The Variance Risk Premium in Equilibrium Models*

Geert Bekaert, Columbia University and the National Bureau of Economic Research,

Eric Engstrom, Board of Governors of the Federal Reserve System

Andrey Ermolov, Gabelli School of Business, Fordham University

May 1, 2020

Abstract

The equity variance risk premium is the expected compensation earned for selling variance risk in equity markets. The variance risk premium is positive and shows moderate persistence. High variance risk premiums coincide with the left tail of the consumption growth distribution shifting down. These facts, together with a positive, yet moderate, difference between the risk-neutral entropy and variance of the aggregate market return, refute the bulk of the extant consumption-based asset pricing models. We introduce a tractable habit model that does fit the data. In the model, the variance risk premium depends positively (negatively) on "bad" ("good") consumption growth uncertainty.

Keywords: variance risk premium, macroeconomic uncertainty, non-Gaussian dynamics

JEL codes: E44, G12, G13

*Contact information: Geert Bekaert - ✉: 3022 Broadway, Uris 411, New York, NY 10027, United States of America, ☎: (+1) 212-854-9156, gb241@gsb.columbia.edu, Eric Engstrom - ✉: Federal Reserve, Washington, DC 20551, United States of America, ☎: (+1) 202-452-3044, eric.c.engstrom@frb.gov, Andrey Ermolov - ✉: 45 Columbus Avenue, room 619, 10023, New York, NY, United States of America, ☎: (+1) 917-969-0060, aermolov1@fordham.edu. The views expressed in this document do not necessarily reflect those of the Board of Governors of the Federal Reserve System, or its staff.

1 Introduction

Consumption-based asset pricing models are typically confronted with a set of salient features regarding the first and second moments of interest rates, equity returns, and in some cases, bond returns including a low risk free rate, a high equity premium and high stock market volatility. At this point, the main paradigms, Campbell and Cochrane (1999)'s habit model, Bansal and Yaron (2004)'s long run risk model and Rietz (1988)'s disaster risk model can all match these stylized facts. A small subset of the vast literature on consumption-based asset pricing has started to explore equity option prices to discipline models and uncover what mechanisms best fit the data. These articles include, inter alia, Du (2011) and Bekaert and Engstrom (2017) for habit models; Bollerslev, Tauchen and Zhou (2009) and Drechsler and Yaron (2011) for long run risk models; Gabaix (2012) and Wachter and Seo (2019) for disaster risk models. Option prices are powerful financial instruments for this purpose, because they reflect at each point in time the conditional expectation of market participants on the equity return distribution, combining their preferences and views on the physical distribution of the underlying returns.

In this article, we explore how these models fit salient facts regarding the variance risk premium and where they fail. The variance risk premium is the expected compensation earned for selling volatility in equity markets. While it can be measured from variance swaps, it can also be measured as the difference between the famous VIX index and an estimate of the conditional variance of equity returns, which is the approach we follow here. Well-known as a "fear index" for asset markets (Whaley, 2000), the VIX uses a weighted average of option prices to approximate the risk neutral variance and is usually higher than the "physical" expected stock market variance. Essentially, this premium reflects out of the money options being more expensive than near the money option prices (Britten-Jones and Neuberger, 2000; Bakshi, Kapadia, and Madan, 2003; Martin 2017). The variance premium also varies considerably through time. In periods of stress, out-of-the money put options that insure against market downturns become relatively more expensive than call options; protection against increases in variance become more expensive as well and the VIX and the variance risk premium both increase.

We confront representative models in the literature with three sets of stylized facts regarding the variance risk premium. The first set comprises time series properties of the variance risk premium. While many models are presumably calibrated to fit these features, the relative fit is still quite different across models. Moreover, the mere fact that the variance risk premium is positive refutes all models that employ only Gaussian shocks to key state variables. Such a result has surfaced in the literature before, but we provide an alternative proof and an alternative result showing that when returns and the pricing kernel are jointly conditionally log-normally distributed, the variance risk premium can only be positive if the equity premium is negative. This result is useful as the standard habit model (Campbell and Cochrane, 1999) and long-run risk model (Bansal and Yaron, 2004, when solved using the typical first order approximation), are conditionally Gaussian.

Our second stylized fact regards the relation between tails in the variance risk premium and consumption growth distributions. For consumption-based asset pricing models to have any chance of explaining options data, there must be correlation between the variance risk premium and consumption growth. However, mimicking the general lack of correlation between asset returns and consumption growth, the correlation between the variance risk premium and consumption growth is negligible using real consumption growth of non-durables and services (1990:M1-2017:M12) to measure consumption growth. The correlation with the variance premium at time t is -0.13 using "future" consumption growth between time t and $t+1$; and -0.05 using consumption growth between time $t-1$ and t correlated with the variance risk premium at time t.

However, there is a strong correlation in the tails of the distribution. When the variance risk premium is relatively high, the consumption growth distribution becomes more negatively skewed, and the quantile shifts for the consumption growth distribution when conditioning on low versus high variance risk premiums are statistically significant for the left tail. These empirical facts pose a challenge for most of the models we examine.

Finally, as our third stylized fact, we rely on Martin (2017) who raises several issues regarding the measurement of risk neutral variances and shows that the difference between the risk-neutral entropy and the risk-neutral variance of returns provides a challenging data moment for various models in the literature. The moderately positive difference between the risk-neutral entropy and variance of returns is an economically important moment, because it tells us that the left tail of the risk-neutral aggregate equity return distribution is heavier than the right tail, but the difference is not large. We examine a wider set of models than Martin (2017) does, taking the opportunity to clarify the empirical and theoretical differences between various concepts of option-implied variances proposed in the literature.

We find that all models are rejected by these stylized facts. Therefore, we introduce a tractable version of the "Bad Environment Good Environment" (BEGE henceforth) framework of Bekaert and Engstrom (2017), which can actually fit the data. We estimate two versions of the model. In the main version, all risk premiums are driven by "good" (positively skewed) and "bad" (negatively skewed) consumption growth volatility state variables, which also drive the variation in stochastic risk aversion ("habit"). In this model the risk-neutral variance and thus the variance risk premium, loads much more heavily on "bad" volatility, than does the physical variance, generating a sizable variance risk premium. The simplest version of the model, as do all other models, does fail to fit the persistence of the variance risk premium, which is surprisingly low (less than 60%). Most asset prices require highly persistent state variables to generate sufficiently variable equity returns and price dividend ratios. While different parameter configurations of our base model can fit the variance risk premium persistence, the fit wit respect to other moments then deteriorates slightly. The model can fit the variance risk premium persistence well, when when we allow for a small pure sentiment shock.

The remainder of the article is organized as follows. Section 2 establishes the stylized facts in the data. Section 3 outlines the various existing consumption based models we examine and how they fit the stylized facts. Section 4 describes our new model and its fit with the data. To provide more over-identification, section 5 investigates how the various models fit the correlation between volatility and equity premiums. Section 6 concludes.

2 The Variance Risk Premium in the Data

2.1 The Variance Risk Premium

The variance risk premium is usually defined as the difference between the risk neutral and physical conditional variance of stock returns. The risk-neutral variance can be computed using option prices or variance swaps (see Bakshi and Madan, 2000; Martin 2017, and Ait-Sahalia, Karaman, and Mancini, 2018). For now, we simply use the square of the well-known VIX index, published by the CBOE, the implied option volatility of the

S&P500 index for contracts with a maturity of one month.[1] The risk-adjusted measure shifts probability mass to states with higher marginal utility (bad states) and this implies that in many realistic economic settings, the variance risk premium is increasing in the economy's risk aversion. Our data start on January 02, 1990 (the start of the model-free VIX series)[2] and covers the period until the end of 2017.

We collect high-frequency (5 minute) returns on the S&P500 index to compute the monthly physical conditional variance, V_t, as:

$$V_t = E_t[RV_{t+1}^{(22)}]. \tag{1}$$

Here $RV_{t+1}^{(22)}$ is the S&P500 realized variance measured as the sum of squared 5 minute returns and close-to-open overnight returns over the next month (22 trading days).[3] The common approach to estimate the conditional variance in (1) uses empirical projections of the realized variance on variables in the information set. Hence, the problem is reduced to one of variance forecasting. Building on Corsi (2009) and Bekaert and Hoerova (2014), we use the one period lagged realized monthly variance, realized variances of the last and last 5 trading days (computed using high-frequency data) and the squared VIX, as predictors. While Bekaert and Hoerova (2014) show that alternative, more complicated models sometimes provide a slightly better fit, this model always provides a very good fit. For robustness, we also consider an AR(1) model for realized variances and fit a simple GJR-GARCH(1,1) model (Glosten, Jagannathan, and Runkle, 1993) on stock returns to extract the conditional variance, with no meaningful differences in results.

We graph the annualized end-of-month variance risk premium in the top panel of Figure 1. The variance risk premium is counter-cyclical peaking in all three recessions but also in 1998 and 2011. The variance risk premium as defined above has unintuitive economic units: for instance, the annualized mean is 0.0196. In some of our empirical work, we therefore employ the "volatility risk premium", the conditional risk neutral

[1]Jiang and Tian (2005) show that the actual computation of the VIX index also introduces errors relative to the theoretical concept.

[2]The CBOE changed the methodology for calculating the VIX, initially measuring implied volatility for the S&P100 index, to be measured in a model-free manner from a panel of option prices (see Bakshi, Madan and Kapadia, 2003, for details) only in September 2003. It then backdated the new model-free index to 1990 using historical option prices.

[3]We use actual simple returns in these computations, whereas some articles suggest using logarithmic returns. However, we find that realized variances using either method are indistinguishable from each other, which is not surprising given the high frequency nature of the returns.

minus physical conditional volatility, that is $VIX_t - V_t^{\frac{1}{2}}$. These numbers, in annualized percent, are easy to interpret. For example, for our sample, the unconditional stock market volatility is 14.64%, the average conditional volatility is 14.42%, and the average volatility premium is 5.36%. The volatility of the volatility premium is 4.02%. We graph the end-of-month volatility risk premium in the bottom panel of Figure 1. To avoid any confusion, we always refer to $VIX_t^2 - V_t$ as the variance risk premium and to $VIX_t - V_t^{\frac{1}{2}}$ as the volatility risk premium.

2.2 Consumption Growth and the Variance Risk Premium

To link the variance risk premium to consumption growth, we obtain U.S. monthly consumption growth for non-durables and services from NIPA from 1990:M1 to 2017:M12. Our goal is to verify the shape of the consumption distribution as a function of variance risk premium realizations. The data paucity necessitates us to contrast just two conditional distributions, depending on either "low" or "high" variance risk premiums.

In Table 1, we show the 10^{th} and 90^{th} percentiles of the consumption growth distribution together with the median, conditional on observing either a high or low variance risk premium. We define a high (low) variance premium as one above (below) the 80^{th} (20^{th}) unconditional percentile in the data over the sample period. Conditioning on more extreme tails is impossible given the scant number of monthly observations we have. Strikingly, the distribution of consumption growth is nearly identical at the median and 90^{th} percentile, but the lower tail is 0.21% lower (2.4 percent lower at an annual rate) in periods of high variance risk premiums. This difference is statistically significant at the 1% level, where the significance is based on a block-bootstrap with the block length of 60 months using 10,000 replications of historical length.

Figure 2 presents a graphical illustration of what is essentially a quantile shift of the negative tail of the consumption growth distribution. It shows that the entire distribution below the median shifts down going from low to high variance risk premiums. The shift at the 20^{th} percentile is also significant (at the 10% level). This downward quantile shift reveals a tantalizing link between the real economy and option prices. It also immediately reveals that the data generating process for consumption growth must accommodate a shift in its distribution over time. Because this empirical fact is an important ingredient in our analysis, we provide some robustness checks in Table 2. In Panels A and B, we show robustness of the result to an alternative choice of the lower/upper percentile, using the $75/25^{th}$ and $85/15^{th}$ percentiles. In Panel C, we use the conditional variance of a GARCH model to compute the variance risk premium and in Panel D the VIX itself.[4] The quantile shift happens in both cases, but is now only statistically significant at the 5% level. Thus, the general result is that in times of great financial uncertainty, as measured by the option implied variance of stock returns or its difference with physical variance, consumption growth appears more left skewed. In the context of a consumption-based asset pricing model, this means that financial uncertainty as priced into options may well reflect real consumption risks.

As a last robustness check, we verify whether the result holds up for a much longer sample by obtaining annual real per capita consumption growth of non-durables and services for the 1929-2017 period. This period has fewer time series observations than our monthly sample, but witnessed multiple, often severe, recessions. The variance risk premium is only directly observable for the 1990-2017 period. In order to obtain estimates of the variance risk premium for the 1929-1989 period, we regress the variance risk premium during 1990-2017 on the sum of squared realized daily returns excluding dividends for the past week, month, and quarter and the price-to-earnings ratio (as the price-dividend ratio exhibits a pronounced time trend during the sample). Note that here we are not able to use the high-frequency realized returns, because they are not available in the early sample. We use realized returns excluding dividends, because it implies a slightly higher explanatory power for the variance risk premium. It does not affect our consumption growth shift results. We save the regression coefficients and standard deviation of the residuals. We then conduct a block-bootstrap analysis, as follows:

1. We block-bootstrap the annual 1929-2017 data using a block length of 5 years.

2. Inside the bootstrap, for observations falling in between 1929 and 1989, we input the variance risk premium using the independent variables at that point of time with the OLS coefficient, previously estimated, plus a randomly sampled Gaussian shock with zero-mean and the standard deviation equal to the standard deviation of the regression residuals above.

3. For the block-bootstrapped data, we compute percentiles of the consumption growth distribution conditional on a high variance risk premium realization, defined as above the 80^{th} percentile of the sampled variance risk premium distribution, and the low

[4]Bekaert and Engstrom (2017) report a similar result for the VIX.

variance risk premium, defined as below the 20^{th} percentile of the sampled variance risk premium distribution.

The point estimates for the conditional consumption growth percentiles are medians across 10,000 block-bootstrap runs. The statistical significance of the difference between consumption growth percentiles conditional on high and low variance risk premium values is computed as the percentage of block-bootstrap runs where the difference is below/above 0. Figure 3 shows the results. The downward shift of the consumption growth distribution when the variance risk premium is high is very similar to what we observe for the most recent monthly data. The shifts are statistically significant at the 5% level for the 10^{th} and 20^{th} percentiles. The 10^{th} percentile shifts from slightly positive consumption growth when the variance premium is low to -4% when the variance premium is high. Generally, the shift is a bit more extreme than with monthly data and starts to be already visible around the 60^{th} percentile.

2.3 Martin's (2017) "bound"

The VIX is actually a weighted average of call and put option prices. Because the weights are proportional to the inverse of the squared strike price (see Bakshi and Madan, 2000, and Britten-Jones and Neuberger, 2000), out of the money put prices receive relatively more weight. Martin (2017) shows that the VIX^2 can be interpreted as twice the risk-neutral entropy of the simple return (the entropy for a variable X is $2 \cdot (\ln[E(X)] - E[\ln(X)])$), and then shows that the risk neutral variance can be approximated using equally weighted call and option prices. He denotes the square root of the risk-neutral variance the "SVIX". While empirically SVIX and VIX are typically close to one another, the VIX is always higher than SVIX. Economically, the difference between VIX and SVIX is an informative moment, because it indicates that the left tail of the risk-neutral distribution of the aggregate equity return is only moderately heavier than the right tail. Martin (2017) then shows that this difference is very difficult to fit by existing consumption-based models. While Martin's article is mostly about using the SVIX index as a lower bound for the equity premium, we use the empirical difference between the VIX and SVIX as a powerful statistic to help refute existing consumption based asset pricing models.

3 The Variance Risk Premium and the Consumption-based Asset Pricing Literature

The previous section has uncovered three sets of stylized facts involving the volatility premium and option implied volatility, which we will demonstrate to be very challenging for existing consumption-based asset pricing models to fit. Here, we survey the various models and the mechanism they use to generate a meaningful volatility premium, and then verify whether they actually fit the stylized facts. We start with a theoretical result showing that log-normal models cannot possibly fit the facts, which is powerful as it immediately refutes the original formulation of the habit and long-run risk models as suitable candidates.

3.1 The Variance Risk Premium under Log-Normality

Campbell and Cochrane's (1999) habit model and the first-order approximation to Bansal and Yaron's (2004) long-run risk model are conditionally Gaussian.[5] Various results in the extant literature (see e.g. Bakshi and Madan, 2006, and Bekaert and Hoerova, 2014) suggest non-Gaussianities in the data generating process for returns are necessary to produce a positive variance risk premium. Drechsler and Yaron (2011) and Martin (2017) prove a more general result, indicating that when returns and the pricing kernel are jointly log-normally distributed, the variance risk premium is exactly zero. In Internet Appendix I, we provide an alternative proof of this theorem, making use of a risk neutral moment generating function.

Another result that can be derived rather straightforwardly and offers useful economic intuition about the variance risk premium is that log-normal models in general cannot generate simultaneously a positive variance risk premium and a positive equity premium. The result is rather intuitive: for the risk neutral variance (which apart from pure physical volatility also reflects covariation of returns with the pricing kernel) to be consistently

[5] In the long-run risk literature, it is customary to assume that endogenous variables are linear functions of the state variables. Pohl, Schmedders, and Wilms (2018) show that this approximation is actually rather poor in many settings. Lorenz, Schmedders, and Schumacher (2020) show this specifically in the context of the Drechsler and Yaron (2011) long-run risk model aimed at explaining the variance risk premium. However, there are almost no published long-run risk papers that use more accurate solution techniques. The solution method for the habit model in Campbell and Cochrane (1999) was shown to be inaccurate by Wachter (2005). In the Campbell and Cochrane-type of models below, we use Wachter's more accurate solution method to solve the model.

above the physical volatility, the covariance between returns and the kernel must be positive. However, if equity returns tend to offer high returns in bad times (states with high marginal utility), the equity premium must be negative. We defer a formal proof to Internet Appendix I.

3.2 Extant Models of the Variance Risk Premium

We reconsider how four different models fit the stylized facts outlined in Section II. The first model is the "vol of vol" model of Bollerslev, Tauchen and Zhou (2009). It is important to mention that this paper is one of the first to show that the variance risk premium predicts stock returns, and the theoretical part of the paper was perhaps not likely meant as the key contribution. They consider a representative agent with Epstein-Zin (1989) preferences and consumption (and dividend) growth featuring stochastic volatility. A second state variable drives time-variation in the volatility of the volatility shocks ("vol of vol"). The second model, Drechsler and Yaron (2011), is a straightforward extension of the long-run risk model meant to fit variance risk premium features. Drechsler and Yaron (2011) add several components to the long-run model including a slow moving component of the volatility and importantly jumps to the long-run risk variable (the conditional mean of consumption growth) and to volatility. We use the model in Wachter (2013) as the representative of the disaster risk paradigm. The model features Epstein-Zin preferences, and (disaster) jumps to the consumption and dividend shocks. While the model was not designed to fit options data, Wachter and Seo (2019) show that this particular specification fits option prices rather well. Finally, we consider the habit model of Bekaert and Engstrom (2017). They add a "bad environment-good environment" structure for consumption growth to Campbell and Cochrane's set up. In this model, consumption growth features "bad" and "good" volatility, with shocks to bad (good) volatility decreasing (increasing) skewness in consumption growth. We now briefly discuss the models in detail.

