Predictability and the Cross-Section of Expected Returns:

A Challenge for Asset Pricing Models

Online Appendix

This version: September 14, 2020

Contents

A1 Robustness of the empirical results

A1.1 Small firms

We start our robustness analyses by checking if our results are potentially driven by small firms. To this end, in each month, we exclude 20% of the stocks with the lowest market capitalization and repeat our sorting exercise. The results shown in Table A1 are very similar to the results for the base case presented in Table 1 of the paper. We therefore conclude that our main findings are not caused by the presence of small stocks in our sample.

A1.2 Return horizons and formation period

We now study whether the choice of the return horizon and the length of the formation period have an impact on the results. In the benchmark case of regressions (1) and (2), we chose return horizons of 4 and 12 months, since the variance risk premium is known to predict returns over shorter horizons of 3-6 months, while the price-dividend ratio is informative over horizons of one year and more. Given a certain formation period (like the 72 months in our benchmark case from Section 3 in the paper), however, a longer return horizon means fewer observations. The choice of the length of the formation period is a trade-off between the need to account for possibly time-varying loadings of expected stock returns on the different risk factors, favoring a shorter formation period, and the need to have sufficient data to run the necessary regressions.

We analyze the power of the two predictors over different horizons by looking at the t-statistics and R^2's of the corresponding regressions. The upper two plots in Figure A1 show the cross-sectional median t-statistics of predictive regressions using the price-dividend ratio and the variance risk premium, respectively, together with 95% confidence bands, averaged over all formation periods. The dotted line represents the t-statistic for the coefficient in the regression for the market return. The lower two plots show the R^2's.

The cross-sectional median t-statistics as well as the median R^2's increase monotonically from one to twelve months for the price-dividend ratio (and even for some more horizons thereafter). For the variance risk premium, the median t-statistic peaks at four months, just as for the market portfolio. The latter result is well-known, see e.g. Bali and Zhou (2016), Bekaert and Hoerova (2014), Bollerslev et al. (2009), Bollerslev et al. (2014), Drechsler and Yaron (2011), or Londono (2015). In addition, the R^2 exhibits a local maximum at the four month-horizon.

Table A4 shows that the choice of the return horizon is not material. We show HML returns and t-statistics for horizons between 3 and 36 months in case of the price-dividend ratio and between 1 and 12 months for the variance risk premium. Furthermore, we vary the formation period to be 48 and 96 months, respectively.

The return spread between the extreme b_ω-portfolios is significantly positive for return horizons between 6 and 18 months. With longer return horizons, the spread becomes insignificant, because there are not enough independent return observations to identify the slope coefficient reliably.

The spread between high- and low-b_ν portfolios is significantly positive for most return horizons less than six months for formation periods of 72 and 96 months. This is in line with the intuition from Figure A1, which showed that the variance risk premium predicts returns over short horizons of up to six months.

A1.3 Alternative regression design

Bansal et al. (2005) run regressions of dividend growth rates on the trailing multi-period moving average of past consumption growth to capture the low frequency component in consumption growth and its relation to the variable to be predicted. Inspired by this approach, we run the regressions

$$r^i_{\tau,\tau+12} - r^f_{\tau,\tau+12} = a_{\omega,i,t} + b_{\omega,i,t} \sum_{j=0}^{11} \omega_{\tau+j} + e^{\omega,i,t}_{\tau+12}, \tag{A1}$$

$$r^i_{\tau,\tau+4} - r^f_{\tau,\tau+4} = a_{\nu,i,t} + b_{\nu,i,t} \sum_{j=0}^{3} \nu_{\tau+j} + e^{\nu,i,t}_{\tau+4}, \tag{A2}$$

and sort stocks into portfolios according to the slope coefficients b_ω and b_ν from these regressions.[1] Returns on b_ω- and b_ν-decile portfolios are presented in Table A2. They are very similar to the results in Table 1. Thus, our findings are robust to variations in the regression design.

A1.4 Rebalancing intervals

A standard procedure in empirical asset pricing is to sort stocks into portfolios only once a year and track these stocks over the subsequent 12 months. For example, Fama and French (1992) form portfolios once a year at the end of June. The reason is that they sort on the book-to-market ratio of equity which is based on balance sheet information available only once per year. Since b_ω and b_ν coefficients are based on past returns, just like momentum, there is no obvious reason to rebalance portfolios less frequently than once per month.

Still, instead of sorting every month, we also perform sorts once per quarter, semi-annually, and annually and hold portfolios until the next sorting date. The results are shown in Table A3. Here, we use the sample from 1933 until 2018. We find that the positive returns on the HML portfolios are not specific to the rebalancing interval. This result is not surprising given that the portfolio composition is very persistent, as discussed in Section 4.1 of the paper.

[1]We thank an anonymous referee for suggesting this robustness test.

A2 The model of Bansal and Yaron (2004)

A2.1 Model setup and solution

In general, the long run risks framework combines two key assumptions. First, there is a representative agent who entertains recursive preferences as developed by Kreps and Porteus (1978), Epstein and Zin (1989), and Weil (1989). This implies that the log pricing kernel $m_{t,t+1}$ is given by

$$m_{t,t+1} = -\delta\theta - \frac{\theta}{\psi}\Delta c_{t+1} + (\theta - 1)r^c_{t,t+1}, \tag{A3}$$

where $\theta = (1-\gamma)/(1-\psi^{-1})$, $\Delta c_{t+1} = \log(C_{t+1}) - \log(C_t)$ denotes log consumption growth, and $r^c_{t,t+1}$ is the log return on aggregate wealth, i.e., the return on the portfolio that pays aggregate consumption as dividends. The investor's preferences are characterized by the time-discount rate δ, the elasticity of intertemporal substitution (EIS) ψ, and the Arrow-Pratt measure of relative risk aversion γ.

The second key assumption is that there are persistent components driving the first two moments of growth rates in aggregate cash-flows. In particular, the changes in log consumption Δc_{t+1}, log aggregate dividends Δd^m_{t+1}, and the state variables are given by

$$\Delta c_{t+1} = \mu_c + x_t + \sqrt{V_t}\,\varepsilon^c_{t+1}, \tag{A4a}$$

$$\Delta d^m_{t+1} = \mu_m + \kappa_m\, x_t + \phi_m\sqrt{V_t}\,\varepsilon^m_{t+1}, \tag{A4b}$$

$$x_{t+1} = \kappa_x x_t + \phi_x\sqrt{V_t}\,\varepsilon^x_{t+1}, \tag{A4c}$$

$$V_{t+1} = \bar{V} + \kappa_V\left(V_t - \bar{V}\right) + \phi_V\sqrt{V_t}\,\varepsilon^V_{t+1}, \tag{A4d}$$

where ε^c, ε^m, ε^x, and ε^V are mutually independent Gaussian innovations. As opposed to the original model of Bansal and Yaron (2004), we use a square-root specification for V, as in Tauchen (2011). This implies that, in the continuous-time limit, the variance process is non-negative. Moreover, the model-implied variance risk premium is time-varying and affine in V_t (see below). We furthermore assume that log growth in dividends on stock i evolves as

$$\Delta d^i_{t+1} \quad = \mu_i + \kappa_i \, x_t + \phi_i \sqrt{V_t} \, \varepsilon^i_{t+1}. \tag{A5}$$

The innovations ε^i may in general have a non-trivial cross-sectional correlation structure, but for simplicity, we assume that they are mutually independent.

We first consider the *aggregate wealth portfolio* that pays aggregate consumption as dividends. For the solution, we rely on linearization techniques, which have recently been criticized for their lack of accuracy, see Kraft et al. (2015) and Pohl et al. (2018). For our purposes, however, the log-linearization is sufficient, since we are interested in the time series relation between different quantities and not in exactly calculating unconditional asset pricing moments. Applying the log-linearization of Campbell and Shiller (1988) to the return on the wealth portfolio r^c gives

$$r^c_{t,t+1} = k_{c,0} + k_{c,1}\omega^c_{t+1} - \omega^c_t + \Delta c_{t+1},$$

where ω^c_t is the log wealth-consumption ratio, $k_{c,1} = \frac{\exp(\bar{\omega^c})}{1+\exp(\bar{\omega^c})}$, and $k_{c,0} = \log(1 + \exp(\bar{\omega^c})) - k_{c,1}\bar{\omega^c}$. Substitution of this and the pricing kernel (see Equation (A3)) into the Euler Equation and conjecturing $\omega^c_t = A_{c,0} + A_{c,x}x_t + A_{c,V}V_t$ yields

$$\begin{aligned} 1 &= \mathbb{E}_t[\exp(m_{t,t+1} + r_{t,t+1})] \\ &= \exp\Big(\theta\delta + (1-\gamma)\mu_c + \theta k_{c,0} + \theta(k_{c,1}-1)A_{c,0} + \theta k_{c,1}A_{c,V}\bar{V} \\ &\qquad + (1-\gamma)x_t + \theta(k_{c,1}\kappa_x - 1)A_{c,x}x_t \\ &\qquad + \theta(k_{c,1}\kappa_V - 1)A_{c,V}V_t + \frac{1}{2}(1-\gamma)^2 V_t + \frac{1}{2}\theta^2 k^2_{c,1}(A^2_{c,x}\phi^2_x + A^2_{c,V}\phi^2_V)V_t\Big). \end{aligned}$$

Solving for $A_{c,0}$, $A_{c,x}$, and $A_{c,V}$ gives

$$A_{c,0} = \frac{1}{1-k_{c,1}}\Big(\delta + (1-1/\psi)\mu_c + k_{c,0} + k_{c,1}A_{c,2}(1-\kappa_V)\bar{V}\Big)$$
$$A_{c,x} = \frac{1-1/\psi}{1-k_{c,1}\kappa_x}.$$

$A_{c,V}$ is one of the solutions of the quadratic equation

$$0 = \frac{1}{2}\theta^2 k_{c,1}^2 A_{c,V}^2 \phi_V^2 + \theta(k_{c,1}\kappa_V - 1)A_{c,V} + \frac{1}{2}(1-\gamma)^2 + \frac{1}{2}\theta^2 k_{c,1}^2 A_{c,x}^2 \phi_x^2. \tag{A6}$$

We follow Tauchen (2011) and select the root for which $\phi_V^2 A_{c,V} \to 0$ when $\phi_V^2 \to 0$. We solve for $A_{c,0}$, $A_{c,x}$, $A_{c,V}$, and $k_{c,0}$ conditional on a candidate $k_{c,1}$ and search for the $k_{c,1}$ that solves

$$k_{c,1} = (\exp(A_{c,0} + A_{c,V}\bar{V}))/(1+\exp(A_{c,0} + A_{c,V}\bar{V})).$$

Substituting the resulting values into Equation (A3) gives the following representation of the pricing kernel:

$$m_{t,t+1} = \mathbb{E}_t[m_{t,t+1}] - \Lambda_c \varepsilon_{t+1}^c - \Lambda_x \varepsilon_{t+1}^x - \Lambda_V \varepsilon_{t+1}^V,$$

where

$$\Lambda_c = \gamma\sqrt{V_t}$$
$$\Lambda_x = (1-\theta)k_{c,1}A_{c,x}\phi_x\sqrt{V_t}$$
$$\Lambda_V = (1-\theta)k_{c,1}A_{c,V}\phi_V\sqrt{V_t}.$$

For a dividend process

$$\Delta d_{t+1}^i = \mu_i + \kappa_i\, x_t + \phi_i\,\sqrt{V_t}\,\varepsilon_{t+1}^i,$$

we can, as before, assume an affine structure of the corresponding price-dividend ratio, i.e.

$\omega_t^i = A_{i,0} + A_{i,x}x_t + A_{i,V}V_t$, and use the Campbell and Shiller (1988) approximation to log-linearize the return on the dividend claim. The corresponding Euler equation gives rise to the following system of equations:

$$A_{i,0} = \frac{1}{1-k_{i,1}}\Big(\theta\delta + \mu_d - \gamma\mu_c + (\theta-1)k_{c,0} + k_{i,0} + (\theta-1)(k_{c,1}-1)A_{c,0} + \Big((\theta-1)k_{c,1}A_{c,V} + k_{i,1}A_{i,V}\Big)(1-\kappa_V)\bar{V}\Big) \tag{A7a}$$

$$A_{i,x} = \frac{\kappa_i - 1/\psi}{1-k_{i,1}\kappa_x} \tag{A7b}$$

$$0 = \frac{1}{2}k_{i,1}^2 A_{i,V}^2\phi_V^2 + \Big((k_{i,1}\kappa_V - 1) + (\theta-1)(k_{c,1}k_{i,1}A_{c,V}\phi_V^2)\Big)A_{i,V} + (\theta-1)(k_{c,1}\kappa_V - 1)A_{c,V} + \frac{1}{2}\gamma^2 + \frac{1}{2}\phi_i^2 + \frac{1}{2}\Big((\theta-1)^2k_{c,1}^2A_{c,x}^2 + k_{i,1}A_{i,x}\Big)^2\phi_x^2 + \frac{1}{2}(\theta-1)^2k_{c,1}^2A_{c,x}^2\phi_V^2. \tag{A7c}$$

We proceed as above in selecting the economically meaningful root. The expected excess return on asset i is

$$\mathbb{E}_t\left[r_{t,t+1}^i\right] - r_{t,t+1}^f = -Cov_t(m_{t,t+1}, r_{t,t+1}^i) - \frac{1}{2}Var_t(r_{t,t+1}^i) = \pi_{i,V}V_t, \tag{A8}$$

where

$$Cov_t(m_{t,t+1}, r_{t,t+1}^i) = \Lambda_c\phi_i\sqrt{V_t} + \Lambda_x k_{i,1}A_{i,x}\phi_x\sqrt{V_t} + \Lambda_V k_{i,1}A_{i,V}\phi_V\sqrt{V_t}$$
$$Var_t(r_{t,t+1}^i) = \phi_i^2V_t + (k_{i,1}A_{i,x}\phi_x)^2V_t + (k_{i,1}A_{i,V}\phi_V)^2V_t$$

such that $\pi_{i,V}$ in Equation (A8) is

$$\pi_{i,V} = (1-\theta)k_{c,1}k_{i,1}A_{c,x}A_{i,x}\phi_x^2 + (1-\theta)k_{c,1}k_{i,1}A_{c,V}A_{i,V}\phi_V^2 - \frac{1}{2}\Big(\phi_i^2 + (k_{i,1}A_{i,x}\phi_x)^2 + (k_{i,1}A_{i,V}\phi_V)^2\Big).$$

Since V_t is the only uncertainty related state variable in this model economy, risk premia on all stocks in the economy are multiples of each other and thus move in lockstep. However, there are differences in the sizes of innovations in expected returns which are quantified by the coefficients $\pi_{i,V}$.

