\right),$$

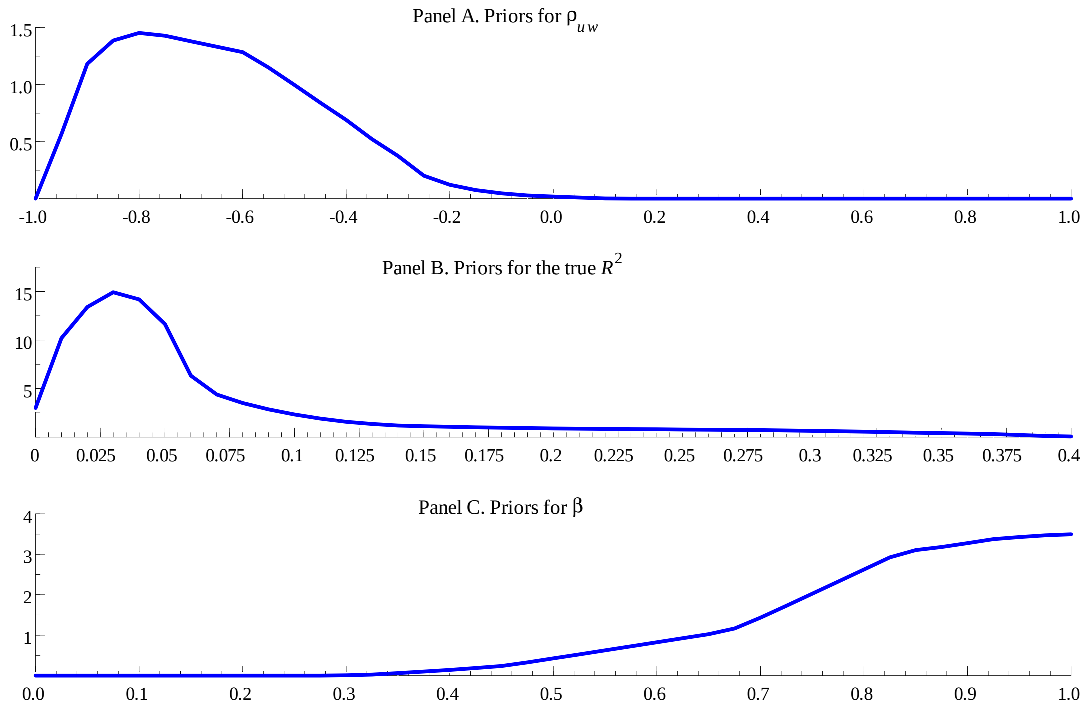
where σ_{uw} is a hyperparameter with a uniform distribution that produces the priors on ρ_{uw} plotted in Figure A1. As discussed in Section B5.1 in the Online Appendix of Pástor and Stambaugh (2009), the prior of Ω_v is determined by the distribution of $C = [\sigma_{vu}, \sigma_{vw}]\Omega_{11}^{-1}$ and $\Psi = \sigma_v^2 - C\Omega_{11}^{-1}C'$. The prior on Ψ is a normal-inverted Wishart, $\Psi \sim IW(0.001, 10)$, and the prior on C is a normal distribution with

$$vec(C) \sim N\left(\begin{bmatrix} -0.033 \\ -0.024 \end{bmatrix}, \begin{bmatrix} 0.00005^2 & 3 \times 10^{-8} \\ 3 \times 10^{-8} & 0.002^2 \end{bmatrix}\right).$$

1.3.2 CV-DC model

For the CV-DC model, we adopt the approach described by Johannes et al. (2014) in their Internet appendix. We train the priors from 1952 to 1954 by regressing excess market returns on a constant and the dividend yield and running an AR(1) process for dividend yield, which corresponds to noninformative priors. Based on these regressions, we have all priors except those for b and σ_η (Equation (13)). We run a rolling window regressions to up to 1954 to have 36 observations of β_t, we then run an AR(1) regression to have prior for b and σ_η.

Figure A1: Prior distributions of parameters



Note: The plots display the prior distributions for the conditional correlation ρ_{uw}, the true R^2 (fraction of the variation in r_{t+1} explained by the observed predictor), and the persistence parameter β.