3.2.1 Bollerslev, Tauchen, and Zhou (2009)

The utility function is:

$$U_t = \left[(1-\delta)C_t^{\frac{1-\gamma}{\theta}} + \delta(E_t\left[U_{t+1}^{1-\gamma}\right])^{\frac{1}{\theta}}\right], \tag{2}$$

where C_t is consumption at time t, $0 < \delta < 1$ reflects the agent's time preferences, γ is the coefficient of relative risk-aversion, $\theta = \frac{1-\gamma}{1-\frac{1}{\psi}}$ and ψ is the intertemporal elasticity of substitution.

The dynamics for log consumption and dividend growth, g_{t+1} and d_{t+1}, respectively, are:

$$\begin{aligned}
&g_{t+1} = d_{t+1} = \mu_g + \sigma_{g,t} z_{g,t+1}, \\
&\sigma^2_{g,t+1} = a_\sigma + \rho_\sigma \sigma^2_{g,t} + \sqrt{q_t} z_{\sigma,t+1}, \\
&q_{t+1} = a_q + \rho_q q_t + \phi_q \sqrt{q_t} z_{q,t+1}, \\
&z_{g,t+1} \sim \mathcal{N}(0,1), z_{\sigma,t+1} \sim \mathcal{N}(0,1), z_{q,t+1} \sim \mathcal{N}(0,1),
\end{aligned} \tag{3}$$

where μ_g is the consumption growth mean, $\sigma^2_{g,t}$ the conditional variance of the consumption growth, and q_t is the conditional variance of the consumption growth variance. It's the latter variable that drives variance risk premium variation in this article.

The model is calibrated monthly to fit reasonable unconditional levels of the equity premium and risk-free rate and the slope coefficient from regressing excess equity returns on the variance of the risk premium. However, for this calibration exercise the authors do not refer to any particular time period. The parameters are reported in Internet Appendix II.

3.2.2 Drechsler and Yaron (2011)

The utility function is:

$$U_t = \left[(1-\delta) C_t^{\frac{1-\gamma}{\theta}} + \delta (E_t [U_{t+1}^{1-\gamma}])^{\frac{1}{\theta}}\right]. \tag{4}$$

There are 5 macroeconomic variables, which dynamics follow:

$$\begin{bmatrix} g_{t+1} \\ x_{t+1} \\ \bar{\sigma}^2_{t+1} \\ \sigma^2_{t+1} \\ d_{t+1} \end{bmatrix} = \begin{pmatrix} g_0 \\ x_0 \\ \bar{\sigma}^2_0 \\ \sigma^2_0 \\ d_0 \end{pmatrix} + \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & \rho_x & 0 & 0 & 0 \\ 0 & 0 & \rho_{\bar{\sigma}^2} & 0 & 0 \\ 0 & 0 & (1-\rho_{\sigma^2}) & \rho_{\sigma^2} & 0 \\ 0 & \phi & 0 & 0 & 0 \end{pmatrix} \begin{bmatrix} g_t \\ x_t \\ \bar{\sigma}^2_t \\ \sigma^2_t \\ d_t \end{bmatrix} + G_t z_{t+1} + \tilde{J}_{t+1}, \tag{5}$$

where g_t is logarithmic consumption growth, x_t is the persistent component of consumption growth and d_t denotes dividend growth. The volatility dynamics is governed by two factors: σ_t^2 represents the conditional volatility and $\bar{\sigma}_t^2$ is the long-run mean component of σ_t^2.

The vector z_{t+1} represents Gaussian innovations with G_t capturing time-variation in volatility:

$$
\begin{aligned}
&z_{t+1} \sim \mathcal{N}(\mathbf{0}_{5\times 1}, \mathcal{I}_{5\times 5}), \qquad (6)\\
&G_t G_t' = diag(\varphi \bullet \sqrt{1-\omega})\,\Omega\, diag(\varphi \bullet \sqrt{1-\omega})' + diag(\varphi \bullet \sqrt{\omega})\,\Omega\, diag(\varphi \bullet \sqrt{\omega})'\sigma_t^2,\\
&\varphi = \begin{pmatrix} \varphi_g \\ \varphi_x \\ \varphi_{\bar{\sigma}^2} \\ \varphi_{\sigma^2} \\ \varphi_d \end{pmatrix}, \omega = \begin{pmatrix} \omega_g \\ \omega_x \\ \omega_{\bar{\sigma}^2} \\ \omega_{\sigma^2} \\ \omega_d \end{pmatrix}, \Omega = \begin{pmatrix} 1 & 0 & 0 & 0 & \Omega_{cd} \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ \Omega_{cd} & 0 & 0 & 0 & 1 \end{pmatrix},
\end{aligned}
$$

where $diag$ is the vector-to-diagonal matrix operator and $\bullet$ is the element-wise multiplication operator.

$\tilde{J}_{t+1}$ is a 5×1 vector of demeaned jump shocks: $\tilde{J}_{t+1} = J_{t+1} - E_t J_{t+1}$. J_{t+1} is a 5×1 vector of compound-Poisson jumps: $J_{t+1,i} = \sum_{j=1}^{N_{t+1}^i} \xi_i^j$, where N_{t+1}^i is the Poisson counting process for the i^{th} jump component and ξ_i^j is the size of the jump that occurs upon the j^{th} increment of N_{t+1}^i. The intensity process for N_{t+1} is $l_1\sigma_t^2$, where $l_1 = \begin{pmatrix} 0 & l_{1,x} & 0 & l_{1,\sigma^2} & 0 \end{pmatrix}'$: there are only jumps in the variance and the persistent component of consumption growth.[6] The persistent component of the expected consumption growth jump size follows an i.i.d. demeaned Gamma distribution multiplied by -1 (that is, the distribution is negatively skewed with limited right and unlimited left tails): $\xi_x \sim -\Gamma(\nu_x, \frac{\mu_x}{\nu_x}) + \mu_x$. The variance jump size follows an i.i.d. Gamma distribution: $\xi_{\sigma^2} \sim \Gamma(\nu_{\sigma^2}, \frac{\mu_{\sigma^2}}{\nu_{\sigma^2}})$. Here, ν_i is the shape and $\frac{\mu_i}{\nu_i}$, $(i = x, \sigma^2)$ are the scale parameters, respectively. These jump variables deliver the potential non-Gaussianities driving variation in the variance risk premium. However, consumption growth itself is conditionally Gaussian in this model.

The model is calibrated monthly to match a wide set of unconditional macroeconomic and financial moments of quarterly US data 1930-2006. The parameters are in Internet

[6] There are no jumps in the long-run mean of the volatility.

Appendix II.

3.2.3 Wachter (2013)

Wachter's model is formulated in continuous time. The utility function is:

$$U_t = E_t \int_t^{\infty} \beta(1-\gamma)U_s(\ln C_s - \frac{1}{1-\gamma}\ln((1-\gamma)U_s)ds. \tag{7}$$

Consumption (C_t) and dividends (D_t) follow:

$$\begin{aligned} dC_t &= \mu C_{t^-}dt + \sigma C_{t^-}dB_t + (e^Z - 1)C_{t^-}dN_t, \\ dD_t &= (\phi\mu + \frac{1}{2}\phi(\phi-1)\sigma^2)D_{t^-}dt + \phi\sigma D_{t^-}dB_t + (e^{\phi Z} - 1)D_{t^-}dN_t, \end{aligned} \tag{8}$$

where B_t is a Brownian motion and N_t is a Poisson process with a time-varying intensity λ_t: $d\lambda_t = \kappa(\bar{\lambda} - \lambda_t)dt + \sigma_\lambda\sqrt{\lambda_t}dB_{\lambda,t}$ with $B_{\lambda,t}$ also a Brownian motion. B_t, $B_{\lambda,t}$, and N_t are independent. Z is a time-invariant distribution independent of B_t, $B_{\lambda,t}$, and N_t which determines the jump (disaster) size. All processes are assumed to be right continuous with left limits. For process x, x_{t^-} denotes $\lim_{s\uparrow t} x_s$ (intuitively, this corresponds to approaching from $s < t$), and x_t denotes $\lim_{s\downarrow t} x_s$ (intuitively, this corresponds to approaching from $s > t$).

Parameters are chosen through a combination of estimation and calibration. First, the distribution of jumps in consumption $(e^Z - 1)$ is estimated from a set of 17 OECD and 5 non-OECD countries between 1870 and 2006. The unconditional jump (disaster) probability, $\bar{\lambda}$, is taken from Barro and Ursua (2008). Second, the remaining parameters are calibrated to match a set of unconditional macroeconomic and financial moments of quarterly US data for the 1947-2010 period. Importantly, the moments are matched for the sample conditional on no disasters, since there have been no consumption disasters in 1947-2010 US data. For this purpose, the model is Euler-discretized and sampled at the monthly frequency and months with no disasters are picked to compute model-implied moments. The parameters are reported in Internet Appendix II.

3.2.4 Bekaert and Engstrom (2017)

The utility function falls into the external habit class:

$$E_t \sum_{j=t}^{\infty} \delta^{j-t} \frac{(C_j - H_j)^{1-\gamma} - 1}{1-\gamma}, \tag{9}$$

where C_j is consumption and H_j is the habit stock with $C_j > H_j$.

Log-consumption and dividend growth (g_{t+1} and d_{t+1}, respectively) follow:

$$\begin{aligned}
g_{t+1} &= \bar{g} + \sigma_{cp}\omega_{p,t+1} - \sigma_{cn}\omega_{n,t+1}, \\
d_{t+1} &= \bar{g} + \sigma_{dp}\omega_{p,t+1} - \sigma_{dn}\omega_{n,t+1}, \\
\omega_{p,t+1} &\sim \Gamma(\bar{p}, 1) - \bar{p}, \\
\omega_{n,t+1} &\sim \Gamma(n_t, 1) - n_t, \\
n_{t+1} &= \bar{n} + \rho_n(n_t - \bar{n}) + \sigma_{nn}\omega_{n,t+1},
\end{aligned} \tag{10}$$

where $\Gamma(x, y)$ is a gamma distribution with shape parameter x and scale parameter y.

Following Campbell and Cochrane (1999), $S_t = \frac{C_t - H_t}{C_t}$, which can be interpreted as the consumption surplus ratio, is modeled in logs as an autoregressive process:

$$\begin{aligned}
s_{t+1} &= \bar{s} + \phi(s_t - \bar{s}) + \lambda_t(g_{t+1} - \bar{g}), \\
\lambda_t &= \begin{cases} \frac{1}{s_t}\sqrt{1 - 2(s_t - \bar{s})} - 1, & \text{if } s_t < s_{t,max} \\ 0, & \text{otherwise} \end{cases}, \\
s_{t,max} &= \bar{s} + \frac{1}{2}(1 - \bar{S}_t^2), \\
\bar{S}_t &= \sqrt{(\sigma_{cp}^2 \bar{p} + \sigma_{cn}^2 n_t)\frac{\gamma}{1 - \phi - \frac{b}{\gamma}}}.
\end{aligned} \tag{11}$$

Note that the surplus ratio shock is perfectly correlated with consumption growth. The modeling of the price of risk variable, λ_t, is analogous to the specification in Campbell and Cochrane (1999), adjusted for the presence of consumption heteroskedasticity.

The model is estimated in 3 steps. First, the consumption growth parameters ($\bar{g}$, σ_{cp}, σ_{cn}, $\bar{p}$, $\bar{n}$, ρ_n, and σ_{nn}) are estimated via classical minimum distance to match unconditional moments of quarterly US consumption growth for the 1958-2013 period. Second, dividend growth parameters (σ_{dp} and σ_{dn}) are estimated to match unconditional dividend growth volatility and the correlation between consumption and dividend growth. Third, the preference parameters (δ, γ, $\bar{s}$, ϕ, and b) are estimated to minimize the distance between the model-implied and 1958-2013 US unconditional asset pricing moments. The parameters are reported in Internet Appendix II.

3.3 General Asset Return Properties

We start by showing how the various models fit the standard salient asset features in Table 3. All the moments shown are annualized monthly values. For the data moments, we show two samples, 1990 to 2017 to correspond to the available data we have for the variance risk premium; and a longer sample extending to 1969. The real risk-free rate is measured as the difference between the monthly nominal risk-free rate from Ibbotson Associates and the monthly counterpart to the Survey of Professional Forecasters expected inflation for the corresponding quarter; specifically, with π the quarterly inflation forecast, we use $(1+\pi)^{\frac{1}{3}}$. The risk-free rate is lower in the short sample (0.64%) than in the longer sample (1.35%) and has a variability of around 2%. For the stock market, we use logarithmic returns on the S&P500 index in excess of the risk-free rate. The equity premium is 4.92% over the short sample and 6.15% over the longer sample with the volatility at about 15%. We also report the mean and variance of the price dividend ratio, noting that this variable has been subject to trending behavior due to tax policy changes in the U.S. (Boudoukh et al., 2007). The standard errors in parentheses are obtained by block-bootstrapping 10,000 times series of historical length using a block length of 60 months, to accommodate persistence in the levels and volatilities of the variables.

Because we use the calibrations/estimations provided in the various original papers, we do not expect the various models to fit all these moments exactly. The fit of the BTZ model is particularly poor, because their calibration focused on the unconditional equity premium and risk-free levels. However, their calibration also implies very unrealistic values for other important moments. These values are not reported in the article but can be directly computed from the formulas in the paper. For instance, the average annualized physical variance of the equity return obtained by plugging the values into equation (12) in the paper is 5.70%, which is about one third of the data counterpart. As another example, plugging the numbers into equation (6) in the paper results in an average annual price dividend ratio of 2.04. A more serious problem with the model part in Bollerslev, Tauchen, and Zhou (2009) is that the model parameters imply that the consumption growth variance ($\sigma^2_{g,t}$) and the variance of consumption growth variance (q_t) both hit the zero-lower bound in more than 10% of the simulations. This results in significant deviations of simulated asset prices from their theoretical counterparts. For instance, theoretically the unconditional equity premium in Table 3 is 7.79%, but the population mean from sampling 100,000 observations is 14.12%. The population value is much higher, because the equity premium is increasing in consumption growth volatility and the volatility of volatility variables, and simulated values for these volatilities are higher than implied by the theoretical model, because the left tails of their distributions are cut due to the zero-lower bound. Analogously, sampling 100,000 observations to infer the population mean for the interest rate delivers an average risk-free rate of -2.95% compared to the theoretical value of 0.69% in Table 3. The population value is much lower for the same reason as before, but now the risk-free rate is decreasing in the consumption growth volatility and the volatility of volatility variable. Apparently, the theoretical asset pricing formulas are unreliable when zero-lower bounds are violated so frequently.

The DY model generally does well with respect to the risk-free rate and equity return moments generating values for the means of both variables within a two standard deviation band around the data moments. It slightly overshoots equity return volatility, but it also underestimates the variability of the price-dividend ratio by an order of magnitude.

The asset pricing statistics for Wachter's model are computed for the population including disasters. Asset pricing statistics for the sample excluding disasters (which Wachter argues to be the best comparison for the post-war US data) are not very different (for instance, the average equity premium is slightly higher and less volatile), because the intensity of the disaster (λ_t) and shocks to it, not the disaster realization itself, are the key variables driving return dynamics. The sample including realized disasters, unlike the sample excluding disasters, implies rather extreme consumption and dividend growth statistics: for instance, the annualized consumption growth volatility in the sample with disasters is 6.22% versus 2.00% in the sample without disasters. However, a consumption growth model without disasters would have no chance to fit the link between consumption growth and variance risk premiums. Table 3 reports the "true" (default-free) risk-free rate in Wachter (2013), whereas Wachter (2013) defines a government bond rate assuming the bond defaults with a probability of 40% when a consumption disaster occurs. This results in a substantially higher "risk-free" rate of 1.00% versus 0.47% in Table 3. Wachter's model generally fits the salient asset return features well, producing average risk free rates and an equity premium close to the data moments. It does generate excessive risk-free rate and equity return volatility, and also overshoots both the mean and volatility of the price-dividend ratio.

The BEGE model fits the risk-free rate and equity return moments, with the exception of generating interest rate volatility that is slightly too low. While doing better than the BTZ and DY models, it still undershoots the mean and the variability of the price-dividend ratio. Overall, while statistically all models are strongly rejected by the data, we argue that DY, Wachter (2013), and BEGE fit general asset prices reasonably well.

3.4 Variance Risk Premium Fit

Table 4 focuses on variance risk premium statistics. As indicated before, the variance risk premium is measured by the difference between VIX^2 and the expected physical return variance. The physical variance is obtained by linearly projecting realized monthly variances computed using high-frequency data onto their one month lagged values, VIX^2, and squared S&P500 daily returns for the previous trading day and previous 5 trading days. This model is one of the better models in Bekaert and Hoerova (2014)'s out-of-sample horse race for forecasting physical realized variances. It simply adds the squared VIX to the well-known "HAR-RV" forecasting model of Corsi (2009). The variance premium is only available in the shorter sample and has a mean of 0.0195 and a volatility of 0.0225, with these moments rather precisely estimated. For ease of economic interpretation, Panel B of Table 4 also reports the volatility premium fit, which simply replaces the variances by volatilities.

To compute the variance risk premium in the models, our first set of statistics use the risk-neutral minus physical annualized variance of the aggregate stock market log-return. Following most of the structural literature, we report the statistics for log instead of gross returns, but the differences between the two concepts are economically small. As discussed in Martin (2017), the VIX^2 actually represents the risk-neutral entropy and always exceeds the risk neutral variance in the data. We return to this issue separately in the next subsection.

Table 4 reveals that the variance risk premium is identically equal to zero in the BTZ model, as it is a log-normal model (see section 3.1). Both the DY and BEGE models generate meaningful variance risk premiums, but they are still too small relative to the data. The Wachter's model generates a much too high volatility risk premium of 15%. It also overshoots the variability of the volatility risk premium, whereas the DY and BEGE models undershoot it.