The coefficient $b_{\omega,i}$ in the linear regression

$$r^i_{t,t+12} - r^f_{t,t+12} = a_{\omega,i} + b_{\omega,i}\omega^m_t + e^{\omega,i,t}_{t+12},$$

equals the unconditional covariance of the excess return on asset i and the price-dividend ratio of the market portfolio divided by the unconditional variance of the price-dividend ratio. These terms can be calculated in closed form within the model:

$$\begin{aligned} b_{\omega,i} &= \frac{Cov\left(r^i_{t,t+12} - r^f_{t,t+12}, \omega^m_t\right)}{Var(\omega^m_t)} = \frac{Cov\left(\mathbb{E}_t\left[r^i_{t,t+12} - r^f_{t,t+12}\right], \omega^m_t\right)}{Var(\omega^m_t)} \\ &= \pi_{i,V} A_{m,V} \frac{Var(V_t)}{Var(\omega^m_t)} \sum_{\tau=0}^{11} \kappa^\tau_V, \end{aligned} \tag{A9}$$

where Cov and Var denote unconditional moments. The unconditional variance of the price-dividend ratio is $Var(\omega^i_t) = A^2_{i,x}Var(x_t) + A^2_{i,V}Var(V_t)$, where $Var(x_t) = \frac{\phi_x \bar{V}}{1-\kappa^2_x}$ and $Var(V_t) = \frac{\phi_V \bar{V}}{1-\kappa^2_V}$. With the definition

$$\lambda_\omega = \left(A_{m,V}\frac{Var(V_t)}{Var(\omega^m_t)}\sum_{\tau=0}^{11}\kappa^\tau_V\right)^{-1}\bar{V} \tag{A10}$$

and using Equation (A9), it is easy to see that

$$\mathbb{E}\left[r^i_{t,t+1} - r^f_{t,t+1}\right] = \pi_{i,V}\bar{V} = \lambda_\omega \cdot b_{\omega,i}.$$

The variance risk premium is given by

$$\nu_t = \mathbb{E}_t^{\mathcal{Q}}[Var_{t+1}(r_{t+2}^m)] - \mathbb{E}_t^{\mathcal{P}}[Var_{t+1}(r_{t+2}^m)],$$

where the conditional return variance is

$$Var_t(r_{t+1}^m) = \left(\phi_m^2 + (k_{m,1}A_{m,x}\phi_x)^2 + (k_{m,1}A_{m,V}\phi_V)^2\right) V_t.$$

$Var_{t+1}(r_{t+2}^m)$ is Gaussian because V_{t+1} is Gaussian. The log pricing kernel is Gaussian as well. We can thus write

$$\begin{aligned}
&\mathbb{E}_t^{\mathcal{Q}}[Var_{t+1}(r_{t+2}^m)] \\
&\quad = \log\left(\mathbb{E}_t^{\mathcal{Q}}[\exp(Var_{t+1}(r_{t+2}^m))]\right) - \frac{1}{2}Var_t(Var_{t+1}(r_{t+2}^m)) \\
&\quad = \log\left(\mathbb{E}_t^{\mathcal{P}}[\exp(m_{t,t+1} + Var_{t+1}(r_{t+2}^m))]\right) - \log\left(\mathbb{E}_t^{\mathcal{P}}[\exp(m_{t,t+1})]\right) - \frac{1}{2}Var_t(Var_{t+1}(r_{t+2}^m)) \\
&\quad = \mathbb{E}_t^{\mathcal{P}}[m_{t,t+1}] + \mathbb{E}_t^{\mathcal{P}}[Var_{t+1}(r_{t+2}^m)] + \frac{1}{2}Var_t(m_{t,t+1}) + \frac{1}{2}Var_t(Var_{t+1}(r_{t+2}^m)) \\
&\qquad + Cov_t(m_{t,t+1}, Var_{t+1}(r_{t+2}^m)) - \mathbb{E}_t^{\mathcal{P}}[m_{t,t+1}] - \frac{1}{2}Var_t(m_{t,t+1}) - \frac{1}{2}Var_t(Var_{t+1}(r_{t+2}^m)).
\end{aligned}$$

It follows that the variance risk premium is

$$\begin{aligned}
\nu_t = B_{m,V}V_t &= Cov_t(m_{t,t+1}, Var_{t+1}(r_{t+2}^m)) \\
&= k_{c,1}(\theta-1)A_{c,V}\left(\phi_m^2 + (k_{m,1}A_{m,x}\phi_x)^2 + (k_{m,1}A_{m,V}\phi_V)^2\right)\phi_V^2\ V_t.
\end{aligned} \tag{A11}$$

Again, the slope coefficient in a predictive regression is given by

$$b_{\omega,i} = \pi_{i,V}B_{m,V}\,\frac{Var(V_t)}{Var(\nu_t^m)}\,\sum_{\tau=0}^{3}\kappa_V^\tau, \tag{A12}$$

and, using the definition

$$\lambda_\nu = \left(B_{m,V} \frac{Var(V_t)}{Var(\nu_t^m)} \sum_{\tau=0}^{3} \kappa_V^\tau \right)^{-1} \bar{V}, \tag{A13}$$

we find that

$$\mathbb{E}\left[r_{t,t+1}^i - r_{t,t+1}^f \right] = \pi_{i,V} \bar{V} = \lambda_\nu \cdot b_{\nu,i}.$$

A2.2 The signs of λ's

Regarding the relation between the agent's preferences and the signs of λ_ν and λ_ω, we can make the following statement:

$$\lambda_\nu < 0 \text{ if and only if } \theta(\theta - 1) < 0.$$

The condition w.r.t. θ holds if either $1 < \gamma < 1/\psi$ or $1 > \gamma > 1/\psi$. Assuming that the investor is sufficiently risk averse ($\gamma > 1$), the statement implies that $\lambda_\nu > 0$ if and only if the investor has a preference for early resolution of uncertainty. To prove this statement we rearrange Equation (A6):

$$\theta A_{c,V} = \frac{1}{1 - k_{c,1}\kappa_V} \left(\frac{1}{2}\theta^2 k_{c,1}^2 A_{c,V}^2 \phi_V^2 + \frac{1}{2}(1-\gamma)^2 + \frac{1}{2}\theta^2 k_{c,1}^2 A_{c,x}^2 \phi_x^2 \right).$$

Hence, $\theta A_{c,V} > 0$, i.e., $A_{c,V}$ is positive if and only if θ is positive because $k_{c,1}$ and κ_V are both between 0 and 1. From Equation (A11) we can deduce that $B_{m,V}$ has the same sign as $(\theta - 1)A_{c,V}$. Together with Equation (A13), this proves the statement.

λ_ω is more difficult to analyze, since $A_{m,V}$ does not necessarily have the same sign as the market price of V-risk. We vary the preference parameters γ and ψ, while keeping all other parameters as reported in Table A5 and analyze the coefficient $A_{m,V}$. Figure A3 shows it as a function of γ and $1/\psi$. The dark area indicates parameter combinations for which $A_{m,V}$, and, hence, λ_ω is negative. The white cross marks the parameter choice in Bansal and Yaron (2004). Those parameter combinations that are located above the 45 degree line indicate a preference for late resolution of uncertainty. The following statement holds irrespective of the choice of the structural parameters.

$$\text{If } \theta(\theta - 1) < 0 \text{ then } \lambda_\omega > 0.$$

If the investor is sufficiently risk averse ($\gamma > 1$), a preference for late resolution of uncertainty (i.e. $1/\psi > \gamma$) implies a positive relation between expected returns $\mathbb{E}_t[r^i_{t,t+h}]$ and coefficients $b_{\omega,i}$ in the cross-section. The converse does not hold in general.

To prove this statement we rewrite Equation (A7c) for the case $i = m$ as

$$0 = aA^2_{m,V} + bA_{m,V} + c,$$

The coefficients are

$$\begin{aligned}
a &= \frac{1}{2}k^2_{m,1}\phi^2_V \\
b &= (k_{m,1}\kappa_V - 1) + (\theta - 1)(k_{c,1}k_{m,1}A_{c,V}\phi^2_V) \\
c &= (\theta - 1)(k_{c,1}\kappa_V - 1)A_{c,V} + \frac{1}{2}\gamma^2 + \frac{1}{2}\phi^2_m + \frac{1}{2}\Big((\theta - 1)^2k^2_{c,1}A^2_{c,x} + k_{m,1}A_{m,x}\Big)^2\phi^2_x \\
&\quad + \frac{1}{2}(\theta - 1)^2k^2_{c,1}A^2_{c,x}\phi^2_V.
\end{aligned}$$

The first term of b is always negative and the second is negative iff $(\theta-1)\theta$ is negative, according to the above argument. In this case, the first term of c is positive, just as the remaining terms which are all quadratic. We can thus conclude that

$$(\theta - 1)\theta < 0 \Rightarrow b < 0 \text{ and } c > 0.$$

The converse does not hold. According to the root selection criterion discussed above, the solution that we consider is

$$A_{m,V} = \frac{1}{2a}\left(-b + \text{sign}(b)\sqrt{b^2 - 4ac}\right).$$

If $(\theta - 1)\theta < 0$ then $b < 0$ and we have to subtract the root. a and c are positive, so $\sqrt{b^2 - 4ac} < -b$, and thus $A_{m,V}$ must be positive.

A2.3 Simulation studies

In this section, we describe how we simulate returns from the model described above and the version of the model of Segal et al. (2015), described in Section A4 of this Online Appendix. We always start by solving the respective model, i.e., we calculate the coefficients A_x, A_V, A_W, B_V, B_W, $\pi_{m,V}$, and $\pi_{m,W}$. We then simulate time series of the state variables, together with consumption growth and dividend growth of the market portfolio. To avoid negative realizations of the variance processes, we simulate these processes on a daily basis. In case a realization of a variance process is still negative, we set it to a very small positive value.

To generate a cross-section of stock returns, we take $\pi_{m,V}$ (and $\pi_{m,W}$ if necessary) as a starting point and generate a cross-section of V- (and W-) loadings by drawing 7,000 normally distributed random numbers $(\varepsilon_i)_{i=1,\dots,I}$ per factor. We set $\pi_{i,V} = f\pi_{m,V} + (\frac{1}{2}\pi_{m,V}) \cdot \varepsilon_i$ (analogously for W), where f is a scaling factor. We proceed similarly to generate return volatility coefficients of all the 7,000 stocks in the cross-section. We then simulate a time-series of excess returns on asset i via the equation

$$r_{i,t+1} - r_{f,t+1} = \pi_{i,V} V_t + \pi_{i,W} W_t + \text{volatility coefficients} \times \text{shocks}.$$

The last term depends on the model specification. We simulate time series with a length of 25 years plus a 10 year burn-in period that is discarded directly after the simulation to make the time series less dependent on the starting values (which we set to the unconditional means of the state variables).

We then compare the value weighted market portfolio from our cross-section with the market portfolio as defined by the dynamics given in the respective model, and use the scaling factors for fine-tuning.

We run predictive regressions with the price-dividend ratio and the variance risk premium as predictors and the market return as the dependent variable. We then scale the return volatility (i.e. we multiply the noise term in the above equation with a constant) until the R^2 of the predictive regression with the variance risk premium is the same as in the data over the four months horizon. We then proceed similarly with the coefficient of x in the price-dividend ratio until the R^2 of the predictive regression with the price-dividend ratio is the same as in the data over the twelve months horizon. We use the same constant to scale the volatility of the stock returns in the cross-section. It is important to notice that we largely reduce the return volatility by this step. This leads to very large t-statistics in some cases (see, e.g., Panel A in Table A6 or Table A8).

A2.4 Quantitative analysis

As detailed above, the calibration of Bansal and Yaron (2004), given in the top line of Table A5, yields $\lambda_\nu > 0$ and $\lambda_\omega < 0$. Note that there is a slight difference between our model and the original one, which is the time varying variance of V. We choose the parameter ϕ_V in Equation (A4a) such that $\phi_V^2 \bar{V}$ matches the constant variance of V in the model of Bansal and Yaron (2004), so that the additional component in our model has only a small impact on unconditional asset pricing moments. Solving the model yields the well-known result that $B_{m,V}$ is negative. This implies that stocks with high b_ω's must on average have lower returns than those with low b_ω's. For the coefficients with respect to the variance risk premium it is the other way around. With these parameters, the market price of V-risk is negative. For the variance risk premium this means the following: Since the variance of V loads positively on V, due to the square-root specification, the variance risk premium, i.e., the $\mathcal{Q}-\mathcal{P}$-difference in the expected return variance, loads positively on V (see the discussion in the above Section A2.1).

The positive λ_ν is in line with the data. However, λ_ω is also positive in the data while it is negative in the model due to the negative $A_{m,V}$. We substantiate this analytic finding with a simulation study, as described in the previous section, and then apply the same algorithm as the one described in Section 3 of the paper. The results, shown in Panel A of Table A6, corroborate the earlier assessment. Thus, we can reject the one-factor long-run risks model with the suggested calibration.

Bansal and Yaron (2004) argue that the sign of $A_{m,V}$ depends on the timing preferences of the representative agent. We thus also consider an agent with a preference for late resolution of uncertainty, in particular $\gamma = 2$ and $\psi = 0.25$. This, indeed, yields $A_{m,V} > 0$ and thus $\lambda_\omega > 0$, but, at the same time $\lambda_\nu < 0$ (see Panel B in Table A6).

Our analysis in the above subsection suggests that there are alternative calibrations of the model which allow both $\lambda_\omega > 0$ and $\lambda_\nu > 0$. With such parameters, we find positive returns on high minus low portfolios, no matter if we use slope coefficients from predictive regressions on the price-dividend ratio or the variance risk premium for sorting. In this case, a high-$b_{\omega,i}$ stock must also be a high $\pi_{i,V}$-stock. The same is true for high-$b_{\nu,i}$ stocks. As a consequence, high-$b_{\omega,i}$ stocks must always also be high-$b_{\nu,i}$ stocks. Thus, in independent double sorts on $b_{\omega,i}$ and $b_{\nu,i}$, all stocks must group in portfolios around the main diagonal. We perform independent double sorts on our simulated returns and find exactly this pattern (see Panel C of Table A7).