2 Interval Method

Understanding the effect of parameter estimates and standard errors on the conditional predictive variance can sharpen our understanding on why the non-negative equity premium restriction reduces conditional volatility in the long run. The conditional predictive variance includes five components (see also Equation (11) in the paper):

$$\begin{aligned} Var(R_{T:T+k}|Z_T) &= E(k\sigma_u^2|Z_T) + E(2k\sigma_u^2\bar{d}_{uw}A(k)|Z_T) \\ &+E(k\sigma_u^2\bar{d}^2B(k)|Z_T) + E\left(\left(\frac{1-\beta^k}{1-\beta}\right)^2 q_T|Z_T\right) \\ &+Var\left(kE_r + \frac{1-\beta^k}{1-\beta}(b_T - E_r)|Z_T\right). \end{aligned}$$

The first four components $E(.)$ capture the effect of parameter estimate on conditional predictive variance. The last term $Var(.)$ is a variance term that represents the effect of parameter uncertainty on the predictive variance. It is clear from the equation that parameter uncertainty is affected by both the parameter estimates and the standard errors of parameter estimates. These effects are likely to be highly nonlinear and interdependent. Even we can use the delta method to linearize them, there is still some interaction components. Although it is largely impossible to precisely isolate these effects, we investigate this question using an interval method. First, we compute $Var(.)$ for the unrestricted and restricted predictive systems and denote them by Var_{PS} and Var_{RPS}, respectively. Second, we calculate $Var(.)$ using the parameter estimates from the unrestricted predictive system and the standard errors of parameter estimates from the restricted predictive system. This term, labeled as $Var1$, captures the effect of the standard errors difference on parameter uncertainty. Third, we compute $Var(.)$ using the parameter estimates from the restricted predictive system and the standard errors of parameter estimates from the unrestricted predictive system. We denote this term as $Var2$, which captures the effect of parameter estimate difference on parameter uncertainty. Finally, we define the interdependent term as

$$Var3 = (Var_{PS} - Var_{RPS}) - Var1 - Var2.$$

Now we can decompose the predictive variance difference between the unrestricted and restricted predictive system $D_{Var(R_{T:T+k}|Z_T)}$ into two terms. Let $D_{E(.)}$ denote the difference of the four expectations terms between the unrestricted and restricted predictive system. We have

$$D_{Var(R_{T:T+k}|Z_T)} = D_E(.) + Var1 + Var2 + Var3.$$

To attribute the difference on the predictive variance to the difference in parameter estimates and the difference in the standard errors of parameter estimates, we calculate two ratios:

$$\frac{Var1}{D_{Var(R_{T:T+k}|Z_T)}} \qquad \text{and} \qquad \frac{Var1 + Var3}{D_{Var(R_{T:T+k}|Z_T)}}$$

These two terms represent the lower and upper bounds of the portion of $D_{Var(R_{T:T+k}|Z_T)}$ attributable to the difference in the standard errors of parameter estimates. We find that the effect of $Var3$ is mild. These two ratios are quite stable for $k > 5$ years, we can roughly attribute 38-45% (29-35%) of the predictive variance difference to the difference in the standard errors of parameter estimates for the predictive system (the CV-DC model). If we take a first-order derivative of $Var(R_{T:T+k}|Z_T)$ with respect to various parameters, it is evident that σ_u^2 and β are the most crucial parameters that cause divergence between the unrestricted and restricted models, in particular, in the long run.

3 The Long-Run Risks Model

3.1 The Model

The long-run risks model views stock price fluctuations as a response to changing expectations of long-run consumption growth and its volatility. This model has four salient features. First, there is a persistent and predictable component of consumption growth. This component can be perceived by economic agents and moves stock prices, even though it is difficult to measure using time-series techniques. Second, the volatility of consumption growth has a persistent component. Third, the stock market is a claim to dividends, which differs from the consumption stream but shares the same persistent component. Fourth, economic agents are equipped with Epstein-Zin-Weil preferences (Epstein and Zin, 1989; Weil, 1989).