Our definition of the variance risk premium is different than the way it is defined in the original BTZ and DY articles. To illustrate the difference, let r_{t+1} be the aggregate equity log-return between time t and $t+1$ while r_{t+2} is the aggregate equity log-return between time $t+1$ and $t+2$. BTZ define the variance risk premium as the difference between the risk neutral and physical variance skipping one period (month): $Var_t^Q(r_{t+2}) - Var_t(r_{t+2})$, instead of $Var_t^Q(r_{t+1}) - Var_t(r_{t+1})$. This allows them to generate a positive variance risk premium instead of the zero value under log-normality, which we report, because returns skipping one period are not log-normal in their model. DY define the variance risk premium as the sum of the difference between the next month's risk-neutral and physical variance, $Var_t^Q(r_{t+1}) - Var_t(r_{t+1})$ (which they call the "level difference"), and the difference between the risk-neutral and physical variance skipping one month, $Var_t^Q(r_{t+2}) - Var_t(r_{t+2})$ (which they call the "drift difference"). This additional "drift difference" component allows DY to generate a much higher average variance risk premium: 12.62% instead of 2.35% in Table 4 (which only takes into the account the "level difference", the $Var_t^Q(r_{t+1}) - Var_t(r_{t+1})$ term). However, both the BTZ and DY definitions of the variance risk premium are inconsistent with the real-world specification of the VIX and variance swaps.

Finally, we report the autocorrelation of the variance risk premium. The premium's autocorrelation is only 0.52 in the data. This low value may reflect measurement error, because we require an empirical model to measure the physical variance, which surely introduces some noise. For instance, Figure 1 illustrates that the variance risk premium is particularly volatile during the Great Recession when financial markets were in turmoil, potentially leading to imprecise measurement. However, even during the period pre-Great Recession (January 1990-November 2007), the variance risk premium autocorrelation was only 0.49. Furthermore, the value of around 0.50 is robust to alternative models of the conditional physical variance.

Nevertheless, all the models generate very persistent autocorrelation. The main reason for this is that the model's state variables, which determine all asset prices, including the variance risk premium, are all very persistent. This persistence is required, so that realistically small shocks to the state variables generate realistically large asset pricing implications.[7]

3.5 Martin (2017) bounds

In Table 5, we report the average VIX and SVIX, computed as proposed by Martin (2017), and the average difference for the 1996-2012 sample, the sample period used in Martin's (2017) article. The difference is 1.41% on average. None of the models can actually fit the Martin's bounds. In line with Result 4 in Martin (2017), BTZ counterfactually implies a negative difference between VIX and SVIX as it is a log-normal model. The rare disasters model of Wachter (2013) substantially overstates the differnce (6.48% versus 1.41% in the data). The BEGE and DY models do generate smallish positive SVIX-VIX differences, with the BEGE model slightly better than the DY model (0.18% versus 0.08%), although the difference is not substantial. Both models still dramatically underfit the VIX-SVIX difference.

The economic intuition behind this result is explained in Section VII of Martin (2017). The VIX^2 can be viewed as a risk-neutral return entropy while $SVIX^2$ is a measure of the risk-neutral return variance; viewed as a portfolio of options, the VIX weights option prices by the inverse of squared strike prices, whereas the SVIX uses equal weighting. Therefore, the VIX^2 is more sensitive to left-tail events (relative to $SVIX^2$), while $SVIX^2$ is more sensitive to right tail events (relative to VIX^2). Thus, the disaster model generating a VIX much higher than the $SVIX$ indicates that left-tail return outcomes are too severe compared to right-tail outcomes in the model, compared to the data. Similarly, the other models fail to generate a large enough VIX compared to the $SVIX$, indicating that the left-tail return outcomes are not severe enough compared to right-tail outcomes in these models compared to the data. Overall, fitting options data requires that the left side of the risk-neutral return distribution is somewhat, but not too much, heavier than the right side of the distribution.

No model that we examined produces the tail behavior implied by the Martin (2017) bounds. Note that BTZ is not able to produce such a behavior even theoretically, as it is log-normal (see section 3.1), whereas the disaster risk model in Wachter (2013) cannot

[7]Another potentially interesting moment is the skewness of the variance or volatility risk premium. However, we find that the skewness of the variance risk premium is not a robust moment, varying between sharply negative and sharply positive depending on how the physical variance is computed, while it is close to zero using our baseline methodology.

produce this behavior without altering its basic asset pricing fit, because a heavy left-tail of consumption growth is required to reproduce all key asset pricing moments, including the equity premium and low risk-free rate. This heavy left-tail then implies that risk-neutral entropy of the aggregate return is much higher than its risk-neutral variance, because the risk-neutral return distribution is strongly left-skewed. Of the four models we consider, only DY and BEGE are potentially able to reproduce the observed tail behavior, albeit under alternative parameterizations than in the original papers.

3.6 Consumption Growth Quantile Shifts and the Variance Premium

The essence of a consumption-based asset pricing model is to link actual consumption data to asset returns. In Section 2, we showed that there is a significant link between the downward shifts in the left tail of consumption growth and the incidence of high variance risk premiums. Here we examine which of the existing models can fit this data pattern. We collect the results in Table 6, reporting the results replicating the data results of Table 1 for samples of 1,000,000 months simulated from the various models. We show the 10^{th}, 50^{th} and 90^{th} percentile of the consumption growth distribution, conditioning on either the variance premium being above its 80^{th} unconditional percentile or below its 20^{th} percentile. We also report the difference in the furthermost right column.

Panel A of Table 6 indicates that BTZ model is not able to replicate the downward shift in the left percentiles of consumption growth following the high variance risk premium. Instead, in the model a high variance risk premium signals a low conditional variance of consumption growth. The underlying mechanism owes to the fact that a positive variance risk premium and equity risk premium cannot coexist in conditionally log-normal models (see Internet Appendix I) because the former (latter) is increasing (decreasing) in the covariance between returns and the stochastic discount factor. Let's denote this covariance as $\sigma_{mr,t}$. Note that $\sigma_{mr,t} = \rho_{mr,t}\sigma_{m,t}\sigma_{r,t}$, where ρ is the correlation coefficient and $\sigma_{m/r}$ represent standard deviations. In the BTZ calibration, $\rho_{mr,t}$ is negative, as it is a necessary condition to achieve a positive equity premium. At the same time, $\sigma_{m,t}$ and $\sigma_{r,t}$ are increasing in the conditional consumption growth volatility, because, in the BTZ model, dividend growth is assumed to equal consumption growth, and with Epstein-Zin preferences the pricing kernel encompasses both consumption growth and the aggregate wealth return. Thus, in BTZ a higher conditional consumption growth volatility decreases $\sigma_{mr,t}$, which then decreases the variance risk premium. Note that the conditional consumption growth volatility in the BTZ model is unrealistically high for both the high variance risk premium and, especially, for the low variance risk premium realizations. This is because the simulated consumption growth volatility is much higher than the consumption growth volatility implied by theory (which is realistic), due to the volatility often hitting the zero-lower bound as discussed in Section 3.3.

Panel B of Table 6 shows that the model of Drechsler and Yaron (2011) also cannot match the asymmetric percentile shifts documented before. Instead, the higher variance risk premium is associated with a higher "symmetric" volatility of consumption growth next period: that is left-tail percentiles shift to the left and right-tail percentiles shift to the right by the same amount. This occurs because in the model the variance risk premium is linearly proportional to σ_t^2 (the consumption growth volatility) and to the intensity of jumps in long-run consumption growth x_t and volatility (see equation (22) in Drechsler and Yaron, 2011). Economically, the variance risk premium is high when σ_t^2 is high, because σ_t^2 is the intensity of future jumps in long-run consumption growth x_t and the volatility of consumption growth, and an Epstein-Zin agent is averse to uncertainty about these jumps. Note from equation (4) in Drechsler and Yaron (2011) that possible jumps in x_t only affect consumption growth between time $t+1$ and $t+2$; however, consumption growth between t and $t+1$ is conditionally Gaussian and, thus, symmetric at time t, making asymmetric percentile shifts impossible. Theoretically, the model of Drechsler and Yaron (2011) could generate the percentile shift in the left tail of consumption growth distribution by skipping one month (that is, for consumption growth between $t+1$ and $t+2$) through the higher probability of a left-skewed jump in long-run consumption growth x_t. However, Panel C of Table 6 documents that this is not the case: while the shift happens, it is economically negligible. The magnitude is small because in long-run risk models, variation in x_t (where the jump happens and which represents predictable consumption growth) contributes little to the total variation in consumption growth. An alternative specification with normally (instead of gamma) distributed jumps in x_t considered by Drechsler and Yaron (2011), cannot generate even the minor shift shown in Panel C of Table 6, because Gaussian jumps are symmetric.

Panel D of Table 6 shows that Wachter's model generates a minuscule shift in the left conditional percentiles of the consumption growth distribution. The shift is much smaller than the one observed in the data. This is because the disasters are very extreme and, thus, affect percentiles of the distribution that are more extreme than the 10^{th} percentile on which we condition. This mismatch between the assumptions regarding consumption data in disaster risk models and option prices is reminiscent of but different than the evidence in Backus, Chernov, and Martin (2011), who show that options data imply less extreme disasters than those implied by disaster consumption models. Panel E of Table 6 shows that the BEGE model replicates the left-tail percentile shifts conditioned on the high variance risk premium reasonably well, although the shifts are of somewhat smaller magnitude than observed in the data. BEGE is able to replicate these shifts, through the bad environment shock $\omega^s_{n,t}$. A large bad environment shock increases the shape parameter of the bad consumption shock, shifting the left-tail percentiles of conditional consumption growth distribution down. Simultaneously, the shock decreases the surplus ratio increasing risk-aversion. These increases in the shape parameter of the bad consumption shock together with the increasing risk aversion then increase the variance risk premium (see Figure 8 in Bekaert and Engstrom, 2017). Note that the BEGE model also generates a small increase in the right-tail percentiles of the conditional consumption growth distribution following a high variance risk premium. This occurs because increasing the "bad" shape parameter also increases the magnitude of the right tail realizations from the bad environment component, although the magnitude is not nearly as strong as for the left tail because the right tail of the bad environment distribution component is finite (see Figure 3 in Bekaert and Engstrom, 2017). The evidence for such a right-tail shift in the data is mixed however (see Table 2).

To summarize, the DY and BTZ models are not able to reproduce the link between the downward shifts in the left tail of consumption growth and the incidence of high variance risk premiums even theoretically, because these models feature conditionally Gaussian one period ahead consumption growth (although DY model features non-Gaussian jumps in the variance of the *future* consumption growth and expected consumption growth). The Wachter (2013) model is not able to replicate the link, because the model requires negative consumption growth realizations which are too severe to fit key asset pricing moments. However, Bekaert and Engstrom's BEGE model does fit the conditional quantile shifts in consumption growth.

4 A New Model

Given that extant models cannot fully fit the empirical facts, we develop an alternative model here. We have established that BTZ is not able to generate a positive variance risk premium, because it is conditionally Gaussian. DY, while featuring a conditionally non-Gaussian stochastic discount factor, is not able to replicate conditional consumption growth quantile shifts, because consumption growth in the model is conditionally Gaussian. The rare disaster model implies that the risk-neutral entropy is too high compared to the risk-neutral variance due to the heavy left-tail of the consumption growth necessary to fit standard asset pricing moments. While Bekaert and Engstrom's (2017) BEGE model fits the variance risk premium moments poorly at the parameters used in the article, it is the only model featuring the economic mechanisms that can match the stylized facts under an alternative parameterization. However, the evaluation of the BEGE model requires a time-consuming numerical solution procedure, greatly decreasing its practical appeal. Thus, we propose a considerably more tractable version of the BEGE model, which allows for quasi closed-form asset pricing solutions, to study if the model is able to fit the stylized facts. In Section 4.1, we outline the model and derive expressions for the risk-neutral and physical variances and the variance risk premium. In Section 4.2, we provide estimation results for the model parameters and consider its fit with the data. Section 4.3 considers a slightly different model, focused on fitting the persistence of the variance risk premium. In Section 4.4, we consider a model incorporating a preference shock.

4.1 A Tractable BEGE-habit Model

The utility function is standard and given by:

$$E_t \sum_{j=t}^{\infty} \beta^{j-t} \frac{(C_j - H_j)^{1-\gamma} - 1}{1-\gamma}, \tag{12}$$

where β is the discount rate, C_j is consumption and H_j is the habit stock with $C_j > H_j$. Log-consumption growth, g_{t+1}, has a constant conditional mean but BEGE shocks:

$$g_{t+1} = \bar{g} + \sigma_{cp}\omega_{p,t+1} - \sigma_{cn}\omega_{n,t+1}, \tag{13}$$

where $\sigma_{cp} > 0$ and $\sigma_{cn} > 0$ and

$$\begin{aligned} \omega_{p,t+1} &\sim \Gamma(p_t, 1) - p_t, \\ \omega_{n,t+1} &\sim \Gamma(n_t, 1) - n_t, \end{aligned} \tag{14}$$

where $\Gamma(x, y)$ is a gamma distribution with shape parameter x and scale parameter y. Given that the mean value of a $\Gamma(x, y)$-distributed variable is $x \cdot y$, $\omega_{p,t+1}$ and $\omega_{n,t+1}$ are zero-mean.

The top plot of Figure 4 Panel A illustrates that the probability density function of $\omega_{p,t+1}$, the "good" component, is bounded from the left and has a long right tail. Similarly, the middle plot of Figure 4 Panel A shows that the probability density function of $-\omega_{n,t+1}$ (the "bad" component) is bounded from the right and has a long left tail. Finally, the bottom plot of Figure 4 Panel A plots the component model which has both tails.

We assume that the shape parameters follow autoregressive processes with the same shocks as the shocks driving consumption growth (see also Gourieroux and Jasiak, 2006):

$$\begin{aligned} p_{t+1} &= \bar{p} + \rho_p(p_t - \bar{p}) + \sigma_{pp}\omega_{p,t+1}, \\ n_{t+1} &= \bar{n} + \rho_n(n_t - \bar{n}) + \sigma_{nn}\omega_{n,t+1}. \end{aligned} \tag{15}$$

Panel B of Figure 4 illustrates possible conditional distributions of consumption growth shocks which could arise as a result of the time variation in the shape parameters in (15). In particular, the probability density function at the top plot of Figure 4 Panel B characterizes the situation where p_t is relatively large and the component distribution has a pronounced right tail, while the probability density function at the bottom plot of Figure 4 corresponds to the case where n_t is relatively large and the component distribution exhibits a pronounced left tail. Consequently, p_t (n_t) acts as good (bad) variance that is associated with positive (negative) skewness in consumption growth.

Analogously to Bekaert, Engstrom, and Grenadier (2010) and Bekaert, Engstrom, and Xu (2019), we model $Q_t = \frac{C_t}{C_t - H_t}$, which can be interpreted as an inverse consumption surplus ratio. Denoting $q_t = \ln(Q_t)$, we assume:

$$q_{t+1} = \bar{q} + \rho_q(q_t - \bar{q}) + \sigma_{qp}\omega_{p,t+1} + \sigma_{qn}\omega_{n,t+1}. \tag{16}$$

The q_t-process drives time-variation in the effective risk aversion of the representative agent. While we assume the process to be an AR(1), the shocks depend on the same shocks as consumption growth does. Economically, we expect $\sigma_{qn} > 0$ and $\sigma_{qp} < 0$. Unlike the models in Campbell and Cochrane (1999) and Bekaert and Engstrom (2017), consumption growth and risk aversion are not perfectly correlated.

We assume the following model for log-dividend growth, d_{t+1}:

$$d_{t+1} = \bar{g} + \gamma_g(\sigma_{cp}\omega_{p,t+1} - \sigma_{cn}\omega_{n,t+1}) \tag{17}$$

We use $\gamma_g = 3.0762$ to match the dividend growth variance, where the monthly dividend growth is computed as the monthly growth of real unsmoothed annual S&P500 dividends.[8] Modelling dividends as levered consumption is popular in the literature with Bollerslev, Tauchen and Zhou (2009) using $\gamma_g = 1$, and Drechsler and Yaron (2011) and Wachter (2013) using $\gamma_g = 3$.

The log-pricing kernel, m_{t+1}, for this model can be written as:

$$m_{t+1} = m_0 + m_q q_t + m_{\omega,p}\omega_{p,t+1} + m_{\omega,n}\omega_{n,t+1}, \tag{18}$$

where:

$$\begin{aligned} m_0 &= \ln\beta - \gamma\bar{g} + \gamma\bar{q}(1-\rho_q), \\ m_q &= -\gamma(1-\rho_q), \\ m_{\omega,p} &= \gamma(\sigma_{qp} - \sigma_{cp}), \\ m_{\omega,n} &= \gamma(\sigma_{qn} + \sigma_{cn}). \end{aligned} \tag{19}$$

In all our estimations we find $m_{\omega,p} < 0$ and $m_{\omega,n} > 0$, that is, good environment shocks decrease marginal utility and bad environment shocks increase marginal utility.

The key formula to derive most of the results is that for a demeaned gamma random variable $X \sim \Gamma(k,\theta) - k\theta$, where k is the shape and θ is the scale parameter, $E(e^X) = e^{-g(\theta)k}$, where the function $g(x)$ is defined as $g(x) = x + \ln(1-x)$. Note that $g(\cdot)$ is always negative. This requires $\theta < 1$. Note that while our fundamental shocks $\omega_{p,t+1}$ and $\omega_{n,t+1}$

[8] An asset pricing fit similar to the one reported below can be obtained with any values of γ_g between 1 and 6.

in (14) formally have $\theta = 1$, in the model they are always premultiplied by σ-coefficients which are less than 1, making the effective scale of these shocks less than 1, because the scale of a gamma distributed variable with unit scale multiplied by σ is σ.

The risk-free rate

The continuously compounded log-risk-free rate, $r_{f,t}$, can be determined the usual way as the negative of the logarithm of the conditional expectation of the exponentiated pricing kernel in (18):

$$r_{f,t} = f_0 + f_q q_t + f_p p_t + f_n n_t, \tag{20}$$

where:

$$\begin{aligned} f_0 &= -\ln\beta + \gamma\bar{g} - \gamma\bar{q}(1-\rho_q), \\ f_q &= \gamma(1-\rho_q), \\ f_p &= g(m_{\omega,p}), \\ f_n &= g(m_{\omega,n}). \end{aligned} \tag{21}$$

Here q_t represents an intertemporal smoothing effect and is positively associated with the short rate ($f_q > 0$), whereas p_t and n_t represent precautionary savings effects and are negatively associated with the short rate ($f_n < 0$ and $f_p < 0$). Note that in our estimation $f_n n_t$ is usually much larger than $f_p p_t$ in magnitude, and therefore the short rate's dependence on n_t is more pronounced than its dependence on p_t.