A3 The model of Drechsler and Yaron (2011)

Drechsler and Yaron (2011) suggest an extension of the long-run risks model which is able to explain the high and time-varying variance risk premium and its predictive power in the data. For this reason, the model is a natural candidate for our analysis. We consider a variant of the original model, in which the variance process V is heteroskedastic. The authors assume large non-Gaussian innovations (jumps) in the two state variables x_t and V_t and a time-varying central tendency of the consumption variance which we call W_t in the following. Both ingredients have been used by other authors before. Jumps were used in general equilibrium contexts by Barro (2006), Rietz (1988), Wachter (2013), and, in a setting with recursive preferences, by Benzoni et al. (2011). A stochastic central tendency has also been considered by Branger and Völkert (2011) and Jin (2015). The dynamics of consumption, dividends, and the state variables are given as follows:

$$\Delta c_{t+1} = \mu_c + x_t + \sqrt{V_t}\varepsilon^c_{t+1} \tag{A14a}$$

$$\Delta d^m_{t+1} = \mu_m + \kappa_m x_t + \phi_m \sqrt{V_t}\varepsilon^m_{t+1} \tag{A14b}$$

$$x_{t+1} = \kappa_x x_t + \phi_x \sqrt{V_t}\varepsilon^x_{t+1} + J_{x,t+1} \tag{A14c}$$

$$V_{t+1} = W_t + \kappa_V (V_t - W_t) + \phi_V \sqrt{V_t}\varepsilon^V_{t+1} + J_{V,t+1} \tag{A14d}$$

$$W_{t+1} = W + \kappa_W (W_t - W) + \phi_W \sqrt{W_t}\varepsilon^W_{t+1} \tag{A14e}$$

Note that the dynamics include jumps in the expected growth rate and in the variance component V. Jumps in the expected growth rate are normally distributed with zero mean and variance σ^2_{Jx}. Jumps in V are exponentially distributed with mean λ_{JV}. The moment generating functions are given by

$$l_{x,t}(u) = \mathbb{E}_t[e^{uJ_{x,t+1}}] = e^{0.5\sigma^2_{J_x}u^2}$$
$$l_{V,t}(u) = \mathbb{E}_t[e^{uJ_{V,t+1}}] = (1 - \lambda_{JV}u)^{-1}$$

By $l^{(j)}_{i,t}(0)$, we denote the j-th derivative of the moment generating function $l_{i,t}(u)$ evaluated at $u = 0$. Furthermore, both jump intensities are affine in V, i.e. $\iota_x V_t$ and $\iota_V V_t$.

In terms of investor preferences, the model makes the same assumptions as the long-run risks model by Bansal and Yaron (2004).

To solve the model, we follow Drechsler and Yaron (2011). We assume that the wealth consumption ratio is affine in the state variables

$$\omega^c_t = A_{c,0} + A_{c,x}x_t + A_{c,V}V_t + A_{c,W}W_t$$

with coefficients

$$A_{c,0} = \frac{1}{1 - k_{c,1}}\left(\delta + k_{c,0} + (1 - \psi^{-1})\mu_c + k_{c,1}A_{c,W}(1 - k_W)\bar{W}\right)$$
$$A_{c,x} = \frac{1 - 1/\psi}{1 - k_{c,1}\kappa_x}$$

$A_{c,V}$ and $A_{c,W}$ solve the quadratic equations

$$0 = \frac{1}{2}(\theta k_{c,1}\phi_V)^2 A^2_{c,V} + \theta(k_{c,1}\kappa_V - 1)A_{c,V} + \frac{1}{2}(\theta k_{c,1}A_{c,x}\phi_x)^2 + \frac{1}{2}(1 - \gamma)^2$$
$$+ \iota_x(l_x(\theta k_{c,1}A_{c,x}) - 1) + \iota_V(l_V(\theta k_{c,1}A_{c,V}) - 1)$$
$$0 = \frac{1}{2}(\theta k_{c,1}\phi_W)^2 A^2_{c,W} + \theta(k_{c,1}\kappa_W - 1)A_{c,W} + (1 - \kappa_V)\theta k_{c,1}A_{c,V}.$$

The log pricing kernel is

$$
\begin{aligned}
m_{t,+} = m_0 + m_x x_t + m_V V_t + m_W W_t - \Lambda_c \sqrt{V_t}\epsilon^c_{t+1} - \Lambda_x \phi_x \sqrt{V_t}\epsilon^x_{t+1} - \Lambda_x J_{x,t+1} \\
- \Lambda_V \phi_V \sqrt{V_t}\epsilon^V_{t+1} - \Lambda_V J_{V,t+1} - \Lambda_W \phi_W \sqrt{W_t}\epsilon^W_{t+1}
\end{aligned}
$$

with coefficients

$$
\begin{aligned}
m_0 &= \theta\delta + (\theta - 1)k_{c,0} + (\theta - 1)(k_{c,1} - 1)A_{c,0} \\
m_x &= (1 - \theta)A_{c,x} \\
m_V &= (1 - \theta)A_{c,V} \\
m_W &= (1 - \theta)A_{c,W}
\end{aligned}
$$

and the market prices of risk are

$$
\begin{aligned}
\Lambda_c &= \gamma \\
\Lambda_x &= (1 - \theta)k_{c,1}A_{c,x} \\
\Lambda_V &= (1 - \theta)k_{c,1}A_{c,V} \\
\Lambda_W &= (1 - \theta)k_{c,1}A_{c,W}.
\end{aligned}
$$

We assume that the dividend growth dynamics of asset i are also specified as in Equation (A14b). The coefficients of the price-dividend ratio are

$$
\begin{aligned}
A_{i,0} &= \frac{1}{1 - k_{i,1}} \left(m_0 + k_{i,0} - \gamma\mu_c + (k_{i,1}A_{i,W} - \Lambda_W)(1 - \pi_W)W + \mu_d\right) \\
A_{i,x} &= \frac{\kappa_i - 1/\psi}{1 - k_{i,1}\kappa_x},
\end{aligned}
$$

and $A_{i,V}$ and $A_{i,W}$ solve the quadratic equations

$$0 = (k_{i,1} - A_{i,V} - \Lambda_V)\kappa_V - A_{i,V} + \frac{1}{2}(\gamma^2(k_{i,1}A_{i,x} - \Lambda_x)^2\phi_x^2 + (k_{i,1}A_{i,V} - \Lambda_V)^2\phi_V^2 + \phi_i^2)$$
$$+ \iota_x(l_x(k_{i,1}A_{i,x} - \Lambda_x) - 1) + \iota_V(l_V(k_{i,1}A_{i,V} - \Lambda_V) - 1) + m_V$$
$$0 = (k_{i,1}A_{i,W} - \Lambda_W)\kappa_W + (k_{c,1}A_{i,W} - \Lambda_V)(1 - \pi_V)$$
$$+ \frac{1}{2}((k_{i,1}A_{i,W} - \Lambda_W)^2\phi_W^2 - A_{i,W} + m_W$$

The expected excess return on asset i is

$$\mathbb{E}[r_{t+1}^i] - r_{t,t+1}^f = \pi_{i,V}V_t + \pi_{i,W}W_t,$$

where the coefficients are given by

$$\pi_{i,V} = k_{i,1}A_{i,x}\Lambda_x\phi_x^2 + k_{i,1}A_{i,V}\Lambda_V\phi_V^2 - \frac{1}{2}((k_{i,1}A_{i,x}\phi_x)^2 + (k_{i,1}A_{i,V}\phi_V)^2 + \phi_i^2)$$
$$+ \iota_x[k_{i,1}A_{i,x}l_x^{(1)}(0) + l_x(-\Lambda_x) - l_x(k_{i,1}A_{i,x} - \Lambda_x)]$$
$$+ \iota_V[k_{i,1}A_{i,V}l_V^{(1)}(0) + l_V(-\Lambda_V) - l_V(k_{i,1}A_{i,V} - \Lambda_V)],$$
$$\pi_{i,W} = k_{i,1}A_{i,W}\Lambda_W\phi_W^2 - \frac{1}{2}(k_{i,1}A_{i,W}\phi_W)^2.$$

The conditional return variance of the market portfolio is $Var_t(r_{t,t+1}^m) = \sigma_{r,V}V_t + \sigma_{r,W}W_t$, with coefficients

$$\sigma_{r,V} = (k_{m,1}A_{m,x}\phi_x)^2 + (k_{m,1}A_{m,V}\phi_V)^2 + \phi_m^2 + \iota_x(k_{m,1}^2A_{m,x}^2l_x^{(2)}(0))$$
$$+ \iota_V(k_{m,1}^2A_{m,W}^2l_V^{(2)}(0)),$$
$$\sigma_{r,W} = (k_{m,1}A_{m,W}\phi_W)^2.$$

The variance risk premium is given by $\nu_t = B_{m,V} V_t + B_{m,W} W_t$ with coefficients

$$
\begin{aligned}
B_{m,V} &= \iota_x^{\mathcal{Q}}(k_{m,1}^2 A_{m,x}^2 l_x^{\mathcal{Q}(2)}(0)) - \iota_x(k_{m,1}^2 A_{m,x}^2 l_x^{(2)}(0)) \\
&\quad + \iota_V^{\mathcal{Q}}(k_{m,1}^2 A_{m,V}^2 l_V^{\mathcal{Q}(2)}(0)) - \iota_V(k_{m,1}^2 A_{m,V}^2 l_V^{(2)}(0)) \\
&\quad + \Big((\kappa_V - 1) - \Lambda_V + \iota_V^{\mathcal{Q}} l_V^{\mathcal{Q}(1)}(0)\Big)\sigma_{r,V}^{\mathcal{Q}} - \Big((\kappa_V - 1) + \iota_V l_V^{(1)}(0)\Big)\sigma_{r,V}, \\
B_{m,W} &= -\Lambda_W \sigma_{r,W} + (1-\kappa_V)\sigma_{r,V}^{\mathcal{Q}} - (1-\kappa_V)\sigma_{r,V}.
\end{aligned}
$$

Here, $\sigma_{r,V}^{\mathcal{Q}}$ denotes the risk-neutral return variance-coefficient in front of V. The diffusive parts of the variance are of course unaffected by the change of measure. However, the risk-neutral jump intensity $\iota^{\mathcal{Q}}$ is higher than the physical jump intensity and the jump size distribution changes. We denote the moment-generating function under the risk-neutral measure by $l^{\mathcal{Q}}$. The change of the probability measure is described in detail by Eraker and Shaliastovich (2008).

Analogous to the model of Bansal and Yaron (2004), the model-implied slope coefficients in predictive regressions are

$$
b_{\omega,i} = \pi_{i,V} A_{m,V} \frac{Var(V_t)}{Var(\omega_t^j)} \sum_{\tau=0}^{h_\omega - 1} \kappa_V^\tau + \pi_{i,W} A_{m,W} \frac{Var(W_t)}{Var(\omega_t^j)} \sum_{\tau=0}^{h_\omega - 1} \kappa_W^\tau \tag{A15a}
$$

$$
b_{\nu,i} = \pi_{i,V} B_{m,V} \frac{Var(V_t)}{Var(\nu_t^j)} \sum_{\tau=0}^{h_\nu - 1} \kappa_V^\tau + \pi_{i,W} B_{m,W} \frac{Var(W_t)}{Var(\nu_t^j)} \sum_{\tau=0}^{h_\nu - 1} \kappa_W^\tau. \tag{A15b}
$$

Branger et al. (2018) calibrate the model extension reported above. The parameters are close to those in Drechsler and Yaron (2011) and yield similar coefficients in the expressions for the price-dividend ratio, the variance risk premium, and the conditional equity premium. Using these parameters, we solve the model according to the procedure discussed above and come up with the numbers in Equation (15). Given these parameters, $A_{m,V}$ and $A_{m,W}$ are both negative. Economically, an increase in consumption volatility V and in the central tendency W of the volatility process are bad news for the risk-averse investor with a preference for early resolution of uncertainty. Thus, the market prices of V- and W-risk are negative, i.e., the wealth consumption ratio and, in line with that, the price-dividend ratio decrease after a positive innovation in V or W. At the same time, there is an increase in risk premia and, thus, expected excess returns on single stocks. This increase is particularly large for relatively risky stocks with high unconditional expected excess returns. As a consequence, stocks with more negative coefficients in predictive regressions on the market price-dividend ratio have higher unconditional returns with this calibration.

A4 The model of Segal et al. (2015)

A4.1 Model setup and solution

The authors assume that the agent has recursive preferences, similar to those in Bansal and Yaron (2004, see Section A2 of this Online Appendix). In terms of cash flow dynamics, the model explicitly distinguishes between positive and negative innovations in consumption and trend consumption growth. Segal et al. (2015) model *good* and *bad uncertainty* as two separate state variables. In the original version the authors assume homoskedastic processes for the two variance components. We consider the following variant of the model with square-root processes for V and W to obtain a time-varying variance risk premium:

$$\Delta c_{t+1} = \mu_c + x_t + \phi_c(\varepsilon^g_{t+1} - \varepsilon^b_{t+1}) \tag{A16a}$$

$$\Delta d^i_{t+1} = \mu_i + \kappa_i x_t + \phi_i \sqrt{\delta_i V_t + (1-\delta_i) W_t}\varepsilon^i_{t+1} \tag{A16b}$$

$$x_{t+1} = \kappa_x x_t + \tau_g V_t - \tau_b W_t + \phi_x(\varepsilon^g_{t+1} - \varepsilon^b_{t+1}) \tag{A16c}$$

$$V_{t+1} = \bar{V} + \kappa_V (V_t - \bar{V}) + \phi_V \sqrt{V_t}\varepsilon^V_{t+1} \tag{A16d}$$

$$W_{t+1} = \bar{W} + \kappa_W (W_t - \bar{W}) + \phi_W \sqrt{W_t}\varepsilon^W_{t+1}, \tag{A16e}$$

where ε^i_{t+1}, ε^V_{t+1}, and ε^W_{t+1} are i.i.d. $\mathcal{N}(0,1)$. ε^g_{t+1} and ε^b_{t+1} are assumed to be independent centered Gamma distributed, i.e. $\varepsilon^g_{t+1} + V_t \sim \Gamma(V_t, 1)$ and $\varepsilon^b_{t+1} + W_t \sim \Gamma(W_t, 1)$. Since the (centered) Gamma-distribution is right-skewed, V_t and W_t characterize the time-varying upside and downside potential in consumption growth.