Bansal and Yaron (2004) describe the dynamics of consumption growth (g_{t+1}) and dividend growth ($g_{d,t+1}$) as follows:

$$\begin{aligned} g_{t+1} &= \mu + x_t + \sigma_t \eta_{t+1}, \\ x_{t+1} &= \rho x_t + \varphi_e \sigma_t e_{t+1}, \\ \sigma^2_{t+1} &= \bar{\sigma}^2 + \nu_1(\sigma^2_t - \bar{\sigma}^2) + \sigma_w w_{t+1}, \\ g_{d,t+1} &= \mu_d + \phi x_t + \varphi_d \sigma_t u_{t+1}, \end{aligned} \tag{A.1}$$

where $\eta_{t+1}, e_{t+1}, w_{t+1}, u_{t+1} \sim i.i.d.N(0,1)$. Here, x_t is the persistently varying component of the expected consumption growth rate. The conditional volatility of consumption, σ^2_{t+1}, is also time-varying and persistent, with an unconditional mean $\bar{\sigma}^2$. The variance process can take negative values, but this will occur only at a low probability if the mean is high enough relative to the volatility of variance. The dividend growth rate imperfectly shares the same persistent and predictable component of consumption growth x_t scaled by parameter ϕ. In addition, the conditional variance of $g_{d,t+1}$ is proportional to the conditional variance of consumption growth, with a scaling parameter φ_d.

Economic agents have Epstein-Zin-Weil preferences with a discount factor δ, relative risk aversion γ, and elasticity of intertemporal substitution ψ. In this setting, the log stochastic discount factor for the economy is given by

$$m_{t+1} = \theta \log \delta - \frac{\theta}{\psi} g_{t+1} + (\theta - 1) r_{a,t+1}, \tag{A.2}$$

where $r_{a,t+1}$ is the return on aggregate wealth and $\theta = (1-\gamma)/(1-1/\psi)$. Bansal and Yaron (2004) use an analytical approximation method to solve the model. They show that the equity premium in the presence of time-varying economic uncertainty is

$$E_t(r_{m,t+1} - r_{f,t}) = \beta_{m,e}\lambda_{m,e}\sigma_t^2 + \beta_{m,w}\lambda_{m,w}\sigma_w^2 - \frac{1}{2}Var_t(r_{m,t+1}), \tag{A.3}$$

where $\beta_{m,e} = \kappa_{1,m}A_{1,m}\varphi_e$, $\beta_{m,w} = \kappa_{1,m}A_{2,m}$, $\lambda_{m,e} = (1-\theta)\kappa_1 A_1 \varphi_e$, $\lambda_{m,w} = (1-\theta)\kappa_1 A_2$, and $Var_t(r_{m,t+1}) = (\beta_{m,e}^2 + \varphi_d^2)\sigma_t^2 + \beta_{m,w}^2\sigma_w^2$. The detailed solutions for A_1, A_2, $A_{1,m}$, and $A_{2,m}$ are defined in the Appendix of Bansal and Yaron (2004).

3.2 Simulation

Our baseline parameter values, which are summarized in Table A1, are consistent with those used in Bansal and Yaron (2004) and Beeler and Campbell (2012). All parameters are given in monthly terms. The monthly persistence of the predictable component of consumption growth is $\rho = 0.979$, implying half-lives between two and three years. We first generate four series of i.i.d. standard normal random variables and use these series to construct the monthly series of consumption, dividends, and state variables based on Equations (A.1). In the simulations, there is a low likelihood that conditional volatility becomes negative. Following Bansal and Yaron (2004) and Beeler and Campbell (2012), we replace negative realizations of conditional variance with a small positive number.

We simulate 1,000 samples, which are generated independently (conditional on the baseline parameters), each consisting of 751 monthly observations to match the sample size in our empirical analysis. To initialize each simulation, we set state variables to their steady-state values and run each simulation for a "burn-in" period of ten years before using the output. For each simulated sample, we estimate the parameters associated with the unrestricted and restricted predictive system. Table A2 reports summary statistics on the key parameters, namely the unconditional equity premium E_r and the persistence parameter β (Panels A and B). Reported statistics are the mean and standard deviation of the parameter estimates over the simulated samples. We note that the parameter estimates are in the ballpark of the estimates reported for our data (Table 2). The unconditional equity premium increases from 0.66% to 0.78% when the NEP condition is imposed to the model.[2] In contrast, the persistence parameter slightly decreases from 0.91 to 0.88. Regarding the CV-DC model (Panels C and D), parameter estimates are barely affected when we impose the NEP condition.

[2]The true (approximate) equity premium given by Equation (A.3) is equal to 0.62%.