Equity pricing

The price-dividend ratio can be computed as:

$$\frac{P_t}{D_t} = E_t \sum_{i=1}^{\infty} e^{\sum_{j=1}^{i} m_{t+j} + d_{t+j}}. \tag{22}$$

By recursively plugging the dividend dynamics from (17) and the stochastic discount factor from (18) into (22), we obtain an expression for the price-dividend ratio of the following form:

$$\frac{P_t}{D_t} = \sum_{i=1}^{\infty} e^{A_i + B_i p_t + C_i n_t + D_i q_t}, \tag{23}$$

where the A, B, C, and D coefficients follow difference equations described in Internet Appendix III. With $pd_t = \ln(\frac{P_t}{D_t})$, we log-linearize equation (23) using a Taylor series approximation to find:

$$pd_t \approx K_0^1 + K_p^1 p_t + K_n^1 n_t + K_q^1 q_t, \tag{24}$$

where we again relegate the actual expressions for the K^1-coefficients to Internet Appendix III.

The aggregate market return is defined as:

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} = \frac{D_{t+1}}{D_t} \frac{1 + \frac{P_{t+1}}{D_{t+1}}}{\frac{P_t}{D_t}}, \tag{25}$$

yielding the log-return, r_{t+1}:

$$r_{t+1} = d_{t+1} + \ln(1 + \frac{P_{t+1}}{D_{t+1}}) - pd_t. \tag{26}$$

Analogously to pd_t, we can linearize $\ln(1 + \frac{P_{t+1}}{D_{t+1}})$ via a Taylor series approximation as:

$$\ln(1 + \frac{P_{t+1}}{D_{t+1}}) \approx K_0^2 + K_p^2 p_{t+1} + K_n^2 n_{t+1} + K_q^2 q_{t+1}, \tag{27}$$

where the K^2 coefficients are reported in Internet Appendix III.

By plugging (17), (24), and (27) into (26) we obtain the following expression for the log-return:

$$r_{t+1} \approx r_0 + r_p p_t + r_n n_t + r_q q_t + r_{\omega,p} \omega_{p,t+1} + r_{\omega,n} \omega_{n,t+1}, \tag{28}$$

where the exact expressions for the coefficients r_0, r_p, r_n, $r_{\omega,p}$, and $r_{\omega,n}$ are once again in Internet Appendix III. In all our estimations we find $r_{\omega,p} > 0$ and $r_{\omega,n} < 0$, that is, good environment shock realizations increase equity returns and bad environment shock realizations decrease them.

Combining (20) and (28), we get the following expression for the equity premium:

$$E_t(r_{t+1} - r_{f,t}) \approx (r_0 - f_0) + (r_p - f_p) p_t + (r_n - f_n) n_t + (r_q - f_q) q_t, \tag{29}$$

where in all our estimations $(r_p - f_p) > 0$ and $(r_n - f_n) > 0$. This makes economic sense, because both $\omega_{p,t+1}$ (which conditional variance is p_t) and $\omega_{n,t+1}$ (which conditional variance is n_t) move the pricing kernel in (18) and the aggregate equity return in (28) in opposite directions. In contrast, $(r_q - f_q)$ is 0 up to approximation error, as q_t does not affect the moments of shocks to the pricing kernel.

The variance risk premium

Following most of the extant literature, we first define the variance risk premium in the model as the difference between risk-neutral and physical variances of log-returns. From (28) it follows that the conditional physical variance of the aggregate market return is:

$$Var_t(r_{t+1}) = r_{\omega,p}^2 p_t + r_{\omega,n}^2 n_t. \tag{30}$$

The risk-neutral (Q-measure) variance of the log-return can be computed by evaluating the first and second derivative of the risk-neutral moment-generating function:

$$\begin{aligned} E_t^Q(r_{t+1}) &= \frac{d}{d\nu}\left[\frac{E_t e^{m_{t+1}+\nu r_{t+1}}}{E_t e^{m_{t+1}}}\right]_{\nu=0}, \\ E_t^Q(r_{t+1}^2) &= \frac{d^2}{d\nu^2}\left[\frac{E_t e^{m_{t+1}+\nu r_{t+1}}}{E_t e^{m_{t+1}}}\right]_{\nu=0}. \end{aligned} \tag{31}$$

Plugging m_{t+1} from (18) and r_{t+1} from (28) into (31) results in:

$$Var_t^Q(r_{t+1}) = (\frac{r_{\omega,p}}{1 - m_{\omega,p}})^2 p_t + (\frac{r_{\omega,n}}{1 - m_{\omega,n}})^2 n_t. \tag{32}$$

Bringing (30) and (32) together, we obtain the expression for the variance risk premium:

$$Var_t^Q(r_{t+1}) - Var_t(r_{t+1}) = r_{\omega,p}^2 \left[\frac{1}{(1 - m_{\omega,p})^2} - 1\right] p_t + r_{\omega,n}^2 \left[\frac{1}{(1 - m_{\omega,n})^2} - 1\right] n_t. \tag{33}$$

The premium in (33) is affected by both bad and good uncertainty. However, as we argued before, good environment shocks decrease the pricing kernel (marginal utility), that is $m_{\omega_p} < 0$. Therefore, the coefficient on p_t is negative. Analogously, bad environment shocks increase the pricing kernel (marginal utility), that is $m_{\omega_n} > 0$. Therefore, bad uncertainty has a larger effect on the risk neutral than on the physical volatility and the reverse is true for good uncertainty, consistent with the intuition that risk neutral pricing shifts mass to high marginal utility states. As a result, bad (good) uncertainty shocks increase (decrease) the variance risk premium.

VIX^2 and $SVIX^2$ studied by Martin (2017) are, in fact, also available in closed form in the model. Relegating the derivation to Internet Appendix III, we find:

$$SVIX_t^2 = Var_t^Q(\frac{R_{t+1}}{R_{f,t}}) = exp(m_0 + 2r_0 - f_0 + [m_q + 2r_q - f_q]q_t + \\ [2r_p - g(m_{\omega,p} + 2r_{\omega,p}) - f_p]p_t + [2r_n - g(m_{\omega,n} + 2r_{\omega,n}) - f_n]n_t) - 1. \quad (34)$$

While the expression looks quite different from (32), we show below that the actual values are very close in our estimation. Next we compute VIX^2 as the risk neutral entropy:

$$VIX^2 = 2 \cdot [\ln E_t^Q(\frac{R_{t+1}}{R_{f,t}}) - E_t^Q(\ln \frac{R_{t+1}}{R_{f,t}})] = 2 \cdot [\ln E_t^Q(R_{t+1}) - E_t^Q(\ln R_{t+1})] = \\ 2 \cdot [r_{f,t} - R_{f,t}E_t(M_{t+1} \ln R_{t+1})] = 2 \cdot [r_{f,t} - exp(r_{f,t})E_t(exp(m_{t+1})r_{t+1})]. \quad (35)$$

Obtaining a closed-form expression for (35), requires computing the expectation of $X \cdot \exp(X)$, where $X \sim \Gamma(k, \theta)$ with k being the shape and θ the scale parameters, respectively. Assuming that $\theta < 1$, which is always satisfied in our estimation, Internet Appendix III demonstrates that:

$$E[X \cdot \exp(X)] = \frac{k\theta}{(1-\theta)^{k+1}}, \quad (36)$$

which, as Internet Appendix III shows, leads to:

$$VIX_t^2 = \\ 2 \cdot [r_{f,t} - e^{r_{f,t}+m_0+m_q q_t} \cdot \{e^{-g(m_{\omega,p})p_t - g(m_{\omega,n})n_t} \cdot (r_0 + r_p p_t + r_n n_t + r_q q_t) + \\ e^{-g(m_{\omega,n})n_t} r_{\omega,p} \cdot (-p_t e^{-g(m_{\omega,p})p_t} + e^{-m_{\omega,p}p_t} \frac{p_t}{(1-m_{\omega,p})^{p_t+1}}) + \\ e^{-g(m_{\omega,p})p_t} r_{\omega,n} \cdot (-n_t e^{-g(m_{\omega,n})n_t} + e^{-m_{\omega,n}n_t} \frac{n_t}{(1-m_{\omega,n})^{n_t+1}})\}]. \quad (37)$$

The last part of the expression looks reminiscent of (32).

4.2 Estimation and Empirical Fit

The model is estimated using the classical minimum distance methodology (Wooldridge, 2002). We match unconditional moments of monthly US real consumption growth of non-durables and services and the financial series used before for the 1990-2017 period. The following unconditional moments are used in the classical minimum distance estimation: consumption growth mean, consumption growth standard deviation, consumption growth skewness (scaled), consumption growth excess kurtosis (scaled), real risk-free rate mean, real risk-free rate standard deviation, the first lag autocorrelation of the real risk-free rate, the average equity premium, physical standard deviation of equity returns, the log-price-dividend ratio mean, standard deviation of the log-price-dividend ratio, the first lag autocorrelation of the log-price-dividend ratio, the mean of the variance risk premium (defined as the risk-neutral variance minus the physical variance of the log-return in the model and VIX^2 minus the physical variance from a Corsi (2009)-type model described above in the data), the standard deviation of the variance risk premium, the difference between the risk-neutral entropy (VIX^2) and the risk-neutral variance ($SVIX^2$). To simplify the estimation, we set $\beta = 1.00$ and $p_t = \bar{p}$. We also set $\bar{q} = 1$, as it is not identified due to always entering asset pricing equations multiplied by γ. We use a diagonal weighting matrix to avoid collinearity issues and achieve a more balanced moment fit. The weighting matrix is the diagonal of the inverse of the covariance matrix estimated from re-sampling 10,000 time series of the historical length with a block length of 60 months.

The first column of Table 7 shows the parameter estimates. In terms of preferences, γ is close to 2.0, but of course this coefficient no longer represents risk aversion. The q_t process is very persistent, as is typical in habit models; ρ_q is the main determinant of the risk-free rate and price-dividend ratio autocorrelation. As expected, $\omega_{p,t}$ shocks decrease risk aversion, that is $\sigma_{qp} < 0$, whereas $\omega_{n,t}$ shocks increase risk aversion, that is $\sigma_{qn} > 0$. The latter coefficient is of a much larger magnitude than the former. Unconditionally, the bad environment process only accounts for 12.3% of the consumption growth variance ($\frac{\sigma_{cn}^2 \bar{n}}{\sigma_{cp}^2 \bar{p} + \sigma_{cn}^2 \bar{n}}$), but, because p_t is constant, n_t drives all of its time variation. Therefore, this ratio increases substantially during recessions. Given that $\bar{p}$, the unconditional shape parameter of the good environment shock, is greater than 10, it follows from the properties of the gamma distribution that good environment shocks are essentially Gaussian. On the contrary, as $\bar{n}$ is less than 1, it follows from the properties of the gamma distribution that bad environment shocks are very skewed and non-Gaussian. The n_t process is highly persistent.

Table 8 shows how the model fits the data moments. On the right hand side, we show the data moments and the corresponding standard errors. While the χ^2 test for the overall fit rejects the model at the 5% significance level, the overall fit is good. In fact, the model moments are always within a two standard error band around the data moments. The moments driving the statistical rejection, such as the mean consumption growth of 0.17% in the model versus 0.20% in the data, have arguably economically sensible values in the model.

Taking a closer look at the various moments, the model does generate an equity premium that is economically smaller than the one observed in the data (but falls within one standard error around it). This is mainly driven by the large data standard error associated with the equity premium, which results in a relatively low weight for this moment in the estimation. Importantly, the model fits the both the mean the variance of the variance risk premium. It also fits the Martin's (2017) bound. The model provides the fit while being entirely consistent with the consumption growth moments. These moments include a small positive unconditional skewness coefficient and excess kurtosis of around 2.0. Recall that the n_t process is very skewed and leptokurtic but unconditionally this process only accounts for a modest fraction of the total consumption variance and helps fit the data during recessions, especially the Great Recession.

To further test the model, we consider two moments not used in the estimation: consumption growth percentile shifts conditional on the variance risk premium and the autocorrelation of the variance risk premium. Table 9 shows the consumption growth percentile shifts conditional on the variance risk premium. The model generates the downward shift in the 10^{th} percentile when the variance risk premium is high, while showing no other shifts, as is true in the data. However, the shift in the 10^{th} percentile is smaller than its data counterpart, although still falling inside two standard deviations from it.

Table 10 shows that the model implied persistence of the variance risk premium is much higher than in the data. The difference is both statistically and economically large.

Tables 7-10 also report the estimation of a model (the "full" model) with a more intricate consumption and dividend growth process:

$$
\begin{aligned}
g_{t+1} &= \bar{g} + \phi_g(n_t - \bar{n}) + \sigma_{cp}\omega_{p,t+1} - \sigma_{cn}\omega_{n,t+1}, \\
d_{t+1} &= \bar{g} + \phi_d(n_t - \bar{n}) + \gamma_g(\sigma_{cp}\omega_{p,t+1} - \sigma_{cn}\omega_{n,t+1},) + \gamma_n(-\sigma_{cn}\omega_{n,t+1}).
\end{aligned} \tag{38}
$$

First, the conditional means of consumption and dividend growth now vary with n_t, as in Segal, Shaliastovich, and Yaron (2015)'s application of the BEGE model of Bekaert and Engstrom (2017) and Bekaert, Engstrom, and Ermolov (2015) to a long-run risk model. Second, apart from the usual leverage term, the dividend growth shock has separate exposure to $\omega_{n,t+1}$, so that it can have different loadings on good and bad consumption growth shocks. We use the dividend growth standard deviation, skewness, and kurtosis as additional moments in the CMD estimation to aid the identification of γ_g and γ_n.

Table 7 shows that the overlapping parameters in the base and full models are almost identical. The conditional means of dividend and consumption growth depend negatively on n_t, in line with the evidence in Segal, Shaliastovich, and Yaron (2015), but only the consumption growth dependence is statistically significant. In contrast, γ_n is borderline significantly negative, so that dividend shocks load heavily on $\omega_{n,t}$ making cash flows riskier than in the base model. The usual leverage coefficient (γ_g) is close to 3.00, as it was in the base model.

Table 8 shows that the fit with the data moments is again very good. With one exception (the standard deviation of equity returns), all moments are within one standard error of their data counterparts. The model generates slight negative skewness for both consumption and dividend growth. Not surprisingly, the test of the over-identifying restrictions now does not reject the model at the 10% level. The equity premium is now 8 basis points higher (almost 1% annualized) than in the base model.

Table 9 reveals that the full model perfectly fits the downward shift in the 10^{th} consumption growth percentile going from low to high variance risk premiums. However, the full model also generates downward shifts at the 50^{th} and 90^{th} percentiles even though these are smaller than the negative tail shift, and still within two standard errors of the shifts observed in the data. This shift is a direct consequence of the dependence of the mean of consumption growth on n_t. Table 10 shows that the full model does not resolve the failure to fit the variance risk premium persistence.

4.3 Fitting the Variance Risk Premium Persistence

The model's failure in fitting the low persistence of the variance risk premium is easily understood. Both state variables driving asset price dynamics (n_t and q_t) are highly persistent. The variance risk premium computed using log returns in (33) in fact only depends on n_t and thus inherits its persistence.

We now investigate if the model is flexible enough to match this moment by explicitly including it into the classical minimum distance estimation. To conserve space, we relegate all the empirical results to Internet Appendix IV, offering a brief discussion of the results here. The parameters are largely the same as before, except for the autoregressive coefficient for n_t, which drops from over 0.99 to less than 0.6, and its innovation standard deviation, which doubles.

Both the base and full models now replicate the low persistence of the variance risk premium, delivering an autocorrelation of about 0.58 versus 0.52 in the data. While the models overall are statistically rejected at the 1% significance level, economically the overall fit is rather adequate. For example, while the standard deviation of the log-price-dividend ratio in the model is above 2 standard deviations of the data counterpart, the model implied value of about 0.42 would be consistent with the data if the payout would be computed as in Boudoukh et al. (2007) and would be lower than the data if the payout would be computed as in Longstaff and Piazzesi (2004). Overall, of the 16 (19) moments used in the estimation for the base (full) model, only four moments are outside a two standard error band around the data moments.

Economically, the most serious violation is that the model-implied variance of the variance risk premium is very low compared to the data (0.0004 in the model versus 0.0019 in the data). Because the variance risk premium is a linear function of n_t, the variance and persistence of the variance risk premium is increasing in the variance and persistence of n_t. Note from (15) that the unconditional variance of n_t is $\frac{\sigma_{nn}^2 \bar{n}}{1-\rho_n^2}$. Decreasing ρ_n, which is required to decrease the persistence of the variance risk premium, also decreases the unconditional variance of n_t and consequently the variance of the variance risk premium. This decrease cannot be offset by changes in other parameters, which are tied to the consumption growth and asset pricing moments.

The model again matches the left tail consumption growth percentile shifts conditional on the variance risk premium observed in the data, although these moments are not used in the estimation. In fact, the quantile shifts generated by the new models are very similar to the ones generated under the previous model specifications.

4.4 Model with a Preference Shock

Given that the macroeconomic model in the previous subsection is, despite an adequate economic fit, strongly statistically rejected, we now introduce a pure preference shock into the model in order to fit the variance risk premium persistence. There are at least two requirements for the preference shock. First, it needs to be non-Gaussian, because, as we have shown, a Gaussian shock does not affect the variance risk premium. Second, the preference shock variance should vary through time and have relatively low persistence, because the variance risk premium persistence is low in the data. This follows from the variance risk premium being a linear function of model's shocks variances (see equations (30) and (32)). Fortunately, such rapidly mean reverting risk aversion is consistent with direct and indirect evidence in the recent asset pricing literature (see Martin, 2017; Bekaert, Engstrom, and Xu, 2019).

We keep the macroeconomic dynamics as in equation (13) and introduce a new preference shock, $\omega_{q,t+1}$, into the log-inverse consumption surplus ratio equation (16):

$$q_{t+1} = \bar{q} + \rho_q(q_t - \bar{q}) + \sigma_{qp}\omega_{p,t+1} + \sigma_{qn}\omega_{n,t+1} + \sigma_{qq}\omega_{q,t+1}. \tag{39}$$

In line with the rest of our model, $\omega_{q,t+1}$ follows a demeaned gamma distribution:

$$\begin{aligned} &\omega_{q,t+1} \sim \Gamma(s_t, 1) - s_t, \\ &s_{t+1} = \bar{s} + \rho_s(s_t - \bar{s}) + \sigma_{sq}\omega_{q,t+1}. \end{aligned} \tag{40}$$

The model is solved in closed form exactly as before except that there is one more state variable, s_t. We again estimate the model, using the more intricate fundamentals process of the full model, via the classical minimum distance methodology, but also explicitly include the variance risk premium autocorrelation at lag 1 as a moment to match.[9]

Table 11 shows the estimated model parameters. The pure preference shock is strongly non-Gaussian ($\bar{s}$ is clearly less than 1), which is required for the shock to have a notable

[9]We did experiment with a model where q_t is also the shape parameter of the $\omega_{q,t+1}$ shock, limiting the number of state variables to 2. However, in such estimations q_t invariably hits its lower boundary of zero.

impact on the variance risk premium. Its persistence is low ($\rho_s = 0.5912$), which helps to fit the low variance risk premium persistence. The parameters imply that unconditionally 15.83% of the q_t process, which also measures stochastic risk aversion, is driven by a "sentiment" shock not correlated with fundamentals. While many of the other parameters are similar to the ones obtained for the full model before (e.g. q_t and n_t are still highly persistent), there are also some distinct changes. For example, the q_t process loads less heavily on $\omega_{n,t}$, but the n_t process is now estimated to be less non-Gaussian with much higher innovation variances. The dependence of the consumption process on n_t (p_t) has also decreased (increased). This is logical given that the preference shock now helps account partly for asset price variation. However, the dividend growth process now loads more heavily on n_t than before, making cash flows more exposed to downside risk.