A further important ingredient of the model is the channel by which V and W impact the trend growth rate x. In a thorough empirical analysis, Segal et al. (2015) use the predictable components of positive and negative realized semivariances of industrial production growth rates to measure V and W in the data. They show that good (bad) uncertainty positively (negatively) predicts future growth in consumption, output, investment, R&D, earnings, and dividends over horizons of 1 to 5 years. These findings strongly suggest positive values for the parameters τ_g and τ_b in Equation (A16c).

We basically apply the same solution technique as the one discussed in Section A2 of this Online Appendix, with the additional feature that the log moment generating function of the Gamma innovations is

$$\log(l_i(u)) = \log(\mathbb{E}_t[e^{u\varepsilon^i_{t+1}}]) = -(\log(1-u)+u) \cdot \begin{cases} V_t \text{ if } i = g \\ W_t \text{ if } i = b \end{cases}$$

For $i = V$ and $i = W$, the shocks ε^i are Gaussian so $\log(l_i(u)) = \frac{1}{2}u^2$ as in the model of Bansal and Yaron (2004). The log wealth consumption ratio can then be solved using the same approach as in Appendix A2. It is given by $\omega^c_t = A_{c,0} + A_{c,x}x_t + A_{c,V}V_t + A_{c,W}W_t$ with coefficients

$$A_{c,0} = \frac{1}{1-k_{c,1}}\left(-\delta + (1-\rho)\mu_c + k_{c,0} + k_{c,1}A_{c,V}(1-\kappa_V)\bar{V} + k_{c,1}A_{c,W}(1-\kappa_W)\bar{W}\right)$$
$$A_{c,x} = \frac{1-1/\psi}{1-k_{1,c}\kappa_x}$$

and $A_{c,V}$ and $A_{c,W}$ solve

$$
\begin{aligned}
0 &= \frac{1}{2}(\theta k_{c,1}\phi_V)^2 A_{c,V}^2 + \theta(k_{1,c}\kappa_V - 1)A_{c,V} \\
&\quad + \theta k_{c,1}A_{c,x}\tau_g - \Big(\log(1-((1-\gamma)\phi_c + \theta k_{c,1}A_{c,x}\phi_x)) + ((1-\gamma)\phi_c + \theta k_{c,1}A_{c,x}\phi_x)\Big) \\
0 &= \frac{1}{2}(\theta k_{c,1}\phi_W)^2 A_{c,W}^2 + \theta(k_{1,c}\kappa_W - 1)A_{c,W} \\
&\quad - \theta k_{c,1}A_{c,x}\tau_b - \Big(\log(1+((1-\gamma)\phi_c + \theta k_{c,1}A_{c,x}\phi_x)) - ((1-\gamma)\phi_c + \theta k_{c,1}A_{c,x}\phi_x)\Big).
\end{aligned}
$$

The log pricing kernel is given in Equation (A3). Substituting the log-linearized return on the wealth portfolio and using the affine guess for the wealth-consumption ratio yields

$$
m_{t,t+1} = m_0 + m_x x_t + m_V V_t + m_W W_t - \Lambda_x(\varepsilon_{t+1}^g - \varepsilon_{t+1}^b) - \Lambda_V \varepsilon_{t+1}^V - \Lambda_W \varepsilon_{t+1}^W
$$

with coefficients

$$
\begin{aligned}
m_0 &= -\delta\theta - \gamma\mu_c + (\theta-1)k_{c,0} + (\theta-1)(k_{c,1}-1)A_{c,0} \\
&\quad + (\theta-1)k_{c,1}A_{c,V}(1-\kappa_V)\bar{V} + (\theta-1)k_{c,1}A_{c,W}(1-\kappa_W)\bar{W} \\
m_x &= -\gamma + (\theta-1)(k_{c,1}\kappa_x - 1)A_{c,x} = -\rho \\
m_V &= (\theta-1)\left((k_{c,1}\kappa_V - 1)A_{c,V} + k_{c,1}A_{c,x}\tau_g\right) \\
m_W &= (\theta-1)\left((k_{c,1}\kappa_W - 1)A_{c,W} - k_{c,1}A_{c,x}\tau_b\right) \\
\Lambda_x &= \gamma\phi_c + (1-\theta)k_{c,1}A_{c,x}\phi_x \\
\Lambda_V &= (1-\theta)k_{c,1}A_{c,V}\phi_V\sqrt{V_t} \\
\Lambda_W &= (1-\theta)k_{c,1}A_{c,W}\phi_W\sqrt{W_t}.
\end{aligned}
$$

The dynamics of dividends of asset i are given in Equation (A16b). The coefficients of the price-dividend ratio can be solved as explained in Appendix A2 and are given by

$$A_{i,0} = \frac{1}{1-k_{i,1}}\left(m_0 + k_{i,1}(1-\kappa_V)A_{i,V}\bar{V} + k_{i,1}(1-\kappa_W)A_{i,W}\bar{W} + \mu_d + k_{i,0}\right)$$
$$A_{i,x} = \frac{\kappa_i - 1/\psi}{1-k_{i,1}\kappa_x},$$

while $A_{i,V}$ and $A_{i,W}$ solve

$$\begin{aligned}
0 &= \frac{1}{2}(k_{i,1}\phi_V)^2 A_{i,V}^2 + \left(k_{i,1}\kappa_V - 1 - k_{i,1}\phi_V\frac{\Lambda_V}{\sqrt{V_t}}\right)A_{i,V} + m_V + k_{i,1}A_{i,x}\tau_g \\
&\quad + \frac{1}{2}\frac{\Lambda_V^2}{V_t} + \frac{1}{2}\phi_i^2\delta_i - \left(\log(1 - k_{i,1}A_{i,x}\phi_x + \Lambda_x) + (k_{i,1}A_{i,x}\phi_x - \Lambda_x)\right) \\
0 &= \frac{1}{2}(k_{i,1}\phi_W)^2 A_{i,W}^2 + \left(k_{i,1}\kappa_W - 1 - k_{i,1}\phi_W\frac{\Lambda_W}{\sqrt{W_t}}\right)A_{i,W} + m_W - k_{i,1}A_{i,x}\tau_g \\
&\quad + \frac{1}{2}\frac{\Lambda_W^2}{W_t} + \frac{1}{2}\phi_i^2(1-\delta_i) - \left(\log(1 + k_{i,1}A_{i,x}\phi_x - \Lambda_x) - (k_{i,1}A_{i,x}\phi_x - \Lambda_x)\right).
\end{aligned}$$

The expected excess return on the market portfolio is

$$\mathbb{E}_t\left[r_{t,t+1}^m\right] - r_{t,t+1}^f = \pi_V V_t + \pi_W W_t,$$

where

$$\begin{aligned}
\pi_{m,V} &= (k_{m,1}\kappa_V - 1)A_{m,V} + m_V + k_{m,1}A_{m,x}\tau_g \\
&\qquad - (\log(1+\Lambda_x) - \Lambda_x) + \frac{1}{2}((1-\theta)k_{c,1}A_{c,V}\phi_V)^2 \\
\pi_{m,W} &= (k_{m,1}\kappa_W - 1)A_{m,W} + m_W - k_{m,1}A_{m,x}\tau_b \\
&\qquad - (\log(1-\Lambda_x) + \Lambda_x) + \frac{1}{2}((1-\theta)k_{c,1}A_{c,W}\phi_W)^2.
\end{aligned}$$

The conditional return variance is $Var_t(r^m_{t,t+1}) = \sigma_{r,V} V_t + \sigma_{r,W} W_t$, where

$$\sigma_{r,V} = \phi_m^2 \delta_m + (k_{m,1} A_{m,x} \phi_x)^2 + (k_{m,1} A_{m,V} \phi_V)^2$$
$$\sigma_{r,W} = \phi_m^2 (1 - \delta_m) + (k_{m,1} A_{m,x} \phi_x)^2 + (k_{m,1} A_{m,W} \phi_W)^2.$$

The variance risk premium can be calculated similarly to the procedure outlined in Section A2 of this Online Appendix. It is given by $\nu_t = B_{m,V} V_t + B_{m,W} W_t$, where

$$B_{m,V} = k_{c,1} \sigma_{r,V} (\theta - 1) A_{c,V} \phi_V^2$$
$$B_{m,W} = k_{c,1} \sigma_{r,W} (\theta - 1) A_{c,W} \phi_W^2.$$

Finally, the slope coefficients from predictive regressions can be calculated using Equations (A15).

A4.2 Quantitative analysis

As a benchmark case for the calibration, we set $\tau_g = \tau_b = 0$, i.e., the trend growth rate is not affected by either variance component. The full set of parameters is given in Table A5. Similarly to Equations (15) in the paper for the model of Drechsler and Yaron (2011), we find that the relation between $b_{\omega,i}$ and expected returns is negative:

$$b_{\omega,i} = -0.37\,\pi_{i,V} - 0.63\,\pi_{i,W}$$
$$b_{\nu,i} = 0.42\,\pi_{i,V} + 0.58\,\pi_{i,W}.$$

The intuition is similar to the Drechsler and Yaron (2011) model. The wealth-consumption ratio decreases in V and W with these parameters, so that the term "good uncertainty" for V does not really appear justified. Accordingly, the coefficients $A_{m,V}$ and $A_{m,W}$ for the price-dividend ratio are negative as well. At the same time, the corresponding coefficients $B_{m,V}$ and $B_{m,W}$ for the variance risk premium are both positive. Simulating returns given these parameters and sorting stocks with respect to b_ω and b_ν implies that returns are decreasing in b_ω, as can be seen in the upper panel of Table A8.

As suggested by Segal et al. (2015), we now assume that τ_g and τ_b are both positive. This choice mitigates (amplifies) the negative effect of increases in V (W) on the wealth-consumption ratio and also on the price-dividend ratio. We choose τ_g large enough to make both valuation ratios appreciate upon increases in V. We furthermore increase the persistence of V, while decreasing that of W. This makes V more and W less important for variations in the price-dividend ratio. At the same time, we increase the volatility of W and decrease that of V, which makes W the dominant driver of the variance risk premium. In this way, the predictive power of the price-dividend ratio is driven by the common variation of the price-dividend ratio and expected excess returns through V and, at the same time, the predictive power of the variance risk premium is driven by the variation in W.[2]

Calculating the slope coefficients given these parameters yields

$$b_{\omega,i} = \quad 0.83\,\pi_{i,V} \; - \; 0.17\,\pi_{i,W} \tag{A17a}$$

$$b_{\nu,i} = -\,0.47\,\pi_{i,V} \; + \; 0.53\,\pi_{i,W}. \tag{A17b}$$

As a consequence, stocks with high expected returns have on average high $b_{\omega,i}$ and also high $b_{\nu,i}$. The simulation results in the lower panel of Table A8 confirm that average returns are increasing in both b_ω and b_ν, just as observed in the data.

[2]The properties of the state variables V and W are similar to the estimates of Zhou and Zhu (2014), who find two volatility components, one of which is persistent and stable, while the other one is transitory and volatile. They are also in line with the stylized empirical fact that the price-dividend ratio (which is assumed to be driven by the more persistent V in this model) predicts returns over longer horizons of several years, while the variance risk premium is mainly a short-term predictor for horizons up to four months. Variation in the price-dividend ratio must therefore be driven by the variation in a persistent state variable, while the state variable that causes the major part of the variation in the variance risk premium must be rather transitory.

Figure A4 shows $b_{\omega,m}$ as a function of τ_g, τ_b, and the parameters of the two variance processes. We set $\tau_g = \tau_b \equiv \tau$ and vary this parameter between 0 and 1.25×10^{-4}. The *variance shifting parameter* is defined such that the values of 0 and 1 correspond to "Case 1" and "Case 2" for the parameters κ_V, ϕ_V, κ_W, and ϕ_W in Table A5, respectively. For values of the variance shifting parameter different from 0 or 1, we linearly inter- and extrapolate between the cases.

A positive $b_{\omega,m}$ indicates that the model is able to produce a positive cross-sectional relation between b_ω and expected returns. The system of equations (A17) is helpful in illustrating the rationale. Here, high-$b_{\omega,i}$ stocks feature, on average, high $\pi_{i,V}$ and low $\pi_{i,W}$. In order to match the positive relation between $b_{\omega,i}$ and unconditional expected returns, the effect of $\pi_{i,V}$ on $b_{\omega,i}$ must dominate. Under the assumption that the π's are uncorrelated in the cross-section and fluctuate around the π's of the market portfolio, this is the case if $b_{\omega,m}$ is positive.

We also find that the parameter $b_{\nu,m}$ is positive for all parameter scenarios considered, which implies that the positive relation between b_ν and expected returns holds consistently. Moreover, since the coefficients of $\pi_{i,V}$ always have opposite signs in the equations for $b_{\omega,i}$ and $b_{\nu,i}$, and the same is true for $\pi_{i,W}$, the negative relation between b_ω and b_ν holds for all the parameter values considered here. Thus, the area of positive values for $b_{\omega,m}$ in Figure A4 indeed represents the scenarios in which the model can reproduce the pairwise cross-sectional correlations between $b_{\omega,i}$, $b_{\nu,i}$, and $\mathbb{E}[r^i_{t+h} - r^f_{t+h}]$.

We also perform independent double sorts based on simulated data to study the cross-sectional relation between b_ω and b_ν. The relative number of stocks in the 25 portfolios is reported in Table A9. Similar to the data, we find that the majority of stocks cluster around the counterdiagonal for both parametrizations. For Case 1 where $\tau_g = \tau_b = 0$, the intuition is similar to the long-run risks model. High b_ω-stocks have low expected return coefficients π_V and/or π_W, because A_V and A_W are negative. Those are exactly the stocks which have low b_ν's. For Case 2, the situation is slightly different: Stocks with high $b_{\omega,i}$ either have a high $\pi_{i,V}$ or a low $\pi_{i,W}$. In any case, a stock with high $\pi_{i,V}$ or a low $\pi_{i,W}$ will have a low $b_{i,\nu}$, since $B_{m,V} < 0$ and $B_{m,W} > 0$.