Table A1: Long-run Risks Model: Calibrated Parameters

Parameters	Symbol	Value
Endowment Process		
Mean Consumption Growth	μ	0.0015
LRR Persistence	ρ	0.9790
LRR Volatility Multiple	φ_e	0.0440
Mean Dividend Growth	μ_d	0.0015
Dividend Leverage	ϕ	3.0
Dividend Volatility Multiple	φ_d	4.5
Baseline Volatility	$\bar{\sigma}$	0.0078
Volatility of Volatility	σ_w	0.0000023
Persistence of Volatility	ν_1	0.987
Preference Parameters		
Time Discount Factor	δ	0.9980
Risk Aversion	γ	8.5
Elasticity of Intertemporal Substitution	ψ	1.5

Note: The table displays the calibration of parameters in the long-run risks model for simulating the equity premium. The endowment process for the model is given by the system (A.1) in the text. All parameters are given in monthly terms. The standard deviation of long-run innovations is equal to the volatility of consumption growth times the long run volatility of multiple (LRR Volatility Multiple), and the standard deviation of dividend growth innovations is equal to the volatility of consumption growth times the volatility multiple for dividend growth (Dividend Volatility Multiple).

Table A2: Parameter Estimates of Predictive Models under the Long-run Risks Model

Parameters	Mean	Std dev.
Panel A: Unrestricted Predictive System		
E_r	0.66%	0.21
β	0.91	0.24
Panel B: Restricted Predictive System		
E_r	0.78%	0.15
β	0.88	0.17
Panel C: Unrestricted CV-DC Model		
α	0.62%	0.46
β_0	0.32	0.10
β_x	0.96	0.29
Panel D: Restricted CV-DC Model		
α	0.65%	0.42
β_0	0.33	0.08
β_x	0.90	0.22

Note: This table presents the results of a Monte Carlo experiment, designed as follows. First, we use the long-run risks model as the true model to generate a non-negative equity premium: 1,000 samples of returns are generated from this model. Next, we estimate the unrestricted and restricted predictive models using the simulated series. Panels A and B report the parameter estimates for the unrestricted and restricted predictive systems, respectively: the unconditional equity premium parameter (E_r) and the persistence parameter (β). Panels C and D report the parameter estimates for the unrestricted and restricted CV-DC models, respectively: the intercept (α), the sensitivity to dividends (β_0), and the dividend persistence parameter (β_x). For each parameter, we report the mean and the standard deviation over the simulated samples.

References

Bansal, R., Yaron, A., 2004. Risks for the long run. Journal of Finance 59, 1481–1509.

Beeler, J., Campbell, J., 2012. The long-run risks model and aggregate asset prices: An empirical assessment. Critical Finance Review 2012, 141–182.

Carter, C., Kohn, R., 1994. On Gibbs sampling for state space models. Biometrika 81, 541–553.

Epstein, L., Zin, S., 1989. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. Econometrica 57, 937–968.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo, J., Berger, J., Dawid, A., Smith, A. (Eds.), Bayesian Statistics. Vol. 4. Oxford University Press, pp. 169–193.

Johannes, M., Korteweg, A., Polson, N., 2014. Sequential learning, predictability, and optimal portfolio returns. Journal of Finance 69, 611–644.

Kim, C.-J., Nelson, C., 1999. State-Space Models with Regime Switching. MIT Press, Cambridge, Massachusetts.

Pástor, L., Stambaugh, R., 2009. Predictive systems: Living with imperfect predictors. Journal of Finance 64, 1583–1627.

Pástor, L., Stambaugh, R., 2012. Are stocks really less volatile in the long run? Journal of Finance 67, 431–477.

Raftery, A., Lewis, S., 1992. How many iterations in the Gibbs sampler? In: Bernardo, J., Berger, J., Dawid, A., Smith, A. (Eds.), Bayesian Statistics. Vol. 4. Oxford University Press, pp. 763–773.

Roberts, G., Gelman, A., Gilks, W., 1997. Weak convergence and optimal scaling of random walk Metropolis algorithm. Annals of Applied Probability 7, 110–120.

Weil, P., 1989. The equity premium puzzle and the risk-free rate puzzle. Journal of Monetary Economics 24, 401–421.

Zhu, X., 2015. Tug-of-war: Time-varying predictability of stock returns and dividend growth. Review of Finance 19, 2359–2399.