Table 12 shows that, while the model is statistically rejected at the 1% significance level, it fits the various moments rather well: the vast majority of the model-implied moments is within one standard deviation of the data counterparts. There is not a single moment outside a two standard deviation band around the data counterpart.

Table 13 indicates that the model matches the left tail consumption growth percentile shifts conditional on the variance risk premium observed in the data, although these moments are once again not used in the estimation. The model-implied shift is now slightly larger than the value observed in the data, which was not the case before. Note that the model also generates a slight negative shift in right tail consumption growth percentiles following a large variance risk premium, although the shift is much smaller than for the left tail consumption growth percentiles. While the data do not generate such a shift, the model shift is within one standard error of what is observed empirically. In sum, the model with a preference shock fits all the salient features of the data.

5 Excess Return Predictability

It is well-known that the variance risk premium predicts stock returns (see Bollerslev, Tauchen, and Zhou, 2009, for the seminal work in this area). In this section, we verify whether the various models can replicate the real world predictability. We do so using a simple univariate regression of future excess stock returns on the variance risk premium. We also go one step further, and test whether the models can match the evidence uncovered in Bekaert and Hoerova (2014). They regress stock returns onto the variance risk premium and the conditional (physical) variance of stock returns. They find that the variance risk premium predicts stock returns with a positive sign, but the stock market variance predicts returns with a negative sign. Significant predictability is present at the one month and three month horizons. This result will be challenging to replicate in many of the models as they lack on obvious channel for uncertainty to negatively predict stock returns. Moreover, it implies that volatility dynamics in the model must be driven by at least two lowly correlated state variables.

Table 14 shows the empirical result for our particular sample for one- and three-month forecasting horizons. The variance risk premium predicts returns more strongly both economically and statistically at the three-month horizon, but the regression coefficient is still significant at the 10% level at the one-month horizon. The evidence regarding conditional physical uncertainty is weaker in this sample. The coefficient is negative at both horizons but not statistically significant.[10] The addition of the uncertainty variable does not alter the variance risk premium coefficient or its significance.

Table 14 also shows the regression coefficients produced by various models. The BTZ model produces the wrong sign both for the variance risk premium and physical uncertainty. This occurs, because as shown in Internet Appendix I, in log-normal models the variables which increase the variance risk premium decrease the equity premium. The DY model moderately and Bekaert and Engstrom (2017) model strongly overshoot the predictive power of the variance premium for stock returns. Note that in the DY model physical and risk-neutral variances are perfectly correlated, as both are linear in σ_t^2, rendering the multivariate regression undefined. The same is true for our base and full models, as the variance risk premium is linear in n_t and thus they generate perfect correlation between the variance risk premium and the conditional variance of stock returns. The preference shock model has two state variables, so that it can potentially replicate the bivariate evidence. The new BEGE models generate regression coefficients that are within a two standard error band of the data values for all coefficients considered but cannot generate a negative value for the physical conditional variance coefficient. The rare disaster model does generate such negative coefficients and also generates coefficients within a two standard error band around all data coefficients. It tends to overshoot (undershoot) the conditional variance (variance risk premium) coefficient.

[10] The t-statistics are higher in absolute magnitude for some alternative measurements of the conditional variance, but it appears that the significance is weakened by the addition of more data after 2010, the end of the Bekaert-Hoerova sample.

In sum, the predictability evidence is not favorable to the extant models, except for the disaster models. The new BEGE model introduced in this article can match the evidence in a statistical fashion. However, as most equilibrium models do, it implies a positive risk-return trade-off with physical risk, whereas this trade-off in the data is weak and maybe even negative. Lochstoer and Muir (2019) explain such evidence using a model where the representative agent underreacts (and ultimately overreacts) to volatility news. The predictability mechanism in our model is consistent with the empirical evidence in Kilic and Shaliastovich (2019), as the predictability is driven by the "bad" uncertainty state variable.

6 Conclusion

In this article, we use properties of the variance risk premium, the premium for selling variance risk in the equity market, to discipline and refute existing consumption-based asset pricing models. The main result from our exercise is that extant models have a hard time matching simultaneously even simple features of asset returns such as the equity premium with features of options prices as reflected in the properties of the variance risk premium, and, importantly, the link between consumption growth and option prices. We therefore introduce a new and tractable model that does fit these facts. The model features a "BEGE" structure with "good" and "bad" volatility driving consumption growth and risk aversion shocks, with bad (good) volatility decreasing (increasing) unscaled skewness of these shocks. The model fits the data even better when a small sentiment shock is allowed.

Bollerslev, Tauchen and Zhou (2009) show that the equity premium and variance premium are correlated. As an "out-of-sample" experiment, we verify how the various models match the evidence on the predictability of the variance risk premium for stock returns. The new BEGE model and Wachter's (2013) model are the only models that can fully match the predictability evidence.

While we examine models representing the various main asset paradigms, research is continually evolving and alternative models may likewise be successful. For instance, Schreindorfer (2019) combines a non-linear data-generating process for consumption growth (a special case of the BEGE fundamentals), with generalized disappointment aversion preferences as in Routledge and Zin (2010) and is able to match several of the data features we consider. However, Drechsler (2018) shows that disappointment aversion preferences imply counterfactual option price dynamics once the disappointment threshold is passed. Our model is not subject to this critique. Schreindorfer (2019) also matches a decomposition of the equity premium in components driven by tail risks and normal risks stressed by Bollerslev and Todorov (2011). Beason and Schreindorfer (2019) show that extant representative agent models fare poorly in matching this evidence but the published BEGE paper by Bekaert and Engstrom (2017) at least matches part of it. We plan to verify how our new BEGE model fares with respect to this evidence in future work.

There are alternative empirical approaches to discipline and refute models that may prove useful for further testing of the new BEGE model. For example, Zviadadze (2018) shows that the term structure of risk in expected returns and cash flow growth is challenging to a number of standard equilibrium models (including disaster risk and long-run risk models). She assigns a large role to the dynamics of consumption variances in matching stock return dynamics. Because time-varying consumption variances play a large role in the BEGE model we introduce, even in the variant with preference shocks, it has potential to be consistent with her findings.

References

Ait-Sahalia, Yacine, Karaman, Mustafa, and Loriano Mancini, 2018, The Term Structure of Variance Swaps and Risk Premia, *Working paper.*

Backus, David, Chernov, Mikhail, and Ian Martin, 2011, Disasters Implied by Equity Index Options, *Journal of Finance*, 66, pp. 1969-2012.

Bakshi, Gurdip, Kapadia, Nikunj, and Dilip Madan, 2003, Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options, *Review of Financial Studies*, 16, pp. 101-143.

Bakshi, Gurdip, and Dilip Madan, 2000, Spanning and Derivative-security Valuation, *Journal of Financial Economics*, 55, pp. 205-238.

Bansal, Ravi, and Amir Yaron, 2004, Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles, *Journal of Finance*, 59, pp. 1481-1509.

Barro, Robert, and Jose Ursua, 2008, Consumption Disasters in the Twentieth Century, *American Economic Review*, 98, pp. 58-63.

Beason, Tyler, and David Schreindorfer, 2019, On Sources of Risk Premia in Representative Agent Models, *Working paper.*

Bekaert, Geert, Engstrom, Eric, and Steven Grenadier, 2010, Stock and Bond Returns with Moody Investors, *Journal of Empirical Finance*, 17, pp. 867-894.

Bekaert, Geert, and Eric Engstrom, 2017, Asset Return Dynamics under Habits and Bad Environment-Good Environment Fundamentals, *Journal of Political Economy*, 125, pp. 713-760.

Bekaert, Geert, Engstrom, Eric, and Andrey Ermolov, 2015, Bad Environments, Good Environments: A Non-Gaussian Asymmetric Volatility Model, *Journal of Econometrics*, 186, pp. 258-275.

Bekaert, Geert, Engstrom, Eric, and Nancy Xu, 2019, The Time Variation in Risk Appetite and Uncertainty, *Working paper.*

Bekaert, Geert, and Marie Hoerova, 2014, The VIX, the Variance Premium and Stock Market Volatility, *Journal of Econometrics* 183, pp. 181-192.

Bollerslev, Tim, Tauchen, George, and Hao Zhou, 2009, Expected Stock Returns and Variance Risk Premia, *Review of Financial Studies*, 22, pp. 4463-4492.

Bollerslev,Tim, and Viktor Todorov, 2011, Tails, Fears, and Risk Premia, *Journal of Finance*, 66, pp. 2165-2211.

Boudoukh, Jacob, Michaely, Robert, Richardson, Matthew, and Michael Roberts, 2007, On the Importance of Measuring Payout Yield: Implications for Empirical Asset Pricing, *Journal of Finance*, 62, pp. 877-915.

Britten-Jones, Mark, and Anthony Neuberger, 2000, Option Prices, Implied Price Processes, and Stochastic Volatility, *Journal of Finance*, 55, pp. 839-866.

Campbell, John, and John Cochrane, 1999, By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior, *Journal of Political Economy*, 107, pp. 205-221.

Cheng, Ing-Haw, 2019, The VIX Premium, *Review of Financial Studies*, 32, pp. 180-227.

Corsi, Fulvio, 2009, A Simple Approximate Long-Memory Model of Realized Volatility, *Journal of Financial Econometrics*, 7, pp. 174-196.

Drechsler, Itamar, and Amir Yaron, 2011, What's Vol Got to Do with It, *Review of Financial Studies*, 24, pp. 1-45.

Drechsler, Itamar, 2018, Discussion of Macroeconomic Tail Risks and Asset Prices by David Schreindorfer, *Society of Financial Studies Cavalcade*, available at `https://sites.google.com/site/idrechsl/`.

Du, Du, 2011, General Equilibrium Pricing of Options with Habit Formation and Event Risks, *Journal of Financial Economics*, 99, pp. 400-426.

Epstein, Larry, and Stanley Zin, 1989,Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework, *Econometrica*, 57, pp. 937-969.

Gabaix, Xavier, 2012, Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance, *Quarterly Journal of Economics*, 127, pp. 645-700.

Glosten, Lawrence, Jagannathan, Ravi, and David Runkle, 1993, On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, *Journal of Finance*, 48, pp. 1779-1801.

Jiang, George, and Yisong Tian, 2005, The Model-Free Implied Volatility and Its Information Content, *Review of Financial Studies*, 18, pp. 1305-1342.

Kilic, Mete, and Ivan Shaliastovich, 2019, Good and Bad Variance Premia and Expected Returns, *Management Science*, 65, pp. 2445-2945.

Lochstoer, Lars, and Tyler, Muir, 2019, Volatility Expectations and Returns, *Working paper*.

Longstaff, Francis, and Monika Piazzesi, 2004, Corporate Earnings and the Equity Premium, *Journal of Financial Economics*, 74, pp. 401-421.

Lorenz, Friedrich, Schmedders, Karl, and Matle Schumacher, 2020, Nonlinear Dynamics in Conditional Volatility, *Working paper*.

Martin, Ian, 2017, What is the Expected Return on the Market?, *Quarterly Journal of Economics*, 132, pp. 367-433.

Pohl, Walter, Schmedders, Karl, and Ole Wilms, 2018, Higher Order Effects in Asset Pricing Models with Long-Run Risks, *Journal of Finance*, 73, pp. 1061-1111.

Rietz, Thomas, 1988, The Equity Risk Premium A Solution, *Journal of Monetary Economics*, 22, pp. 117-131.

Routledge, Bryan, and Stanley Zin, 2010, Generalized Disappointment Aversion and Asset Prices, *Journal of Finance*, 65, pp. 1303-1332.

Schreindorfer, David, 2019, Macroeconomic Tail Risks and Asset Pricies, *Review of Financial Studies*, forthcoming.

Wachter, Jessica, 2005, Solving Models with External Habit, *Finance Research Letters*, 2, pp. 210-226.

Wachter, Jessica, 2013, Can Time-varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?, *Journal of Finance*, 68, pp. 987-1035.

Wachter, Jessica, and Sang Byung Seo, 2019, Option Prices in a Model With Stochastic Disaster Risk, *Management Science*, 65, pp. 3449-3469.

Whaley, Robert, 2000, The Investor Fear Gauge, *The Journal of Portfolio Management*, 26, pp. 12-17.

Wooldridge, 2002, Econometric Analysis of Cross Section and Panel Data, MIT Press, pp. 445-446.

Zviadadze, Irina, 2018, Term Structure of Risk in Expected Returns, *Working paper*.

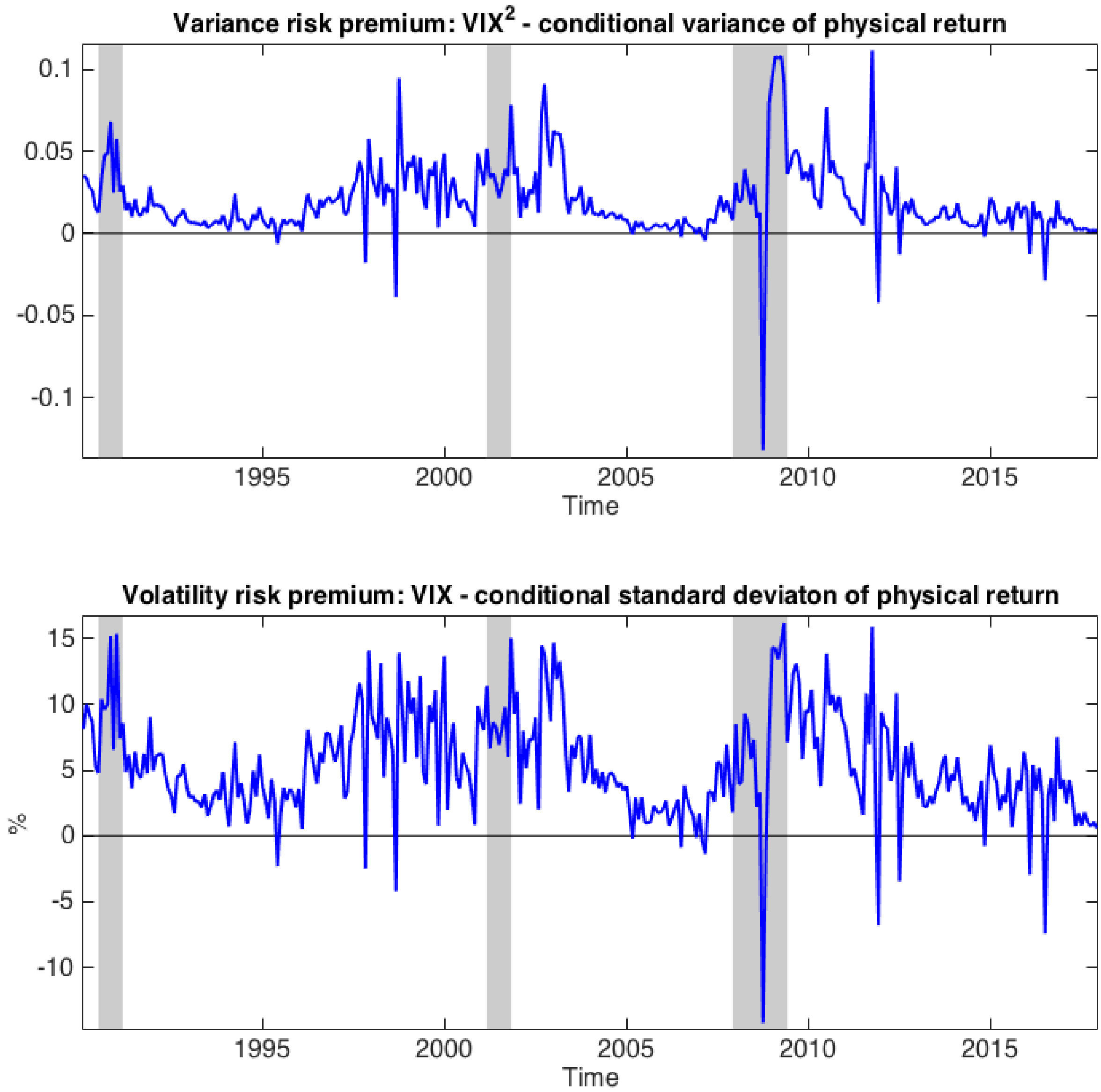

Figure 1 – Variance and Volatility Risk Premia. Data is monthly from January 1990 to December 2017. The variance risk premium is the difference between annualized VIX^2 and expected physical variance of the $S\&P500$ return from Corsi (2009)-type model. The volatility risk premium is the difference between annualized VIX and expected physical standard deviation of the $S\&P500$ return from Corsi (2009)-type model. NBER recessions are shaded.