In Section 3.3 we also found clusters near the main diagonal and we also observed a positive rank correlation between b_ω and b_ν in some months. We observe a similar pattern in Table A9. Furthermore, the rank correlation between b_ω and b_ν is positive in 15% of the months. The reason is that our rolling window approach estimates covariances on rather short samples. Despite the fact that innovations in V and W are theoretically uncorrelated, they can appear correlated in small samples. Imagine a perfect correlation, e.g., $V_t = W_t$ for all t in a particular 72-month window. The variance risk premium in t would be $(B_{m,V} + B_{m,W})W_t$, with a negative $B_{m,V}$ and a positive $B_{m,W}$, which is greater than $B_{m,V}$ in absolute terms. A similar argument holds for the price-dividend ratio, where the positive $A_{m,V}$ is greater in magnitude than the negative $A_{m,W}$. In that estimation window, a high $\hat{b}_\omega$-stock would also be a high $\hat{b}_\nu$-stock. In periods with accidentally high correlation between the state variables, we cannot identify exposures to the two state variables separately, because the price-dividend ratio and the variance risk premium move in lockstep and, thus, do not represent independent sources of information.

While the model perfectly explains the pairwise cross-sectional correlations between $b_{\omega,i}$, $b_{\nu,i}$, and $\mathbb{E}[r^i_{t+h} - r^f_{t+h}]$, it implies that the slope coefficient $b_{\omega,m}$ of the market portfolio is positive, since $\pi_{m,V}$ is positive and larger than $\pi_{m,W}$. Moreover, it is reasonable to assume that this is also true for the majority of stocks in the cross-section, since the value-weighted sum of the $\pi_{i,V}$'s must equal $\pi_{m,V}$. This pattern is at odds with the finding from Section 4.1 that the majority of slope coefficients and especially the slope coefficient of the market itself is negative in the large majority of periods. As a summary we can state that the model can explain the main empirical facts in a natural way but fails when it comes to explaining the details. Another state variable could be added to the model to also match the signs of the slope coefficients.

A5 The model of Bekaert and Engstrom (2017)

A5.1 Model setup and solution

Following Campbell and Cochrane (1999), the agent maximizes the intertemporal utility function

$$\mathbb{E}_t\left[\sum_{j=0}^{\infty}\exp(-j\delta)\frac{(C_{t+j}-H_{t+j})^{1-\gamma}-1}{1-\gamma}\right]. \tag{A18}$$

The preference parameters δ and γ control the agent's pure time preference rate and her risk attitude. More specifically, Campbell and Cochrane (1999) show that the Arrow-Pratt measure of relative risk aversion is time-varying and equal to $\gamma\exp(-s_t)$, with s_t being defined below. H_t denotes the time-varying habit or subsistence level, such that $C_t > H_t$. Time variation in H is modeled indirectly via the time variation in the log surplus consumption ratio

$$s_t = \log(C_t - H_t) - \log(C_t), \tag{A19}$$

which is assumed to follow

$$s_{t+1} = \bar{s} + \kappa_s(s_t - \bar{s}) + \lambda_t(\Delta c_{t+1} - \mu_c). \tag{A20}$$

λ_t is always positive such that the surplus consumption ratio increases in consumption growth. In particular,

$$\lambda_t = max\left(0, \frac{\sqrt{1-2(s_t-\bar{s})}-1}{\sqrt{(\phi_{cg}^2 V_t + \phi_{cb}^2 W_t)\frac{\gamma}{1-\kappa_s - b/\gamma}}}\right), \tag{A21}$$

which is a natural generalization of the specification in Campbell and Cochrane (1999), in the sense that the preference parameter b can be different from zero which allows the risk-free interest rate to depend directly on s_t (see Bekaert and Engstrom 2017, for details).

With these preferences, the log pricing kernel is given by

$$m_{t+1} = -\delta - \gamma \Delta c_{t+1} - \gamma (s_{t+1} - s_t). \tag{A22}$$

Consumption growth is not i.i.d. as in Campbell and Cochrane (1999), but driven by *good* and *bad* shocks. The dynamics are given by

$$\Delta c_{t+1} = \mu_c + \phi_{cg} \varepsilon^g_{t+1} - \phi_{cb} \varepsilon^b_{t+1} \tag{A23a}$$

$$V_{t+1} = \bar{V} + \kappa_V (V_t - \bar{V}) + \phi_V \varepsilon^g_{t+1} \tag{A23b}$$

$$W_{t+1} = \bar{W} + \kappa_W (W_t - \bar{W}) + \phi_W \varepsilon^b_{t+1} \tag{A23c}$$

$$\Delta d^i_{t+1} = \mu_i + \phi_{ig} \varepsilon^g_{t+1} - \phi_{ib} \varepsilon^b_{t+1}, \tag{A23d}$$

where ε^g_{t+1} and ε^b_{t+1} are independent centered Gamma distributed as described in Section A4 for the case of the Segal et al. (2015) model. In contrast to the latter model, however, there are no separate shocks in the dynamics of V, W, and d^i in Equations (A23a) to (A23d), i.e., ε^g and ε^b are the only shocks in the model. We assume that the dynamics of firm i's dividends are structurally equal to the market claim in Bekaert and Engstrom (2017). The authors discuss in great detail how these dynamics lead to plausible time variation in the volatility, skewness, and kurtosis of consumption growth, while keeping the mean growth rate of consumption constant at μ_c.

Bekaert and Engstrom (2017) estimate the parameters in Equations (A23a) to (A23c) based exclusively on consumption data. The exposures ϕ_{mg} and ϕ_{mb} of the market claim (the case $i = m$) in Equation (A23d) are then set to match dividend growth volatility and the correlation between consumption and dividend growth. There are two pairs of parameters which achieve this goal. Bekaert and Engstrom (2017) choose $\phi_{mg} = -0.0055$ and $\phi_{mb} = 0.0217$, which imply that the aggregate dividend claim hedges consumption shocks through ε^g. The alternative pair of parameters with $\phi_{mg} = 0.0081$ and $\phi_{mb} = -0.0138$ is discussed below. Moreover, Bekaert and Engstrom (2017) consider a parsimonious special case of the model with $\kappa_V = \phi_V = 0$, implying that V_t is constant. For the positive relation between $b_{\omega,i}$ and expected returns, however, a time-varying V turns out to be crucial, so we consider the general model.

We choose model parameters that are equal to those in Bekaert and Engstrom (2017) with a few exceptions which are discussed now. Since we assume that V is not constant, we have to choose the persistence κ_V of V and the multiplier ϕ_V of the shock ε_g. We choose these parameters such that V has properties that are similar to W. More precisely, we choose $\kappa_V = \kappa_W = 0.91$ and ϕ_V such that the variance of V is equal to the variance of W.

To solve the model, we first calculate the price-dividend ratio of asset i on a grid for the state variables. To define reasonable edges of the grid, we simulate a sample of 100,000 periods and use the most extreme values. We use a grid with 36 grid points per dimension. We start with a (constant) guess of the price-dividend ratio for each grid point and iterate backward using the Euler Equation in the form

$$\exp(w_t^i) = \mathbb{E}_t\left[\exp\left(m_{t+1} + \log(\exp(w_{t+1}^i) + 1) + \Delta d_{t+1}^i\right)\right].$$

We calculate transition probabilities between grid points and calculate the expectation numerically. Once we have the price-dividend ratio on the grid, we use quadrature methods to calculate expectations under $\mathbb{P}$ and $\mathbb{Q}$ on the grid. Here, we interpolate between grid points and use a subgrid to avoid running out of the grid for the price-dividend ratio. In this way we calculate conditional expected excess returns and the variance risk premium.

A5.2 Quantitative analysis

Before we start the analysis of the implications for the cross-section of expected returns, we want to shed light on a few special aspects of the Bekaert and Engstrom (2017) model. First, both high-V states and high-W states are bad states of the world, i.e., the wealth-consumption ratio decreases in both V and W. This seems surprising in the case of V, since it represents the time-varying upside potential in consumption growth and could thus be labeled "good uncertainty". However, this intuition is misleading. The variance of the ε^g-shocks is equal to V_t and thus obviously increasing in V. Although the distribution of consumption growth is more right-skewed at high levels of V, the negative influence of increased volatility prevails.[3] This is in line with our first calibration of the Segal et al. (2015) model, where $\tau_g = \tau_b = 0$.

Second, ε^g-shocks have ambivalent consequences for the investor. She appreciates the associated increase in consumption and surplus, but she does not appreciate the increase in V (the wealth-consumption ratio goes down). With habit preferences, the wealth-consumption ratio does not impact the pricing kernel (as it would with Epstein and Zin (1989) preferences), and so the market price of ε^g-risk is undoubtedly positive. This leads to the rather uncommon feature that a positive shock to a source of randomness with a positive market price of risk leads to a decline in the wealth-consumption ratio. In simple words, the investor likes *going* to the high-V state, but she does not like *being* in the high V-state.

Third, the properties of the dividend claim are similar to those of the consumption claim,

[3]For this feature, the habit preferences are crucial. With time-additive preferences, the wealth-consumption ratio increases in V and W. As discussed by, e.g., Bansal and Yaron (2004), it is important to distinguish between risk aversion and the (inverse of the) elasticity of intertemporal substitution when it comes to a model's ability to explain asset prices and their dynamics.

if both ϕ_{mg} and ϕ_{mb} are positive. In this case, the price-dividend ratio decreases and risk premia increase in V and W, similar to the model of Drechsler and Yaron (2011). By setting ϕ_{mg} (or ϕ_{mb}) to a negative value, dividends hedge consumption-shocks through ε^g (ε^b), which is particularly valuable in high-V states (high-W states). Thus, the price-dividend ratio increases and risk premia decrease in V (W), exactly if ϕ_{mg} (ϕ_{mb}) is negative. In the original calibration by Bekaert and Engstrom (2017), ϕ_{mg} is negative and ϕ_{mb} is positive. The alternative calibration where ϕ_{mg} is positive and ϕ_{mb} is negative has very similar implications, with V and W switching roles in all of the following arguments.

Figure A5 shows conditional expected returns on the market and four alternative dividend claims, the market price-dividend ratio, and the variance risk premium, all as functions of the state variables V, W, and s. When we show the impact of one state variable, the other two are set to their steady states values.

We start with an analysis of the impact of V. The aggregate price-dividend ratio is increasing in V, which is due to the fact that ϕ_{mg} is negative. This in turn implies that dividends are a hedge against consumption shocks through ε^g, and this hedge is particularly valuable when V is high. The conditional expected return on the market portfolio is decreasing in V. All else equal, this pattern leads to a negative relation between the price-dividend ratio and expected returns. This negative relation is also found for the other two state variables. The price-dividend ratio is decreasing (increasing) and expected returns are increasing (largely decreasing) in W (s). As a consequence, the slope coefficient $b_{\omega,m}$ is unambiguously negative.

To generate a cross-section of dividend claims, we vary the parameters ϕ_{ig} and ϕ_{ib}. For the market portfolio we set $\phi_{mg} = -0.0055$ and $\phi_{mb} = 0.0217$. To specifically consider the effect of V on the relation between slope coefficients and expected returns, we first consider two stocks i and j with $\phi_{ig} = -0.0075$ and $\phi_{jg} = -0.0035$ and set ϕ_{ib} and ϕ_{jb} equal to the exposure ϕ_{mb} of the market. With these parameters, stock i is an even better hedge for ε^g-risk than the market.

Because of that, the unconditional expected return on that stock is lower than that on the market portfolio. Moreover, its conditional expected return is more sensitive to innovations in V, i.e., an increase in V lets the expected return on that asset go down by even more than that on the market portfolio. In contrast to that, the expected return on stock j is less sensitive to changes in V. Thus, the slope coefficient $b_{\omega,j}$ will be less negative than $b_{\omega,m}$. In line with the data, stock j will also have a relatively high unconditional expected return.

The variance risk premium of the market is decreasing in V and increasing in W, which implies a positive $b_{\nu,m}$ and a negative cross-sectional correlation between b_ν and b_ω, just as documented for our sample in Section 3 of the paper. However, we find $b_{\nu,i} > b_{\nu,j}$, which is at odds with the positive cross-sectional relation between b_ν and expected returns that we observe in the data. Borrowing the intuition from Section 5.2, $\pi_{i,V}$ is smaller than $\pi_{j,V}$, while $\pi_{i,W}$ and $\pi_{j,W}$ are the same. As argued in Section 5.2, we have to consider the exposures to a second state variable (i.e., allow cross-sectional variation in π_W) to explain the cross-sectional relations between b_ω, b_ν, and expected returns.

To study the effect of W, we consider two stocks k and ℓ with $\phi_{kb} = 0.0117$ and $\phi_{\ell b} = 0.0317$, while setting $\phi_{kg} = \phi_{\ell g} = \phi_{mg}$. In this scenario, the impact of V on expected returns is very similar for the two stocks and the market portfolio (the respective curves differ only in level, with stock ℓ having higher expected returns, but not in slope).

When we look at the impact of W, on the other hand, we see that the expected return on stock ℓ increases more strongly than that on stock k. Since the aggregate price-dividend ratio decreases and the variance risk premium increases in W, $b_{\omega,k}$ is larger than $b_{\omega,\ell}$, while $b_{\nu,k}$ is smaller than $b_{\nu,\ell}$. Since the exposure of ℓ to W is stronger than that of k, the former has a higher unconditional expected excess return. Thus, when stocks in the cross-section differ only with respect to ϕ_b, the model can only explain the positive relation between b_ν and expected returns, but fails to provide a rationale for the positive relation between b_ω and expected returns.

As argued in Section 5.2, it is crucial to combine the two channels, i.e., allow cross-sectional variation in exposures to good and bad shocks. The negative correlation between b_ω and b_ν is implied by innovations in V and W alike and, thus, always holds. Whether the model produces positive correlations between b_ω and expected returns (driven by V) and, at the same time, between b_ν and expected returns (driven by W), depends on the cross-sectional distribution of the parameters ϕ_{ig} and ϕ_{ib} and the exact properties of the state variables.