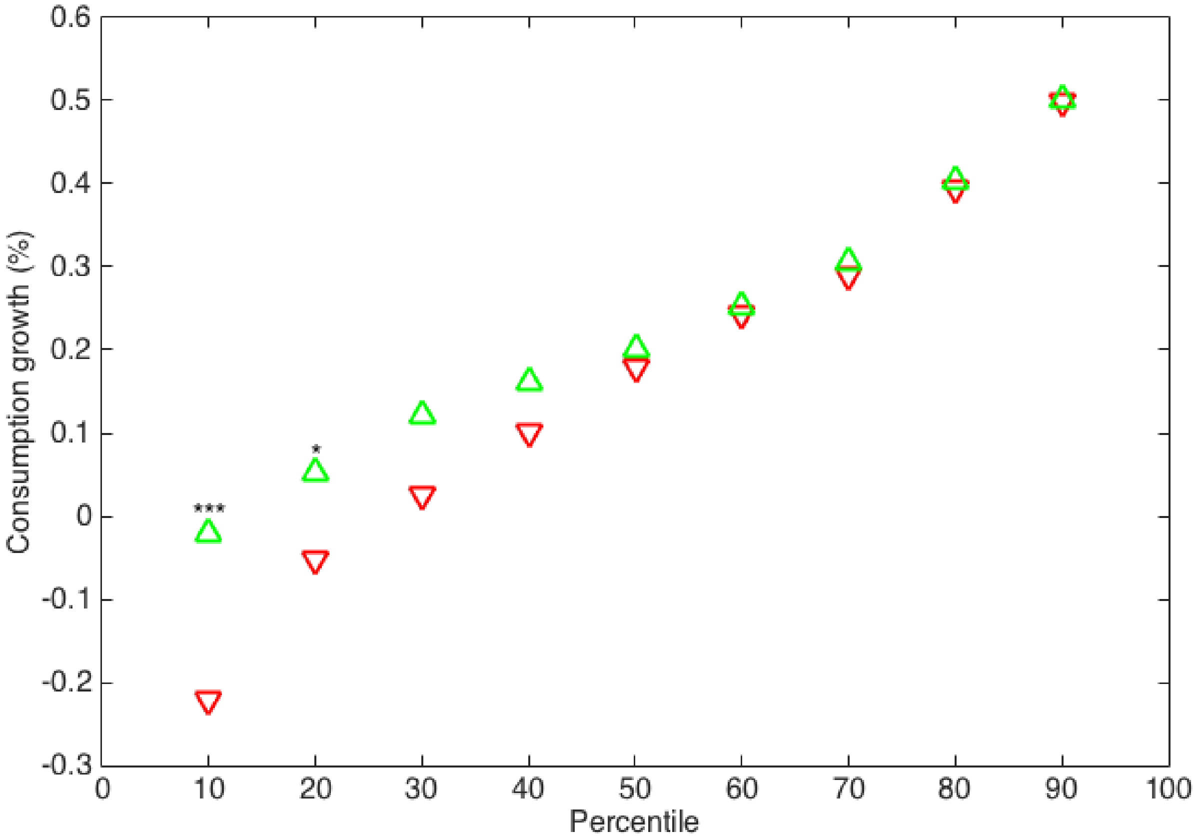

Figure 2 – Percentiles of Next Month Consumption Growth Conditional on the Current Variance Risk Premium in the US data. Data is monthly from January 1990 to December 2017. Green up-pointing triangles correspond to quantiles conditional on the low variance risk premium, and red down-pointing triangles correspond to quantiles conditional on the high variance risk premium. The variance risk premium is the difference between VIX^2 and expected physical variance of the market return obtained from a Corsi (2009)-type model. High variance risk premium is defined as the variance risk premium above 80^{th} unconditional percentile in the data and low variance risk premium as the variance risk premium below 20^{th} unconditional percentile in the data. *** and * correspond to statistical significance at the 1% and 10% levels, respectively. Statistical significance is determined based on block-bootstrap standard errors computed by re-sampling 10,000 time series of historical length with the block length of 60 months.

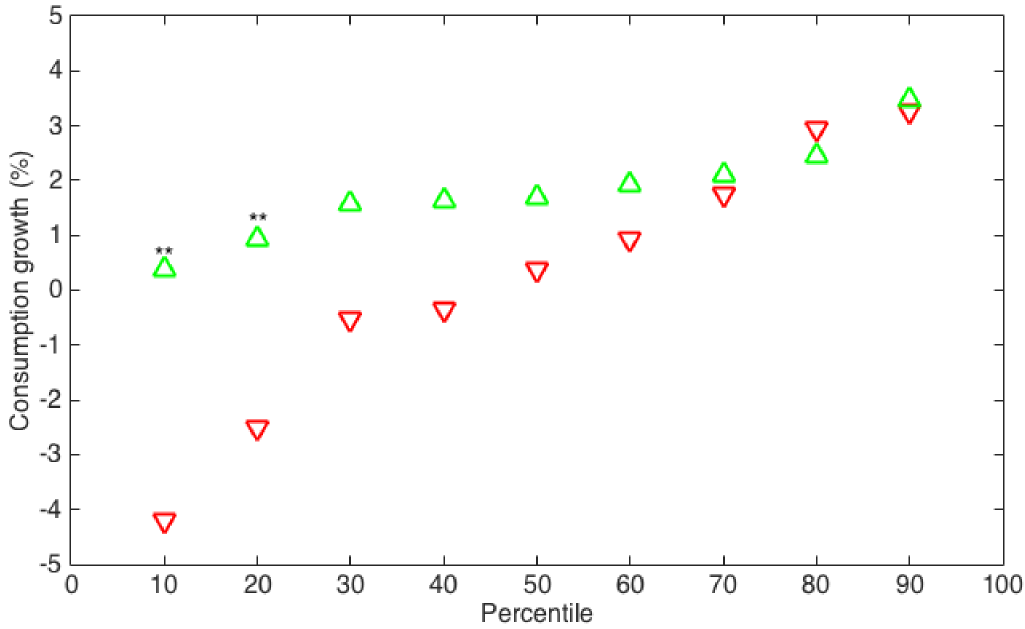

Figure 3 – Percentiles of Next Year Consumption Growth Conditional on the Current Variance Risk Premium in the US data. Data is annually from 1929 to 2017. Green up-pointing triangles correspond to quantiles conditional on the low variance risk premium, and red down-pointing triangles correspond to quantiles conditional on the high variance risk premium. High variance risk premium is defined as the variance risk premium above 80^{th} unconditional percentile in the data and low variance risk premium as the variance risk premium below 20^{th} unconditional percentile in the data. ** corresponds to statistical significance at the 5% level, respectively. Statistical significance is determined based on block-bootstrap computed by re-sampling 10,000 time series of historical length with the block length of 5 years.

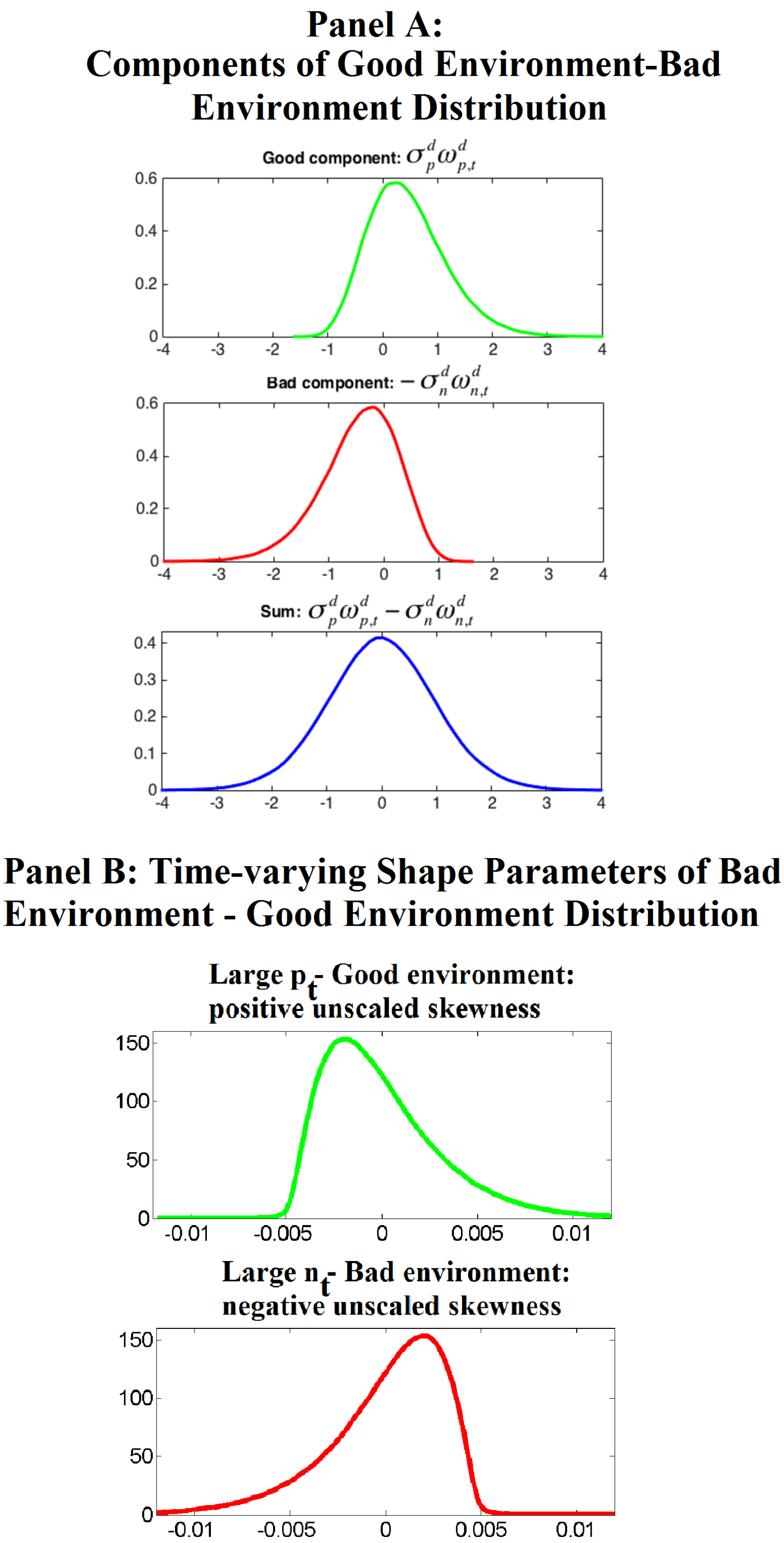

Figure 4 – Bad Environment - Good Environment Distribution. Graphs are probability density functions.

Table 1 – Percentiles of Next Month Consumption Growth Conditional on the Current Variance Risk Premium in US data. Data is monthly from January 1990 to December 2017. The variance risk premium is the difference between VIX^2 and expected physical variance of the market return obtained from a Corsi (2009)-type model. High variance risk premium is defined as the variance risk premium above 80^{th} unconditional percentile in the data and low variance risk premium as the variance risk premium below 20^{th} unconditional percentile in the data. Standard errors in parentheses are block-bootstrap standard errors computed from 10,000 re-samples of historical length with a block length of 60 months. *** corresponds to statistical significance at the 1% level.

	High variance risk premium	Low variance risk premium	High-Low Difference
10^{th} percentile	-0.23%	-0.02%	-0.21%***
	(0.07%)	(0.03%)	(0.07%)
50^{th} percentile	0.18%	0.20%	-0.02%
	(0.06%)	(0.03%)	(0.05%)
90^{th} percentile	0.50%	0.50%	0.00%
	(0.07%)	(0.03%)	(0.07%)

Table 2 – Percentiles of Next Month Consumption Growth Conditional on the Current Variance Risk Premium in US data: Alternative Specifications. Data is monthly from January 1990 to December 2017. In Panels A and B the variance risk premium is the difference between VIX^2 and expected physical variance from a Corsi (2009)-type model. In Panel C the variance risk premium is computed using conditional physical variance from GJR-GARCH model of Glosten, Jagannathan and Runkle (1993). In Panel C, high variance risk premium is defined as the variance risk premium above 80^{th} unconditional percentile in the data and low variance risk premium as the variance risk premium below 20^{th} unconditional percentile in the data. Standard errors in parentheses are block-bootstrap standard errors computed from 10,000 re-samples of historical length with a block length of 60 months. *** and ** correspond to statistical significance at the 1% and 5% levels, respectively.

Panel A: High and low variance risk premia are defined as 75^{th} and 25^{th} unconditional percentiles, respectively			
	High variance risk premium	Low variance risk premium	High-Low Difference
10^{th} percentile	-0.25%	-0.04%	-0.20%***
	(0.06%)	(0.03%)	(0.06%)
50^{th} percentile	0.18%	0.20%	-0.02%
	(0.05%)	(0.02%)	(0.05%)
90^{th} percentile	0.51%	0.50%	0.00%
	(0.06%)	(0.02%)	(0.06%)
Panel B: High and low variance risk premia are defined as 85^{th} and 15^{th} unconditional percentiles, respectively			
	High variance risk premium	Low variance risk premium	High-Low Difference
10^{th} percentile	-0.26%	-0.01%	-0.24%***
	(0.07%)	(0.03%)	(0.08%)
50^{th} percentile	0.15%	0.21%	-0.06%
	(0.07%)	(0.04%)	(0.06%)
90^{th} percentile	0.45%	0.50%	-0.05%
	(0.09%)	(0.04%)	(0.08%)
Panel C: Variance risk premium is defined as VIX-GJR-GARCH conditional variance			
	High variance risk premium	Low variance risk premium	High-Low Difference
10^{th} percentile	-0.22%	-0.04%	-0.18%**
	(0.07%)	(0.05%)	(0.09%)
50^{th} percentile	0.19%	0.20%	-0.01%
	(0.08%)	(0.03%)	(0.09%)
90^{th} percentile	0.51%	0.50%	0.01%
	(0.09%)	(0.04%)	(0.10%)
Panel D: Consumption growth conditional on VIX^2 alone			
	High VIX^2	Low VIX^2	High-Low Difference
10^{th} percentile	-0.25%	-0.05%	-0.20%**
	(0.07%)	(0.04%)	(0.08%)
50^{th} percentile	0.17%	0.21%	-0.04%
	(0.08%)	(0.03%)	(0.08%)
90^{th} percentile	0.49%	0.50%	-0.01%
	(0.09%)	(0.04%)	(0.09%)

Table 3 – General Asset Pricing Fit. Values are unconditional moments. Moments are annualized monthly values. In data, the aggregate stock market is approximated by S&P500 index. Real risk-free rate is approximated as the difference between monthly nominal risk-free rate from Ibbotson Associates and the monthly counterpart of Survey of Professional Forecasters expected inflation for the corresponding quarter. Block-bootstrap standard errors in parentheses are obtained by re-sampling 10,000 time series of the historical length with a block length of 60 months. ** and *** correspond to the rejection of the model at the 5% and 1% significance levels, respectively. BTZ refers to Bollerslev, Tauchen, and Zhou (2009), DY to Drechsler and Yaron (2011), Wachter to Wachter (2013), and BEGE to Bekaert and Engstrom (2017).

	BTZ	DY	Wachter	BEGE	US: 1990-2017	US: 1969-2017
$E(r_{f,t})$	0.69%	0.95%	0.47%	1.24%	0.64%	1.35%
					(0.60%)	(0.52%)
$Std(r_{f,t})$	9.86%	2.12%	2.73%	1.46%	1.92%	2.30%
					(0.18%)	(0.21%)
$E(r_t - r_{f,t})$	7.79%	6.04%	4.92%	6.15%	4.92%	6.15%
					(2.69%)	(2.02%)
$Std(r_t)$	5.70%	18.10%	21.31%	16.81%	14.64%	15.64%
					(1.33%)	(0.86%)
$E(\frac{P_t}{D_t})$	2.04	19.96	93.77	19.53	51.95	40.74
					(4.69)	(4.94)
$Std(\frac{P_t}{D_t})$	0.09	2.96	26.72	5.35	14.30	17.25
					(2.61)	(2.86)
χ^2-test (1990-2017)	3468***	190***	130***	144***		
χ^2-test (1990-2017): no $\frac{P}{D}$ stats	2645***	12**	43***	20***		
χ^2-test (1969-2017)	1647***	83***	177***	123***		
χ^2-test (1969-2017): no $\frac{P}{D}$ stats	1611***	15**	64***	21***		

Table 4 – Variance and Volatility Risk Premium Moments Fit. Values are computed from monthly data and annualized. The model variance risk premium is defined as the difference between the conditional risk-neutral and physical variances of the next month's log-market return. The model volatility risk premium is defined as the difference between the conditional risk-neutral and physical standard deviations of the next month's log-market return. In data the risk-neutral log-return variance is proxied by VIX^2 and physical log-return variance by the value implied by a Corsi (2009)-type model using high-frequency data. In data the risk-neutral log-return volatility is proxied by VIX and physical log-return standard deviation by the value implied by a Corsi (2009)-type model using high-frequency data. The block-bootstrap standard error in parentheses are obtained by re-sampling 10,000 time series of the historical length with a block length of 60 months. BTZ refers to Bollerslev, Tauchen, and Zhou (2009), DY to Drechsler and Yaron (2011), Wachter to Wachter (2013), and BEGE to Bekaert and Engstrom (2017).

Panel A: Variance Risk Premium Moments Fit					
	BTZ	DY	Wachter	BEGE	US: 1990-2017
Mean	0.0000	0.0093	0.1208	0.0064	0.0195
					(0.0026)
Standard deviation	0.0000	0.0199	0.1025	0.0023	0.0225
					(0.0051)
Lag 1 autocorrelation	1.00	0.90	0.99	0.95	0.52
					(0.09)
Panel B: Volatility Risk Premium Moments Fit					
	BTZ	DY	Wachter	BEGE	US: 1990-2017
Mean	0.00%	2.35%	15.08%	1.87%	5.36%
					(0.58%)
Standard deviation	0.00%	1.94%	9.52%	0.39%	4.02%
					(0.33%)
Lag 1 autocorrelation	1.00	0.89	0.99	0.97	0.45
					(0.10)

Table 5 – Martin (2017) Bound Fit. Standard errors in parentheses are computed via bootstrapping the historical time series 10,000 times. VIX and SVIX are defined as in Martin (2017). Model values are obtained by simulating 100,000 monthly observations for each model. BTZ refers to Bollerslev, Tauchen, and Zhou (2009), DY to Drechsler and Yaron (2011), Wachter to Wachter (2013), and BEGE to Bekaert and Engstrom (2017).

	BTZ	DY	Wachter	BEGE	US: 1996-2012
$\overline{VIX}$	19.58%	16.70%	38.79%	17.94%	22.32%
Standard error					(0.59%)
$\overline{SVIX}$	19.61%	16.62%	32.31%	17.76%	20.91%
Standard error					(0.53%)
$\overline{VIX - SVIX}$	-0.03%	0.07%	6.48%	0.18%	1.41%
Standard error					(0.06%)

Table 6 – Percentiles of Next Month Consumption Growth Conditional on the Current Variance Risk Premium Fit. In models the variance risk premium is defined as the difference between the conditional risk-neutral and physical variances of log-market returns. In data the risk-neutral log-return variance is proxied by VIX^2 and physical log-return variance by the value implied by a Corsi (2009)-type model using high-frequency data. High variance risk premium is defined as the variance risk premium above 80^{th} unconditional percentile and low variance risk premium as the variance risk premium below 20^{th} unconditional percentile. Values are computed numerically by simulating time series of 1,000,000 months under each model. BTZ refers to Bollerslev, Tauchen, and Zhou (2009), DY to Drechsler and Yaron (2011), Wachter to Wachter (2013), and BEGE to Bekaert and Engstrom (2017). For the High-Low Difference, *** indicates that the model value is more than 2.58 standard deviations away from the data counterpart.

	High variance risk premium	Low variance risk premium	High-Low Difference
Panel A: BTZ			
10^{th} percentile	-1.62%	-11.91%	10.29%***
50^{th} percentile	0.15%	0.15%	0.00%
90^{th} percentile	1.92%	12.21%	−10.29%***
Panel B: DY			
10^{th} percentile	-1.06%	-0.47%	−0.59%***
50^{th} percentile	0.16%	0.16%	0.00%
90^{th} percentile	1.38%	0.79%	0.59%***
Panel C: DY skipping forward one month			
10^{th} percentile	-1.03%	-0.47%	−0.56%***
50^{th} percentile	0.16%	0.16%	0.00%
90^{th} percentile	1.34%	0.79%	0.55%***
Panel D: Wachter			
10^{th} percentile	-0.55%	-0.53%	−0.02%***
50^{th} percentile	0.20%	0.21%	-0.01%
90^{th} percentile	0.94%	0.94%	0.00%
Panel E: BEGE			
10^{th} percentile	-0.31%	-0.20%	-0.11%
50^{th} percentile	0.17%	0.16%	0.01%
90^{th} percentile	0.56%	0.51%	0.05%
Panel F: US 1990-2017			
10^{th} percentile	-0.23%	-0.02%	-0.21%
50^{th} percentile	0.18%	0.20%	-0.02%
90^{th} percentile	0.50%	0.50%	0.00%

Table 7 – Classical Minimum Distance Parameter Estimates. Values are monthly. The weighting matrix is the inverse of the covariance matrix estimated from re-sampling 10,000 time series of the historical length with the block length of 60 months. A diagonal weighting matrix is used. Standard errors are in parentheses.

	Base model	Full model
	Preferences	
β	1.0000	1.0000
	(fixed)	(fixed)
γ	1.9870	2.0593
	(0.5972)	(0.6480)
$\bar{q}$	1.0000	1.0000
	(fixed)	(fixed)
ρ_q	0.9904	0.9873
	(0.0121)	(0.0147)
σ_{qp}	$-2.64 \cdot 10^{-5}$	$-2.39 \cdot 10^{-5}$
	(0.0011)	(0.0009)
σ_{qn}	0.1140	0.1203
	(0.0327)	(0.0330)
	Consumption growth	
$\bar{g}$	0.0017	0.0019
	(0.0002)	(0.0002)
σ_{cp}	0.0007	0.0006
	(0.0002)	(0.0002)
σ_{cn}	0.0035	0.0033
	(0.0005)	(0.0005)
$\bar{p}$	11.0848	16.6721
	(4.8705)	(6.1193)
$\bar{n}$	0.0621	0.0822
	(0.0211)	(0.0223)
ρ_n	0.9954	0.9915
	(0.0164)	(0.0205)
σ_{nn}	0.0327	0.0394
	(0.0159)	(0.0168)
ϕ_g		-0.0056
		(0.0023)
	Dividend growth	
ϕ_d		-0.0007
		(0.0019)
γ_g		2.9255
		(0.3628)
γ_n		0.7336
		(0.3930)

Table 8 – Classical Minimum Distance Moments Fit. Values are monthly. The weighting matrix is the inverse of the covariance matrix estimated from re-sampling 10,000 time series of the historical length with the block length of 60 months. The diagonal weighting matrix is used. In data the risk-neutral log-return variance is proxied by VIX^2 and physical log-return variance by the value implied by a Corsi (2009)-type model using high-frequency data. ** and *** correspond to the rejection at the 5% and 1% significance levels, respectively.

Moment	Base model	Full model	Data	Data standard error
		Consumption growth		
$E(g_t)$	0.0017	0.0019	0.0020	0.0002
$Std(g_t)$	0.0024	0.0024	0.0024	0.0002
$Skw(g_t)$	0.1170	-0.0100	0.1163	0.3141
$ExKur(g_t)$	2.0166	1.8513	2.0186	0.7741
		Risk-free rate		
$E(r_{f,t})$	0.0013	0.0007	0.0005	0.0004
$Std(r_{f,t})$	0.0016	0.0016	0.0016	0.0001
$Corr(r_{f,t}, r_{f,t+1})$	0.9858	0.9791	0.9735	0.0093
		Dividend growth		
$Std(d_t)$		0.0074	0.0075	0.0015
$Skw(d_t)$		-0.3444	-0.9717	0.6496
$ExKur(d_t)$		3.5383	3.0077	1.4094
		Equity		
$E(r_t - r_{f,t})$	0.0020	0.0028	0.0041	0.0023
$Std(r_t)$	0.0462	0.0455	0.0426	0.0039
$E(pd_t)$	6.4515	6.4326	6.3969	0.0644
$Std(pd_t)$	0.3002	0.2384	0.2796	0.0377
$Corr(pd_t, pd_{t+1})$	0.9897	0.9846	0.9917	0.0079
		Options		
$Var^Q(r_t) - Var(r_t)$	0.0015	0.0016	0.0016	0.0003
$Std(Var_t^Q(r_{t+1}) - Var_t(r_{t+1}))$	0.0020	0.0017	0.0019	0.0003
$VIX^2 - SVIX^2$	0.0007	0.0007	0.0006	0.0001
χ^2-statistic	12.5042**	6.2485		
p-value	1.49%	10.10%		

Table 9 – Percentiles of Next Month Consumption Growth Conditional on the Current Variance Risk Premium: Model Implications. The model variance risk premium is defined as the difference between the conditional risk-neutral and physical variances of the next month's log-market return. In data the risk-neutral log-return variance is proxied by VIX^2 and physical log-return variance by the value implied by a Corsi (2009)-type model using high-frequency data. High variance risk premium is defined as the variance risk premium above 80^{th} unconditional percentile and low variance risk premium as the variance risk premium below 20^{th} unconditional percentile. Values are computed numerically by simulating time series of 100,000 months under each model. Data block-bootstrap standard errors in parentheses are obtained by re-sampling 10,000 time series of the historical length with a block length of 60 months. *** indicates the statistical significance of the data value at the 1% level.

Panel A: US data 1990M1-2017M2			
	High variance risk premium	Low variance risk premium	High-Low difference
10^{th} percentile	-0.23%	-0.02%	-0.21%***
			(0.07%)
50^{th} percentile	0.18%	0.20%	-0.02%
			(0.05%)
90^{th} percentile	0.50%	0.50%	0.00%
			(0.07%)
Panel B: Base model			
	High variance risk premium	Low variance risk premium	High-Low difference
10^{th} percentile	-0.17%	-0.10%	-0.07%
50^{th} percentile	0.16%	0.15%	0.01%
90^{th} percentile	0.50%	0.48%	0.02%
Panel C: Full model			
	High variance risk premium	Low variance risk premium	High-Low difference
10^{th} percentile	-0.24%	-0.04%	-0.20%
50^{th} percentile	0.11%	0.21%	-0.10%
90^{th} percentile	0.44%	0.52%	-0.08%

Table 10 – Variance Risk Premium Monthly Autocorrelation. The data values are for the difference between the VIX^2 and the conditional physical variance estimate from a Corsi (2009) model. Data standard errors in parentheses are obtained by re-sampling 10,000 time series of the historical length with a block length of 60 months. The model estimates are for the difference between the conditional risk-neutral and physical variances of log-returns.

Model		US data 1990M1-2017M12	
Base	Full	Estimate	Standard error
0.9954	0.9915	0.5197	0.0943

Table 11 – Classical Minimum Distance Parameter Estimates for the Model with a Preference Shock. Values are monthly. The weighting matrix is the inverse of the covariance matrix estimated from re-sampling 10,000 time series of the historical length with the block length of 60 months. A diagonal weighting matrix is used. Standard errors are in parentheses.

Preferences	
β	1.0000
	(fixed)
γ	2.1432
	(0.6441)
$\bar{q}$	1.0000
	(fixed)
ρ_q	0.9926
	(0.0211)
σ_{qp}	$-2 \cdot 10^{-5}$
	(0.0010)
σ_{qn}	0.0204
	(0.0096)
σ_{qq}	0.2085
	(0.0584)
$\bar{s}$	0.0037
	(0.0018)
ρ_s	0.5912
	(0.0179)
σ_{sq}	0.0554
	(0.0272)
Consumption growth	
$\bar{g}$	0.0018
	(0.0002)
σ_{cp}	0.0017
	(0.0003)
σ_{cn}	0.0012
	(0.0004)
$\bar{p}$	0.9003
	(0.3710)
$\bar{n}$	2.1432
	(0.8499)
ρ_n	0.9938
	(0.0295)
σ_{nn}	0.1705
	(0.0514)
ϕ_n	-0.0002
	(0.0001)
Dividend growth	
ϕ_d	-0.0001
	(0.0002)
γ_g	0.9570
	(0.4158)
γ_n	3.5825
	(1.0001)

Table 12 – Classical Minimum Distance Moments Fit for the Model with a Preference Shock. Values are monthly. The weighting matrix is the inverse of the covariance matrix estimated from re-sampling 10,000 time series of the historical length with the block length of 60 months. A diagonal weighting matrix is used. In the data the risk-neutral log-return variance is proxied by VIX^2 and the physical log-return variance by the value implied by a Corsi (2009)-type model using high-frequency data. *** corresponds to the rejection at the 1% significance level.

Moment	Model	Data	Data standard error
Consumption growth			
$E(g_t)$	0.0018	0.0020	0.0002
$Std(g_t)$	0.0024	0.0024	0.0002
$Skw(g_t)$	0.1097	0.1163	0.3141
$ExKur(g_t)$	2.0303	2.0186	0.7741
Risk-free rate			
$E(r_{f,t})$	0.0009	0.0005	0.0004
$Std(r_{f,t})$	0.0018	0.0016	0.0001
$Corr(r_{f,t}, r_{f,t+1})$	0.9686	0.9735	0.0093
Dividend growth			
$Std(d_t)$	0.0081	0.0075	0.0015
$Skw(d_t)$	-1.2818	-0.9717	0.6496
$ExKur(d_t)$	2.6162	3.0077	1.4094
Equity			
$E(r_t - r_{f,t})$	0.0025	0.0041	0.0023
$Std(r_t)$	0.0419	0.0426	0.0039
$E(pd_t)$	6.4555	6.3969	0.0644
$Std(pd_t)$	0.2791	0.2796	0.0377
$Corr(pd_t, pd_{t+1})$	0.9919	0.9917	0.0079
Options			
$Var^Q(r_t) - Var(r_t)$	0.0015	0.0016	0.0003
$Std(Var_t^Q(r_{t+1}) - Var_t(r_{t+1}))$	0.0015	0.0019	0.0003
$AC_1(Var_t^Q(r_{t+1}) - Var_t(r_{t+1}))$	0.5939	0.5197	0.0943
$VIX^2 - SVIX^2$	0.0007	0.0006	0.0001
χ^2-statistic	8.6043***		
p-value	0.37%		

Table 13 – Percentiles of Next Month Consumption Growth Conditional on the Current Variance Risk Premium: Model with a Preference Shock. The model variance risk premium is defined as the difference between the conditional risk-neutral and physical variances of the next month's log-market return. In data the risk-neutral log-return variance is proxied by VIX^2 and physical log-return variance by the value implied by a Corsi (2009)-type model using high-frequency data. High variance risk premium is defined as the variance risk premium above 80^{th} unconditional percentile and low variance risk premium as the variance risk premium below 20^{th} unconditional percentile. Model values are computed numerically by simulating time series of 100,000 months. Data block-bootstrap standard errors in parentheses are obtained by re-sampling 10,000 time series of the historical length with a block length of 60 months. *** indicates the statistical significance of the data value at the 1% level.

Panel A: Model with a preference shock			
	High variance risk premium	Low variance risk premium	High-Low difference
10^{th} percentile	-0.25%	0.08%	-0.34%
50^{th} percentile	0.11%	0.19%	-0.08%
90^{th} percentile	0.44%	0.48%	-0.05%
Panel B: US data 1990M1-2017M2			
	High variance risk premium	Low variance risk premium	High-Low difference
10^{th} percentile	-0.23%	-0.02%	-0.21%***
			(0.07%)
50^{th} percentile	0.18%	0.20%	-0.02%
			(0.05%)
90^{th} percentile	0.50%	0.50%	0.00%
			(0.07%)

Table 14 – Variance Risk Premium and Excess Equity Return Predictability. The values are slope coefficients from monthly regressions. The data values are from regressing log aggregate excess equity returns on the variance risk premium, VP_t, and the conditional physical variance estimate from a Corsi (2009)-type model using high-frequency data, CV_t. Data standard errors in parentheses are obtained by re-sampling 10,000 time series of the historical length with a block length of 60 months. The model estimates are from regressing log aggregate equity excess returns on the difference between the variance risk premium of log-market returns, VP_t, and conditional physical variance of log-returns, CV_t. The model values are obtained by simulating 100,000 monthly samples. BTZ refers to Bollerslev, Tauchen, and Zhou (2009), DY to Drechsler and Yaron (2011), Wachter to Wachter (2013), and BEGE to Bekaert and Engstrom (2017). * and *** correspond to the statistical significance at the 10% and 1% levels, respectively.

Specification 1: $r_{t+1 \text{ month}} - r_{f,t} = \beta_0 + \beta_{VP} \cdot VP_t + \epsilon_{t+1}$								
	Extant models				New Model			Data
	BTZ	DY	Wachter	BEGE	base	full	preference shock	
β_{VP}	-0.0639	0.0579	0.0047	0.1024	0.0131	0.0166	0.0119	0.0378*
								(0.0217)
Specification 2: $r_{t+1 \text{ month}} - r_{f,t} = \beta_0 + \beta_{VP} \cdot VP_t + \beta_{CV} \cdot CV_t + \epsilon_{t+1}$								
	Extant models				New Model			Data
	BTZ	DY	Wachter	BEGE	base	full	preference shock	
β_{VP}	-0.0002	-	0.0100	0.0996	-	-	0.0137	0.0374*
								(0.0206)
β_{CV}	0.0181	-	-0.0300	0.0006	-	-	0.0018	-0.0156
								(0.0148)
Specification 3: $r_{t+3 \text{ months}} - r_{f,t} = \beta_0 + \beta_{VP} \cdot VP_t + \epsilon_{t+1}$								
	Extant models				New Model			Data
	BTZ	DY	Wachter	BEGE	base	full	preference shock	
β_{VP}	-0.1853	0.1582	0.0146	0.2890	0.0395	0.0519	0.0234	0.0842***
								(0.0309)
Specification 4: $r_{t+3 \text{ months}} - r_{f,t} = \beta_0 + \beta_{VP} \cdot VP_t + \beta_{CV} \cdot CV_t + \epsilon_{t+1}$								
	Extant models				New Model			Data
	BTZ	DY	Wachter	BEGE	base	full	preference shock	
β_{VP}	-0.0072	-	0.0270	0.3044	-	-	0.0279	0.0838***
								(0.0303)
β_{CV}	0.0507	-	-0.0734	-0.0033	-	-	0.0135	-0.0170
								(0.0315)

Internet Appendix I: The Variance Risk Premium under Log-Normality

Denote by M_{t+1} the gross pricing kernel at time $t+1$, and R_{t+1} the gross equity return at time $t+1$. In no arbitrage economies, the usual pricing condition implies $E_t[M_{t+1}R_{t+1}] = 1$. We use lowercase letters to indicate the natural logarithms of upper case variables.

Proposition 1. *If M_{t+1} and R_{t+1} are conditionally log-normal, then the variance risk premium for log-returns is 0, that is, the physical variance of log-returns is equal to their risk-neutral variance.*

Proof. Suppose that m_{t+1} and r_{t+1} both have linear dependence on a multivariate conditionally Gaussian state vector, X_{t+1}:

$$\begin{aligned} &X_{t+1} \sim \mathcal{N}(0, \Sigma_t), \\ &m_{t+1} = \bar{m}_t + \beta_m X_{t+1}, \\ &r_{t+1} = \bar{r}_t + \beta_r X_{t+1}. \end{aligned} \tag{41}$$

Note that the moment-generating function for returns under the physical measure is defined as:

$$\begin{aligned} mgf_t^P(r_{t+1};\nu) =& E_t[exp(\nu r_{t+1})] = E_t[exp(\nu \bar{r}_t + \nu \beta_r' X_{t+1})] \\ &= exp(\nu \bar{r}_t + \frac{1}{2}\nu^2 \beta_r' \Sigma_t \beta_r). \end{aligned} \tag{42}$$

Calculating the first derivative of $mgf_t^P(r_{t+1};\nu)$ with respect to ν and evaluating at $\nu = 0$ yields the conditional mean of returns:

$$E_t[r_{t+1}] = \bar{r}_t. \tag{43}$$

Calculating the second derivative of $mgf_t^P(r_{t+1};\nu)$ with respect to ν and evaluating at $\nu = 0$ yields the uncentered second moment, $E_t[r_{t+1}^2]$. Using the definition of variance, this leads to the expected expression for variance under the physical measure:

$$Var_t[r_{t+1}] = E_t[r_{t+1}^2] - E_t[r_{t+1}^2]^2 = \beta_r' \Sigma_t \beta_r. \tag{44}$$

Next, to calculate risk-neutral conditional moments, we need to evaluate the risk-neutral moment-generating function, which is defined as:

$$mgf_t^Q(r_{t+1};\nu) = \frac{E_t[exp(m_{t+1} + \nu r_{t+1})]}{E_t[exp(m_{t+1})]}. \tag{45}$$

This simplifies to:

$$\begin{aligned} \frac{E_t[exp(m_{t+1} + \nu r_{t+1})]}{E_t[exp(m_{t+1})]} &= E_t[exp(\bar{m}_t + \nu\bar{r}_t + (\beta_m' + \nu\beta_r')X_{t+1}] \\ &= exp(\nu\bar{r}_t + \frac{1}{2}\nu^2\beta_r'\Sigma_t\beta_r + \frac{1}{2}\beta_m'\Sigma_t\nu\beta_r + \frac{1}{2}\beta_r'\Sigma_t\nu\beta_m). \end{aligned} \tag{46}$$

We proceed, again, by evaluating successive derivatives of $mgf_t^Q(r_{t+1};\nu)$ with respect to ν and evaluating at $\nu = 0$. Eventually, we arrive at:

$$Var_t^Q[r_{t+1}] = \beta_r\Sigma_t\beta_r, \tag{47}$$

so that $Var_t^Q[r_{t+1}] = Var_t^P[r_{t+1}]$.