Importantly, the key feature that allows the model to explain the empirical patterns in expected returns and slope coefficients is the negative coefficient ϕ_{mg} (or, alternatively, a negative value for ϕ_{mb}). It implies that there are shocks in the economy which lead to increases in aggregate consumption and decreases in aggregate dividends and vice versa. There is no obvious micro foundation for this channel. As a result, our empirical findings remain difficult to reproduce in an endowment economy, unless one is willing to rely on somewhat counterintuitive assumptions about the cash-flow dynamics. A production economy may be necessary to provide a stronger economic intuition.

A6 The model of Croce (2014)

The agent is assumed to have recursive preferences, similar to those used in the model of Bansal and Yaron (2004, see Section A2 of this Online Appendix). The difference is that time-t utility is not only a function of consumption but also of leisure, aggregated using a standard CES utility function. More precisely,

$$U_t = \left[(1-e^{-\delta})\tilde{C}_t^{1-1/\psi} + e^{-\delta}\left(\mathbb{E}_t\left[U_{t+1}^{1-\gamma}\right]\right)^{1/\theta}\right]^{1/(1-1/\psi)}, \text{ where} \tag{A24a}$$

$$\tilde{C}_t = \left[o\,C_t^{1-1/\eta} + (1-o)(A_{t-1}l_t)^{1-1/\eta}\right]^{1/(1-1/\eta)} \tag{A24b}$$

Here, A denotes aggregate productivity, l denotes leisure, η is the elasticity of substitution between consumption and leisure, and o is a weighting parameter. The other parameters can be interpreted as in Appendix A2.

The aggregate capital stock is given by K_t. Output Y_t is produced from capital using a constant returns-to-scale production function of the form

$$Y_t = K_t^{\alpha} \left[A_t n_t\right]^{1-\alpha},$$

where n_t denotes hours worked. Apart from hours worked, the way how capital is utilized to produce output depends on aggregate productivity A_t which is subject to stochastic shocks which are specified exogenously. The log growth rate of aggregate productivity follows

$$\log(A_{t+1}/A_t) = \mu + x_t + \phi_a e^{V_t} \varepsilon^a_{t+1} \tag{A25a}$$

$$x_{t+1} = \kappa_x x_t + \phi_x e^{V_t} \varepsilon^x_{t+1} \tag{A25b}$$

$$V_{t+1} = \kappa_V V_t + \phi_V \varepsilon^V_{t+1} \tag{A25c}$$

Croce allows ε^x_{t+1} and ε^V_{t+1} to be correlated while the correlation of the two shocks with ε^a_{t+1} is assumed to be equal to zero.

In this setup, the agent optimizes the intertemporal utility function in Equation (A24a) by choosing optimal leisure l, labor input n, consumption C and investment I subject to the constraints

$$Y_t \geq C_t + I_t, \tag{A26a}$$

$$1 \geq n_t + l_t, \tag{A26b}$$

$$K_{t+1} \leq (1 - \delta_K) K_t + I_t - G_t K_t. \tag{A26c}$$

The parameter δ_K in (A26c) denotes the depreciation rate and G_t denotes capital adjustment costs which are specified as in Jermann (1998):

$$G_t = \frac{I_t}{K_t} - \left[\frac{\alpha_1}{1 - \frac{1}{\xi}} \left(\frac{I_t}{K_t} \right)^{1-1/\xi} + \alpha_0 \right]$$

Here, the parameter ξ controls how costly it is to adjust capital, and α_0 and α_1 are chosen to make the capital adjustment cost and its first derivative equal to zero in the steady state.

The model is calibrated using parameters that are standard in the long-run risks literature, in particular, a risk aversion coefficient of 10 and an EIS of 2. The full set of parameters can be found in Table 3 in Croce (2014). The model is solved in dynare++ and the code is provided on Max Croce's website at `https://sites.google.com/view/mmcroce/chronological-order`.

References

Bali, T. and H. Zhou (2016): "Risk, Uncertainty, and Expected Returns," *Journal of Financial and Quantitative Analysis*, 51, 707–735.

Bansal, R., R. Dittmar, and C. Lundblad (2005): "Consumption, Dividends, and the Cross-Section of Equity Returns," *Journal of Finance*, 60, 1639–1672.

Bansal, R. and A. Yaron (2004): "Risks for the Long-Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59, 1481–1509.

Barro, R. (2006): "Rare Disasters and Asset Markets in the Twentieth Century," *Quarterly Journal of Economics*, 2006, 823–866.

Bekaert, G. and E. Engstrom (2017): "Asset Return Dynamics Under Habits and Bad Environment–Good Environment Fundamentals," *Journal of Political Economy*, 125, 713–760.

Bekaert, G. and M. Hoerova (2014): "The VIX, the Variance Premium and Stock Market Volatility," *Journal of Econometrics*, 183, 181–192.

Benzoni, L., P. Collin-Dufresne, and R. Goldstein (2011): "Explaining Asset Pricing Puzzles Associated with the 1987 Market Crash," *Journal of Financial Economics*, 101, 552–573.

Bollerslev, T., G. Tauchen, and H. Zhou (2009): "Expected Stock Returns and Variance Risk Premia," *Review of Financial Studies*, 22, 4463–4492.

Bollerslev, T., J. Xu, and H. Zhou (2014): "Stock Return Predictability and Variance Risk Premia: Statistical Inference and International Evidence," *Journal of Financial and Quantitative Analysis*, 49, 633–661.

Branger, N., P. Rodrigues, and C. Schlag (2018): "The Role of Volatility Shocks and Rare Events in Long-Run Risk Models," *Journal of Economic Dynamics and Control*, 68, 95–122.

Branger, N. and C. Völkert (2011): "What is the Equilibrium Price of Variance Risk? A Long-Run Risk Model with Two Volatility Factors," Working Paper.

Campbell, J. and J. Cochrane (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107, 205–251.

Campbell, J. and R. Shiller (1988): "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1, 195–228.

Croce, M. (2014): "Long-run Productivity Risk: A New Hope for Production-based Asset Pricing?" *Journal of Monetary Economics*, 66, 13–31.

Drechsler, I. and A. Yaron (2011): "What's Vol Got to Do With It," *Review of Financial Studies*, 24, 1–45.

Epstein, L. and S. Zin (1989): "Substitution, Risk Aversion, and the Temporal Behaviour of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57, 937–969.

Eraker, B. and I. Shaliastovich (2008): "An Equilibrium Guide to Designing Affine Pricing Models," *Mathematical Finance*, 18, 519–543.

Fama, E. and K. French (1992): "The Cross-Section of Expected Stock Returns," *Journal of Finance*, 47, 427–465.

Jermann, U. (1998): "Asset Pricing in Production Economies," *Journal of Monetary Economics*, 41, 257–275.

Jin, J. (2015): "Jump-Diffusion Long-Run Risks Models, Variance Risk Premium, and Volatility Dynamics," *Review of Finance*, 19, 1223–1279.

Kraft, H., T. Seiferling, and F. Seifried (2015): "Asset Pricing and Consumption-Portfolio Choice with Recursive Utility and Unspanned Risk," Working Paper.

Kreps, D. and E. Porteus (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46, 185–200.

Londono, J. (2015): "The Variance Risk Premium around the World," Working Paper.

Newey, W. and K. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Pohl, W., K. Schmedders, and O. Wilms (2018): "Higher-Order Effects in Asset-Pricing Models with Long-Run Risks," *Journal of Finance*, 73, 1061 – 1111.

Rietz, T. (1988): "The Equity Risk Premium a Solution," *Journal of Monetary Economics*, 22, 117–131.

Segal, G., I. Shaliastovich, and A. Yaron (2015): "Good and Bad Uncertainty: Macroeconomic and Financial Market Implications," *Journal of Financial Economics*, 117, 369–397.

Tauchen, G. (2011): "Stochastic Volatility in General Equilibrium," *Quarterly Journal of Finance*, 1, 707–731.

Wachter, J. (2013): "Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?" *Journal of Finance*, 68, 987–1035.

Weil, P. (1989): "The Equity Premium Puzzle and the Risk Free Rate Puzzle," *Journal of Monetary Economics*, 24, 401–421.

Zhou, G. and Y. Zhu (2014): "Macroeconomic Volatilites and Long-Run Risks of Asset Prices," *Management Science*, 61, 413–430.

Table A1: **Portfolio returns for single sorts on $b\omega$ and b_ν (excluding small firms)**

	low	2	3	4	5	6	7	8	9	high	HML	t-stat
Panel A: Sample 1990 - 2018												
Sorted on b_ω	0.10	0.45	0.43	0.65	0.61	0.60	0.35	0.36	0.51	0.69	0.60	1.98
Sorted on b_ν	0.30	0.40	0.54	0.51	0.57	0.61	0.38	0.35	0.52	0.79	0.49	1.33
Panel B: Sample 1933 - 2018												
Sorted on b_ω	0.69	0.83	0.82	0.89	0.92	0.96	0.85	0.87	0.92	1.02	0.33	2.54
Sorted on b_ν	0.70	0.83	0.84	0.81	0.82	0.88	0.92	0.84	0.94	0.99	0.29	2.16

The table shows average value-weighted returns on portfolios sorted on b_ω and b_ν, respectively. b_ω and b_ν are estimated as described in Section 3.1 of the paper. In each month, firms with a market capitalization below the 20% quantile of the cross-sectional distribution of market capitalizations in that month are excluded. Portfolios are held over one month, returns are expressed in percentage points. We show the returns on the decile portfolios as well as on the HML portfolio long in the highest and short in the lowest decile portfolio. t-statistics are adjusted following Newey and West (1987). The sort on b_ν in Panel B relies on the backwards extrapolation of the time series for the variance risk premium as described in Section 3.2 of the paper.

Table A2: **Portfolio returns for single sorts on $b\omega$ and b_ν (alternative estimation)**

	low	2	3	4	5	6	7	8	9	high	HML	t-stat
Panel A: 1990/01 - 2018/06												
Sorted on b_ω	-0.03	0.43	0.57	0.55	0.62	0.44	0.44	0.45	0.51	0.83	0.86	2.64
Sorted on b_ν	0.19	0.43	0.69	0.61	0.63	0.58	0.43	0.45	0.32	0.68	0.49	1.35
Panel B: 1933/01 - 2018/06												
Sorted on b_ω	0.65	0.81	0.90	0.91	0.96	0.93	0.89	0.92	0.90	0.94	0.29	2.13
Sorted on b_ν	0.61	0.73	0.83	0.91	0.92	0.97	0.95	0.95	0.92	0.90	0.29	1.97

The table shows average value-weighted returns on portfolios sorted on b_ω and b_ν, respectively. b_ω and b_ν are estimated via the regressions (A1) and (A2). Portfolios are held over one month, returns are expressed in percentage points. We show the returns on the decile portfolios as well as on the HML portfolio long in the highest and short in the lowest decile portfolio. t-statistics are adjusted following Newey and West (1987). The sort on b_ν in Panel B relies on the backwards extrapolation of the time series for the variance risk premium as described in Section 3.2 of the paper.

Table A3: **Portfolio returns for single sorts on $b\omega$ and b_ν (different rebalancing intervals)**

	low	2	3	4	5	6	7	8	9	high	HML	t-stat
Panel A: Sorted on b_ω												
quarterly	0.72	0.83	0.84	0.89	0.92	0.92	0.88	0.89	0.96	0.99	0.27	2.01
semiannual	0.73	0.83	0.88	0.88	0.94	0.95	0.88	0.90	0.95	1.02	0.29	2.30
annual	0.79	0.06	1.05	1.05	1.04	1.03	0.98	0.95	1.12	1.11	0.32	2.63
Panel B: Sorted on b_ν												
quarterly	0.71	0.84	0.85	0.83	0.83	0.81	0.95	0.85	0.97	1.03	0.32	2.38
semi-annual	0.77	0.84	0.88	0.84	0.84	0.85	0.94	0.87	0.94	0.98	0.22	1.69
annual	0.93	0.99	1.04	1.02	0.90	0.99	0.99	0.98	1.04	1.15	0.22	1.70

The table shows average value-weighted returns on portfolios sorted on b_ω and b_ν, respectively. b_ω and b_ν are estimated as described in Section 3.1 of the paper. Portfolios are held over three months, six months, or 12 months. Returns are expressed in percentage points per month. We show the returns on the decile portfolios as well as on the HML portfolio long in the highest and short in the lowest decile portfolio. t-statistics are adjusted following Newey and West (1987). The sample period is 1933/01 - 2018/06. The sort on b_ν in Panel B relies on the backwards extrapolation of the time series for the variance risk premium as described in Section 3.2 of the paper.

Table A4: **HML returns using different prediction horizons and formation periods**

Panel A: HML_ω (sample 1933-2018)

Form. period	3	6	9	12	15	18	21	24	27	30	33	36
48	0.25	0.29	0.44	0.39	0.43	0.36	0.32	0.32	0.24	0.18	0.23	0.24
	[1.88]	[2.19]	[3.07]	[2.61]	[2.80]	[2.29]	[1.94]	[1.91]	[1.47]	[1.06]	[1.31]	[1.42]
72	0.23	0.26	0.40	0.31	0.20	0.32	0.25	0.29	0.31	0.26	0.16	0.18
	[1.72]	[2.02]	[2.86]	[2.31]	[1.54]	[2.41]	[1.84]	[2.10]	[2.23]	[1.82]	[1.19]	[1.35]
96	0.20	0.27	0.39	0.37	0.24	0.30	0.21	0.12	0.00	0.03	-0.00	-0.04
	[1.56]	[2.09]	[2.97]	[2.79]	[1.90]	[2.30]	[1.62]	[0.91]	[0.04]	[0.29]	[-0.03]	[-0.37]

Panel B: HML_ν (sample 1933-2018)

Form. period	1	2	3	4	5	6	7	8	9	10	11	12
48	0.07	0.17	0.27	0.20	0.20	0.06	-0.05	0.04	0.03	0.05	0.11	-0.03
	[0.50]	[1.14]	[1.90]	[1.38]	[1.40]	[0.45]	[-0.35]	[0.26]	[0.23]	[0.32]	[0.74]	[-0.19]
72	0.13	0.25	0.28	0.30	0.35	0.16	-0.05	0.10	0.10	0.09	0.13	0.06
	[0.93]	[1.68]	[2.05]	[2.13]	[2.52]	[1.17]	[-0.35]	[0.69]	[0.78]	[0.70]	[0.97]	[0.43]
96	0.17	0.28	0.30	0.36	0.35	0.21	0.03	0.10	-0.01	0.05	0.06	-0.01
	[1.36]	[1.98]	[2.22]	[2.64]	[2.57]	[1.64]	[0.20]	[0.76]	[-0.07]	[0.36]	[0.48]	[-0.05]

The table shows the average returns on HML_ω and HML_ν portfolios using formation periods of 48, 72, and 96 months, and return horizons between 3 and 36 months (1 and 12 months) in Panel A (Panel B). HML returns are average monthly returns on portfolios long in the highest and short in the lowest decile portfolio. t-statistics are adjusted following Newey and West (1987). The sample period is 1933/01 - 2018/06. The sorts on b_ν in Panel B rely on on the backwards extrapolation of the time series for the variance risk premium as described in Section 3.2 of the paper.

Table A5: **Model parameters**

Preferences			Consumption			Dividends			V			W		
Variant of the Bansal and Yaron (2004) model														
γ	ψ	δ	μ_c	κ_x	ϕ_x	μ_m	κ_m	ϕ_m	$\bar{V}$	κ_V	ϕ_V			
10	1.5	.0020	.0015	.979	.044	.0015	3	4.5	$.0078^2$	.987	.0378			
2	.25													
10	.25													
Variant of the Drechsler and Yaron (2011) model														
γ	ψ	δ	μ_c	κ_x	ϕ_x	μ_m	κ_m	ϕ_m		κ_V	ϕ_V	$\bar{W}$	κ_W	ϕ_W
10	2	.0010	.0016	.971	.032	.0016	2.5	5.7		.7985	.35	$.0066^2$	.985	.1
			ι_x	$\mu_{J,x}$	$\sigma^2_{J,x}$				ι_V	$\lambda_{J,V}$				
			.0667	0	.0027				.0667	2.6				
The model of Segal et al. (2015)														
γ	ψ	δ	μ_c	κ_x	ϕ_x	μ_m	κ_m	ϕ_m	$\bar{V}$			$\bar{W}$		
10	1.5	.0020	.0015	.979	$.044\phi_c$	.0015	3	$4\phi_c$	.5			.5		
			ϕ_c	τ_g	τ_b	δ_m				κ_V	ϕ_V		κ_W	ϕ_W
		Case 1:	.0078	0	0	.5				.987	.0378		.987	.0378
		Case 2:	.0078	$.25\phi_x$	$.25\phi_x$	.5				.990	$\frac{2}{3}.0378$		.900	$\frac{3}{2}.0378$
The model of Bekaert and Engstrom (2017)														
γ		δ	μ_c	ϕ_{cg}	ϕ_{cb}	μ_m	ϕ_{mg}	ϕ_{mb}	$\bar{V}$	κ_V	ϕ_V	$\bar{W}$	κ_W	ϕ_W
11.43		.0001	.0014	.00067	.0019	.0014	-.0055	.0217	11.43	.91	.87	1.56	.91	.32
$\bar{s}$	κ_s	b												
-1.5585	.9963	.0099												

The table shows the parameters we choose to calibrate the models. We consider three parameterizations of the Bansal and Yaron (2004) model and two parameterizations of the Segal et al. (2015) model. Here, we show the alternative parameters one below the other. All other parameters remain unchanged.

Table A6: **Returns on sorted portfolios based on simulated data from the Bansal and Yaron (2004) model**

	low	high	HML
Panel A: $\gamma = 10$**,** $\psi = 1.5$			
Sorted on b_ω	8.27	7.03	-1.23
	[25.99]	[19.24]	[-3.86]
Sorted on b_ν	5.23	9.85	4.62
	[18.84]	[29.10]	[34.87]
Panel B: $\gamma = 2$**,** $\psi = 0.25$			
Sorted on b_ω	4.52	4.62	0.09
	[3.90]	[3.98]	[10.14]
Sorted on b_ν	4.65	4.49	-0.16
	[4.00]	[3.88]	[-11.97]
Panel C: $\gamma = 10$**,** $\psi = 0.25$			
Sorted on b_ω	4.45	5.10	0.65
	[4.00]	[4.58]	[9.94]
Sorted on b_ν	4.23	5.34	1.11
	[3.85]	[4.74]	[12.13]

The table shows value-weighted returns on portfolios that are sorted w.r.t. b_ω and b_ν. All returns are annualized and expressed in percentage points. Column 3 shows returns on the portfolios that are long the high b portfolio and short the low b portfolio. Newey and West (1987) t-statistics are in parentheses. The return data, the price-dividend ratio, and the variance risk premium are simulated as explained in Appendix A2.3 after solving the model as detailed in Section A2.1 with the parameters listed in Table A5. In Panel A, we set $\gamma = 10$ and $\psi = 1.5$, in Panel B, we set $\gamma = 2$ and $\psi = 0.25$, and in Panel C, we set $\gamma = 10$ and $\psi = 0.25$.

Table A7: **Number of stocks in independently double-sorted portfolios based on simulated data from the Bansal and Yaron (2004) model**

Panel A: $\gamma = 10$, $\psi = 1.5$					
	low b_ω	2	3	4	high b_ω
low b_ν	2.96	2.31	2.45	3.90	8.39
2	2.25	3.27	4.56	5.99	3.94
3	2.45	4.58	5.00	4.51	2.46
4	3.94	5.89	4.54	3.29	2.34
high b_ν	8.41	3.95	2.45	2.30	2.87
Panel B: $\gamma = 2$, $\psi = 0.25$					
	low b_ω	2	3	4	high b_ω
low b_ν	0.43	0.93	1.96	4.74	11.94
2	0.96	2.45	4.83	7.02	4.73
3	2.01	4.76	6.42	4.84	1.96
4	4.73	7.07	4.78	2.47	0.94
high b_ν	11.86	4.78	1.99	0.94	0.42
Panel C: $\gamma = 10$, $\psi = 0.25$					
	low b_ω	2	3	4	high b_ω
low b_ν	11.95	4.74	1.96	0.93	0.43
2	4.73	7.02	4.83	2.45	0.96
3	1.96	4.83	6.43	4.76	2.01
4	0.94	2.47	4.78	7.07	4.73
high b_ν	0.42	0.94	1.99	4.78	11.87

The table shows the average relative numbers of stocks in portfolios that are independently double-sorted w.r.t. b_ω and b_ν, expressed in percentage points. The holding period is one month. The return data, the price-dividend ratio, and the variance risk premium are simulated as explained in Section A2.3 after solving the model as detailed in Section A2.1 with the parameters listed in Table A5. In Panel A, we set $\gamma = 10$ and $\psi = 1.5$, in Panel B, we set $\gamma = 2$ and $\psi = 0.25$, and in Panel C, we set $\gamma = 10$ and $\psi = 0.25$.

Table A8: **Returns on sorted portfolios based on simulated data from the Segal et al. (2015) model**

	low	high	HML
Panel A: Case 1			
Sorted on b_ω	16.84	10.77	-6.07
	[126.67]	[72.19]	[-31.13]
Sorted on b_ν	8.88	17.97	9.09
	[81.02]	[191.51]	[83.11]
Panel B: Case 2			
Sorted on b_ω	7.36	9.79	2.43
	[21.12]	[38.14]	[4.05]
Sorted on b_ν	7.95	9.57	1.62
	[70.73]	[60.79]	[6.29]

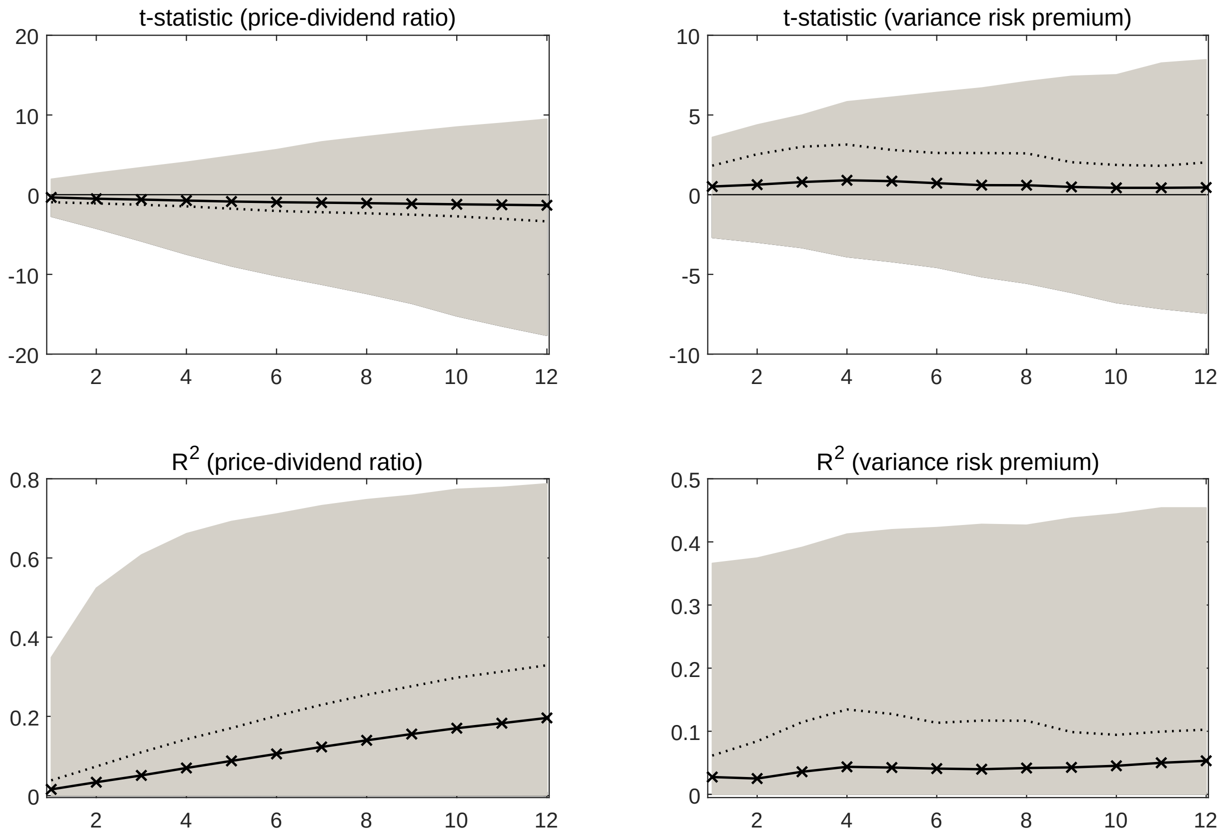
The table shows value-weighted returns on portfolios that are sorted w.r.t. b_ω and b_ν. All returns are annualized and expressed in percentage points. Column 3 shows returns on the portfolios that are long the high b portfolio and short the low b portfolio. Newey and West (1987) t-statistics are in parentheses. The return data, the price-dividend ratio, and the variance risk premium are simulated as explained in Section A2.3 after solving the model as detailed in Section A4.1 with the parameters listed in Table A5. For the returns in Panel A, we use the parameters denoted *Case 1* in Table A5 and for the returns in Panel B, we use the parameters denoted *Case 2* in Table A5.

Table A9: **Number of stocks in independently double-sorted portfolios based on simulated data from the Segal et al. (2015) model**

Panel A: Case 1					
	low b_ω	2	3	4	high b_ω
low b_ν	0.48	1.38	2.74	5.05	10.35
2	1.36	3.11	4.64	5.76	5.13
3	2.76	4.63	5.22	4.69	2.71
4	5.13	5.74	4.67	3.12	1.33
high b_ν	10.27	5.14	2.73	1.39	0.47
Panel B: Case 2					
	low b_ω	2	3	4	high b_ω
low b_ν	1.66	1.32	1.41	2.52	13.09
2	1.28	1.80	3.08	11.26	2.58
3	1.38	3.05	11.13	3.08	1.35
4	2.56	11.31	3.02	1.85	1.30
high b_ν	13.15	2.52	1.37	1.28	1.69

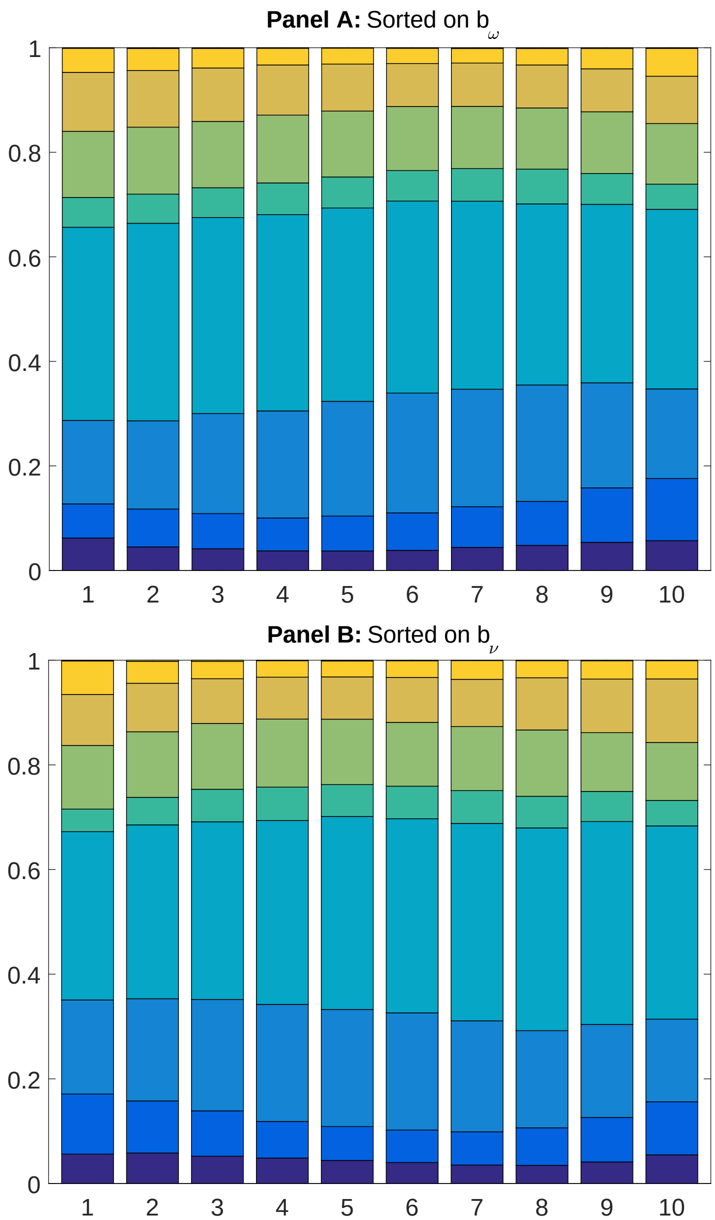
The table shows the average relative numbers of stocks in portfolios that are independently double-sorted w.r.t. b_ω and b_ν, expressed in percentage points. The holding period is one month. The return data, the price-dividend ratio, and the variance risk premium are simulated as explained in Section A2.3 after solving the model as detailed in Section A4.1 with the parameters listed in Table A5. For the returns in Panel A, we use the parameters denoted *Case 1* in Table A5 and for the returns in Panel B, we use the parameters denoted *Case 2* in Table A5.

Figure A1: **Predictive power for different returns horizons**



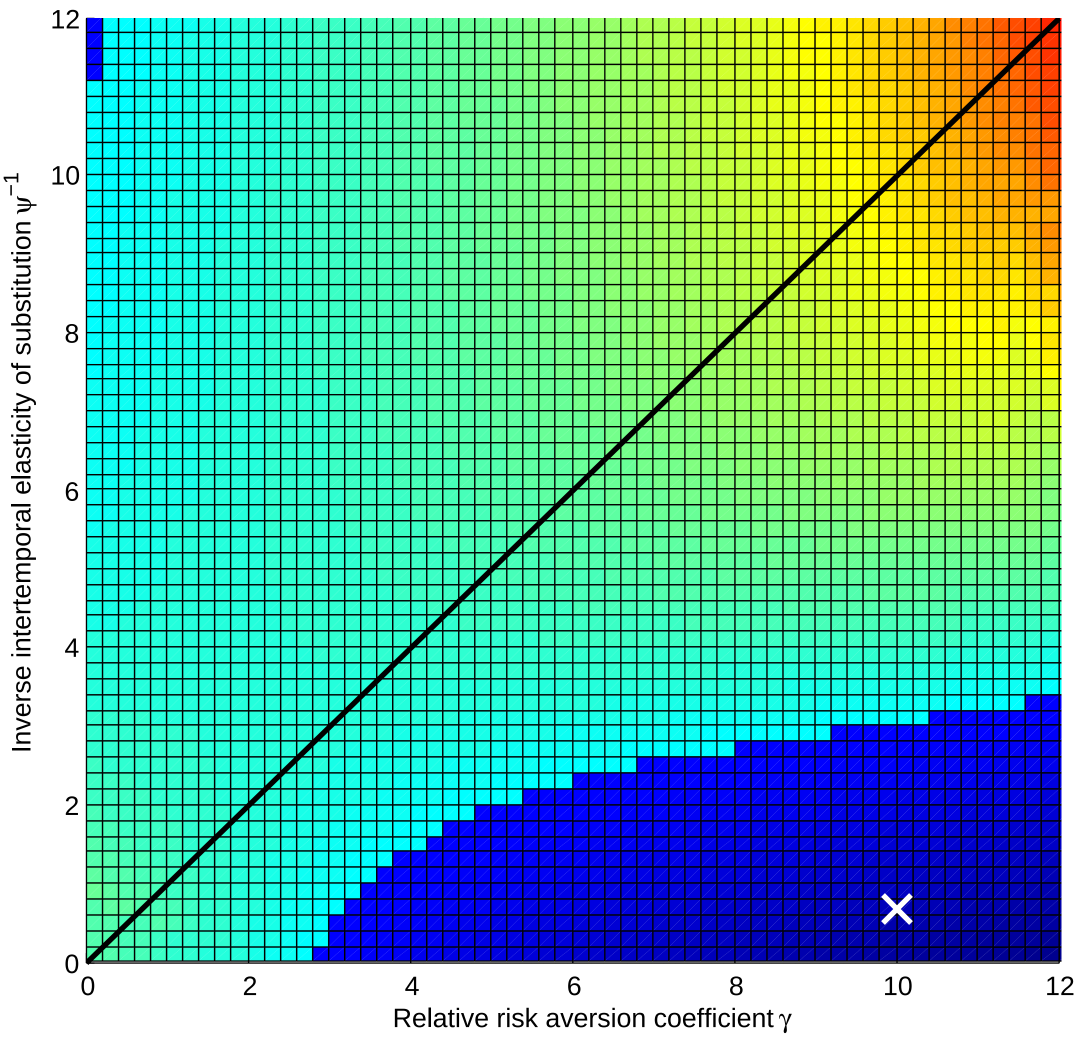
The figure shows t-statistics (upper row of graphs) and R^2's (lower row of graphs) of predictive regressions of excess returns on the price-dividend ratio (left column of graphs) and the variance risk premium (right column of graphs) as a function of the return horizon in months. Asterisks denote the time-series averages of the cross-sectional medians of the distribution of t-statistics and R^2's. The grey areas indicate time-series averages of 95% confidence bands. The dotted lines indicate the time series average of the respective quantity for predictive regressions of excess return on the aggregate stock market.

Figure A2: **Industry composition of sorted portfolios**



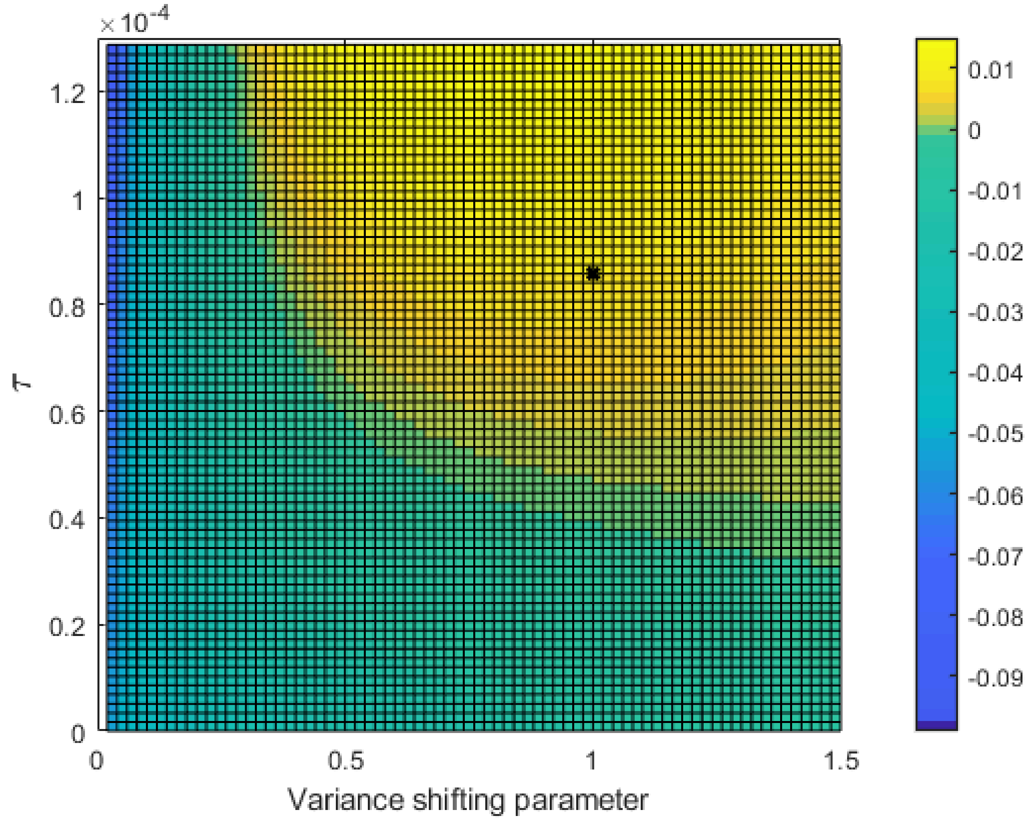
The figure shows the time series averages of the proportions of stocks from different industries in the ten portfolios sorted on b_ω (upper graph) and sorted on b_ν (lower graph). We categorize stocks w.r.t. their month t SIC code as reported in CRSP. The dark blue areas at the bottom of the graphs display stocks with SIC codes between 0 and 999, the dark red area on top (which is barely visible due to the small relative number of stocks in that category) displays stocks with SIC codes between 9000 and 9999. We exclude stocks with SIC codes between 6000 and 6999 (financials) and between 4900 and 4999 (utilities).

Figure A3: **The coefficient $A_{m,V}$ in the model of Bansal and Yaron (2004)**



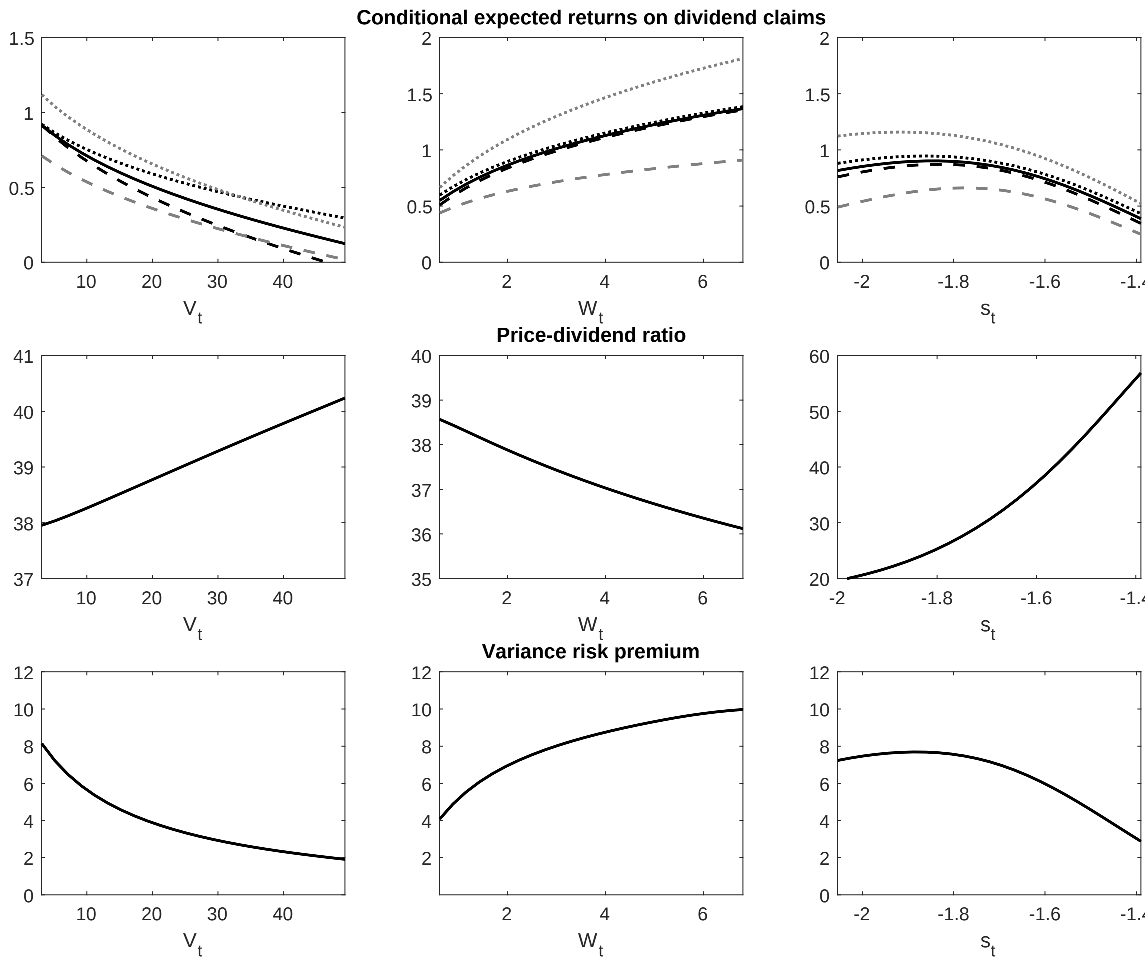
The figure shows the coefficient $B_{m,V}$ for different choices of the preference parameters γ and ψ^{-1}. All other model parameters are chosen as reported in Table A5. The dark blue area indicates the region of the parameter space in which the coefficient $B_{m,V}$ is negative.

Figure A4: **Relation between b_ω and expected returns in the model of Segal et al. (2015)**



The figure shows the coefficient $b_{\omega,m}$ for different choices of the model parameters. We set $\tau_g = \tau_b \equiv \tau$ and vary this parameter between 0 and 1.25×10^{-4} (vertical axis). The variance shifting parameter (horizontal axis) is defined such that values of 0 and 1 correspond to "Case 1" and "Case 2" for the parameters κ_V, ϕ_V, κ_W, and ϕ_W in Table A5, respectively. For values of the variance shifting parameter different from 0 or 1, we linearly inter- and extrapolate between the two cases. The resulting values for $b_{\omega,m}$ are indicated by the colors shown on the scale to the right of the plot. All other model parameters are set as in Table A5.

Figure A5: **Conditional expected excess returns, market price-dividend ratios and variance risk premia in the model of Bekaert and Engstrom (2017)**



The figure shows conditional expected excess returns on different assets (upper row of graphs) as well as the price-dividend ratio and the variance risk premium of the market portfolio (middle and lower row of graphs, respectively), each as implied by the model of Bekaert and Engstrom (2017). In the graphs for expected excess returns, the solid black line denotes the aggregate market portfolio. The black dashed, black dotted, grey dashed, and grey dotted lines represents stocks i, j, k, and ℓ as described in Section A5.2. In the left, middle, and right column of graphs, the quantities are shown as functions of the state variables V, W, and s, respectively. When one state variable is varied, the other two are held fixed at their respective steady state values. Excess returns are expressed in percentage points per month. The price-dividend ratio is annualized, i.e., monthly dividends are multiplied by 12. The variance risk premium is expressed in basis points per month.