Proposition 2. *If M_{t+1} and R_{t+1} are conditionally log-normal, then the conditional equity premium and the conditional variance risk premium can not simultaneously be positive.*

*P*roof: Given that M_{t+1} and R_{t+1} are conditionally log-normal:

$$\begin{pmatrix} m_{t+1} \\ r_{t+1} \end{pmatrix} \sim \mathcal{N}(\begin{pmatrix} \bar{m}_t \\ \bar{r}_t \end{pmatrix}, \begin{pmatrix} \sigma_{m,t}^2 & \sigma_{mr,t} \\ \sigma_{mr,t} & \sigma_{r,t}^2 \end{pmatrix}),$$

the physical return variance can be computed as:

$$Var_t(R_{t+1}) = E_t(R_{t+1}^2) - E_t(R_{t+1})^2 = e^{2\bar{r}_t + 2\sigma_{r,t}^2} - e^{2\bar{r}_t + \sigma_{r,t}^2} = e^{2\bar{r}_t + \sigma_{r,t}^2}(e^{\sigma_{r,t}^2} - 1). \tag{48}$$

The risk-neutral expectation is:

$$E_t^Q(R_{t+1}) = \frac{E_t(M_{t+1}R_{t+1})}{E_t(M_{t+1})}.$$

The risk-neutral variance can now be computed as:

$$
\begin{aligned}
&E_t(M_{t+1}) = E_t(e^{m_{t+1}}) = e^{\bar{m}_t + 0.5\sigma^2_{m,t}}, \\
&E_t(M_{t+1}R_{t+1}) = E_t(e^{m_{t+1}+r_{t+1}}) = e^{\bar{m}_t + \bar{r}_t + 0.5\sigma^2_{m,t} + 0.5\sigma^2_{r,t} + \sigma_{mr,t}}, \\
&E^Q_t(R_{t+1}) = e^{\bar{r}_t + 0.5\sigma^2_{r,t} + \sigma_{mr,t}}, \\
&E_t(M_{t+1}R^2_{t+1}) = E_t(e^{m_{t+1}+2r_{t+1}}) = e^{\bar{m}_t + 2\bar{r}_t + 0.5\sigma^2_{m,t} + 2\sigma_{mr,t} + 2\sigma^2_{r,t}}, \\
&E^Q_t(R^2_{t+1}) = e^{2\bar{r}_t + 2\sigma_{mr,t} + 2\sigma^2_{r,t}}, \\
&Var^Q_t(R_{t+1}) = E^Q_t(R^2_{t+1}) - E^Q_t(R_{t+1})^2 = e^{2\bar{r}_t + 2\sigma_{mr,t} + 2\sigma^2_{r,t}} - e^{2\bar{r}_t + \sigma^2_{r,t} + 2\sigma_{mr,t}} = \\
&e^{2\bar{r}_t + \sigma^2_{r,t} + 2\sigma_{mr,t}}(e^{\sigma^2_{r,t}} - 1).
\end{aligned}
\tag{49}
$$

Comparing the final line of (49) to the final line of (48), we can see that in order for the variance risk premium to be positive ($Var^Q_t(R_{t+1}) > Var_t(R_{t+1})$) as in the data, it must be the case that $\sigma_{mr,t} > 0$.

However, note that $\sigma_{mr,t} > 0$ implies (counterfactually) a negative equity premium. This can be seen from the typical Euler equation using properties of the log-normal distribution:

$$
\begin{aligned}
&E_t(M_{t+1}R_{t+1}) = 1, \\
&e^{\bar{m}_t + \bar{r}_t + 0.5\sigma^2_{r,t} + 0.5\sigma^2_{m,t} + \sigma_{mr,t}} = 1, \\
&e^{\bar{m}_t + \bar{r}_t + 0.5\sigma^2_{r,t} + 0.5\sigma^2_{m,t}} = e^{-\sigma_{mr,t}}, \\
&E_t(R_{t+1})E_t(M_{t+1}) = e^{-\sigma_{mr,t}}, \\
&\ln E_t(R_{t+1}) - r_{f,t} = -\sigma_{mr,t},
\end{aligned}
\tag{50}
$$

where $r_{f,t}$ is the log-risk-free rate.

Internet Appendix II: Parameterization of Existing Models

Model Parameters. The notation follows original articles. Parameterizations are monthly except for the annual parameterization in Wachter (2013). For Bollerslev, Tauchen, and Zhou (2009) κ_1 is the constant in Campbell and Shiller (1988) log-linearization of the aggregate market return: $r_{t+1} = \kappa_0 + \kappa_1 \omega_{t+1} - \omega_t + d_{t+1}$.

			Panel A: Bollerslev, Tauchen, and Zhou (2009)				
Preferences	δ	γ	ψ	κ_1			
	0.997	10	1.5	0.9			
g_t	μ_g						
	0.0015						
$\sigma^2_{g,t}$	a_σ	ρ_σ					
	$1.34 \cdot 10^{-6}$	0.978					
q_t	a_q	ρ_q	ϕ_q				
	$2 \cdot 10^{-7}$	0.8	0.0010				
			Panel B: Drechsler and Yaron (2011)				
Preferences	δ	γ	ψ				
	0.999	9.5	2.0				
g_t	g_0	φ_c	ω_c				
	0.0016	0.0066	0.5				
x_t	x_0	ρ_x	φ_x	ω_x	$l_{1,x}$	μ_x	ν_x
	0	0.976	0.0002	1	0.0667	0.0008	1
$\bar{\sigma}^2_t$	$\bar{\sigma}^2_0$	$\rho_{\bar{\sigma}^2}$	$\varphi_{\bar{\sigma}^2}$	$\omega_{\bar{\sigma}^2}$			
	0.015	0.985	0.1	0			
σ^2_t	σ^2_0	ρ_{σ^2}	φ_{σ^2}	ω_{σ^2}	l_{1,σ^2}	μ_{σ^2}	ν_{σ^2}
	0	0.87	0.35	1	0.0667	2.55	1
d_t	d_0	ϕ	φ_d	ω_d	Ω_{cd}		
	0.0016	2.5	0.0376	0.125	0.2		
			Panel C: Wachter (2013)				
Preferences	β	γ					
	0.012	3					
C_t	μ	σ	Z				
	0.0252	0.02	Table 8 in Barro and Ursua (2008)				
D_t	ϕ						
	2.6						
λ_t	λ	κ	σ_λ				
	0.0355	0.08	0.067				
			Panel D: Bekaert and Engstrom (2017)				
Preferences	δ	γ	$\bar{s}$	ϕ	b		
	0.9999	11.43	-1.5585	0.9963	0.0099		
g_t	$\bar{g}$	σ_{cp}	σ_{cn}	$\bar{p}$	$\bar{n}$	ρ_n	σ_{nn}
	0.0014	0.0007	0.0019	11.43	1.56	0.91	0.32
d_t	σ_{dp}	σ_{dn}					
	-0.0055	0.0217					

Internet Appendix III: Model Solution

The price-dividend ratio is:

$$
\begin{aligned}
\frac{P_t}{D_t} &= \sum_{i=1}^{\infty} e^{A_i + B_i p_t + C_i n_t + D_i q_t}, \\
A_1 &= \ln\beta + (1-\gamma)\bar{g} + \gamma(1-\rho_q)\bar{q}, \\
B_1 &= -g(m_{\omega,p} + \alpha\sigma_{cp}), \\
C_1 &= -g(m_{\omega,n} - \alpha\sigma_{cn}), \\
D_1 &= -\gamma(1-\rho_q), \\
A_n &= A_{n-1} + A_1 + B_{n-1}\bar{p}(1-\rho_p) + C_{n-1}\bar{n}(1-\rho_n) + D_{n-1}\bar{q}(1-\rho_q), \\
B_n &= B_{n-1}\rho_p - g(m_{\omega,p} + \alpha\sigma_{cp} + B_{n-1}\sigma_{pp} + D_{n-1}\sigma_{qp}), \\
C_n &= C_{n-1}\rho_n - g(m_{\omega,n} - \alpha\sigma_{cn} + C_{n-1}\sigma_{nn} + D_{n-1}\sigma_{qn}), \\
D_n &= D_{n-1}\rho_q + D_1.
\end{aligned}
\tag{51}
$$

The Taylor series log-linearization of (51) is:

$$
\begin{aligned}
pd_t &\approx K_0^1 + K_p^1 p_t + K_n^1 n_t + K_q^1 q_t, \\
K_p^1 &= \frac{\sum_{i=1}^{\infty} B_i \exp(A_i + B_i\bar{p} + C_i\bar{n} + D_i\bar{q})}{\sum_{i=1}^{\infty} \exp(A_i + B_i\bar{p} + C_i\bar{n} + D_i\bar{q})}, \\
K_n^1 &= \frac{\sum_{i=1}^{\infty} C_i \exp(A_i + B_i\bar{p} + C_i\bar{n} + D_i\bar{q})}{\sum_{i=1}^{\infty} \exp(A_i + B_i\bar{p} + C_i\bar{n} + D_i\bar{q})}, \\
K_q^1 &= \frac{\sum_{i=1}^{\infty} D_i \exp(A_i + B_i\bar{p} + C_i\bar{n} + D_i\bar{q})}{\sum_{i=1}^{\infty} \exp(A_i + B_i\bar{p} + C_i\bar{n} + D_i\bar{q})}, \\
K_0^1 &= \ln[\sum_{i=1}^{\infty} \exp(A_i + B_i\bar{p} + C_i\bar{n} + D_i\bar{q})] - K_p^1\bar{p} - K_n^1\bar{n} - K_q^1\bar{q}.
\end{aligned}
\tag{52}
$$

The Taylor series linearization of $\ln(1+\frac{P_{t+1}}{D_{t+1}})$ is:

$$
\begin{aligned}
&\ln(1+\frac{P_{t+1}}{D_{t+1}}) \approx K_0^2 + K_p^2 p_{t+1} + K_n^2 n_{t+1} + K_q^2 q_{t+1},\\
&K_p^2 = \frac{\sum_{i=1}^{\infty} B_i \exp(A_i + B_i\bar{p} + C_i\bar{n} + D_i\bar{q})}{1+\sum_{i=1}^{\infty}\exp(A_i + B_i\bar{p} + C_i\bar{n} + D_i\bar{q})},\\
&K_n^2 = \frac{\sum_{i=1}^{\infty} C_i \exp(A_i + B_i\bar{p} + C_i\bar{n} + D_i\bar{q})}{1+\sum_{i=1}^{\infty}\exp(A_i + B_i\bar{p} + C_i\bar{n} + D_i\bar{q})},\\
&K_q^2 = \frac{\sum_{i=1}^{\infty} D_i \exp(A_i + B_i\bar{p} + C_i\bar{n} + D_i\bar{q})}{1+\sum_{i=1}^{\infty}\exp(A_i + B_i\bar{p} + C_i\bar{n} + D_i\bar{q})},\\
&K_2^0 = \ln[1+\sum_{i=1}^{\infty}\exp(A_i + B_i\bar{p} + C_i\bar{n} + D_i\bar{q})] - K_p^2\bar{p} - K_n^2\bar{n} - K_q^2\bar{q}.
\end{aligned}
\tag{53}
$$

The expression for the log-return is:

$$
\begin{aligned}
&r_{t+1} \approx r_0 + r_p p_t + r_n n_t + r_q q_t + r_{\omega,p}\omega_{p,t+1} + r_{\omega,n}\omega_{n,t+1},\\
&r_0 = \bar{g} + K_0^2 - K_0^1 + K_p^2\bar{p}(1-\rho_p) + K_n^2\bar{n}(1-\rho_n) + K_q^2\bar{q}(1-\rho_q),\\
&r_p = K_p^2\rho_p - K_p^1,\\
&r_n = K_n^2\rho_n - K_n^1,\\
&r_q = K_q^2\rho_q - K_q^1,\\
&r_{\omega,p} = \alpha\sigma_{cp} + K_p^2\sigma_{pp} + K_q^2\sigma_{qp},\\
&r_{\omega,n} = -\alpha\sigma_{cn} + K_n^2\sigma_{nn} + K_q^2\sigma_{qn}.
\end{aligned}
\tag{54}
$$

$SVIX^2$ can be computed as:

$$
\begin{aligned}
&SVIX^2 = \sigma_{Q,t}^2(\frac{R_{t+1}}{R_{f,t}}) = \frac{1}{R_{f,t}^2}\sigma_{Q,t}^2(R_{t+1}) = \frac{1}{R_{f,t}^2}(E_t^Q(R_{t+1}^2) - (E_t^Q R_{t+1})^2) =\\
&\frac{1}{R_{f,t}^2}(R_{f,t}E_t(M_{t+1}R_{t+1}^2) - (R_{f,t}E_t(M_{t+1}R_{t+1}))^2) = \frac{1}{R_{f,t}^2}(R_{f,t}E_t(M_{t+1}R_{t+1}^2) - R_{f,t}^2) =\\
&\frac{E_t(M_{t+1}R_{t+1}^2)}{R_{f,t}} - 1 = E_t[exp(m_{t+1} + 2r_{t+1} - r_{f,t})] - 1.
\end{aligned}
\tag{55}
$$

By plugging (18), (20) and (28) into (55) we obtain:

$$
\begin{aligned}
&SVIX_t^2 = exp(m_0 + 2r_0 - f_0 + [m_q + 2r_q - f_q]q_t + \\
&[2r_p - g(m_{\omega,p} + 2r_{\omega,p}) - f_p]p_t + [2r_n - g(m_{\omega,n} + 2r_{\omega,n}) - f_n]n_t) - 1.
\end{aligned} \tag{56}
$$

Computation of $E[X \cdot \exp(X)]$, where $X \sim \Gamma(k, \theta)$ with k being the shape and θ the scale parameters, respectively:

$$
\begin{aligned}
&E[X \cdot \exp(X)] = \frac{1}{\Gamma(k)\theta^k} \int_0^\infty x^{k-1} e^{-\frac{x}{\theta}} x e^x dx = \frac{1}{\Gamma(k)\theta^k} \int_0^\infty x^{(k+1)-1} e^{-\frac{x}{\frac{\theta}{1-\theta}}} dx = \\
&\frac{1}{\Gamma(k)\theta^k} \cdot \Gamma(k+1) \cdot (\frac{\theta}{1-\theta})^{k+1} \cdot \underbrace{\frac{1}{\Gamma(k+1) \cdot (\frac{\theta}{1-\theta})^{k+1}} \cdot \int_0^\infty x^{(k+1)-1} e^{-\frac{x}{\frac{\theta}{1-\theta}}} dx}_{=1,\text{ as pdf of gamma distribution with shape } k+1 \text{ and scale } \frac{\theta}{1-\theta}} = \\
&\frac{\Gamma(k+1) \cdot (\frac{\theta}{1-\theta})^{k+1}}{\Gamma(k)\theta^k} = \frac{k\theta}{(1-\theta)^{k+1}},
\end{aligned} \tag{57}
$$

where the last step uses the property of the gamma function that $\Gamma(k+1) = k \cdot \Gamma(k)$ and the second line requires $\theta < 1$ in order for $\frac{1}{\Gamma(k+1)\cdot(\frac{\theta}{1-\theta})^{k+1}} \cdot \int_0^\infty x^{(k+1)-1} e^{-\frac{x}{\frac{\theta}{1-\theta}}} dx$ to be a proper probability density function. This condition is always satisfied in our estimation. Now VIX_t^2 can be computed by plugging (18), (20), and (28) into (35):

$$
\begin{aligned}
&VIX_t^2 = 2 \cdot [r_{f,t} - e^{r_{f,t}} \cdot \\
&E_t\{e^{m_0 + m_q q_t + m_{\omega,p}\omega_{p,t+1} + m_{\omega,n}\omega_{n,t+1}} \cdot (r_0 + r_p p_t + r_n n_t + r_q q_t + r_{\omega,p}\omega_{p,t+1} + r_{\omega,n}\omega_{n,t+1})\}] = \\
&2 \cdot [r_{f,t} - e^{r_{f,t} + m_0 + m_q q_t} \cdot E_t\{e^{m_{\omega,p}\omega_{p,t+1} + m_{\omega,n}\omega_{n,t+1}} \cdot (r_0 + r_p p_t + r_n n_t + r_q q_t) + \\
&e^{m_{\omega,p}\omega_{p,t+1} + m_{\omega,n}\omega_{n,t+1}} \cdot (r_{\omega,p}\omega_{p,t+1} + r_{\omega,n}\omega_{n,t+1})\}] = 2 \cdot [r_{f,t} - e^{r_{f,t} + m_0 + m_q q_t} \cdot \\
&\{e^{-g(m_{\omega,p})p_t - g(m_{\omega,n})n_t} \cdot (r_0 + r_p p_t + r_n n_t + r_q q_t) + E_t[e^{m_{\omega,p}\omega_{p,t+1}} e^{m_{\omega,n}\omega_{n,t+1}} r_{\omega,p}\omega_{p,t+1}] + \\
&E_t[e^{m_{\omega,p}\omega_{p,t+1}} e^{m_{\omega,n}\omega_{n,t+1}} r_{\omega,n}\omega_{n,t+1}]\}] = 2 \cdot [r_{f,t} - e^{r_{f,t} + m_0 + m_q q_t} \cdot \{e^{-g(m_{\omega,p})p_t - g(m_{\omega,n})n_t} \cdot \\
&(r_0 + r_p p_t + r_n n_t + r_q q_t) + e^{-g(m_{\omega,n})n_t} r_{\omega,p} E_t[e^{M_{\omega,p}\omega_{p,t+1}}\omega_{p,t+1}] + e^{-g(m_{\omega,p})p_t} r_{\omega,n} E_t[e^{M_{\omega,n}\omega_{n,t+1}}\omega_{n,t+1}]\}] = \\
&2 \cdot [r_{f,t} - e^{r_{f,t} + m_0 + m_q q_t} \cdot \{e^{-g(m_{\omega,p})p_t - g(m_{\omega,n})n_t} \cdot (r_0 + r_p p_t + r_n n_t + r_q q_t) + \\
&e^{-g(m_{\omega,n})n_t} r_{\omega,p} \cdot (-p_t e^{-g(m_{\omega,p})p_t} + e^{-m_{\omega,p}p_t} \frac{p_t}{(1 - m_{\omega,p})^{p_t+1}}) + \\
&e^{-g(m_{\omega,p})p_t} r_{\omega,n} \cdot (-n_t e^{-g(m_{\omega,n})n_t} + e^{-m_{\omega,n}n_t} \frac{n_t}{(1 - m_{\omega,n})^{n_t+1}})\}],
\end{aligned} \tag{58}
$$

where the last step follows from (57).

Internet Appendix IV: Macroeconomic Models Fitting the Variance Risk Premium Persistence

The estimated models are as in section 4.1, but the variance risk premium persistence is now explicitly included as a moment into the classical minimum distance estimation.

Classical Minimum Distance Parameter Estimates. Values are monthly. The weighting matrix is the inverse of the covariance matrix estimated from re-sampling 10,000 time series of the historical length with the block length of 60 months. A diagonal weighting matrix is used. Standard errors are in parentheses.

	Base model	Full model
	Preferences	
β	1.0000	1.0000
	(fixed)	(fixed)
γ	2.1853	2.1694
	(0.5118)	(0.5760)
$\bar{q}$	1.0000	1.0000
	(fixed)	(fixed)
ρ_q	0.9971	0.9873
	(0.0106)	(0.0124)
σ_{qp}	-0.0005	-0.0003
	(0.0010)	(0.0011)
σ_{qn}	0.1150	0.1153
	(0.0296)	(0.0307)
	Consumption growth	
$\bar{g}$	0.0016	0.0016
	(0.0002)	(0.0002)
σ_{cp}	0.0006	0.0004
	(0.0002)	(0.0002)
σ_{cn}	0.0042	0.0040
	(0.0005)	(0.0005)
$\bar{p}$	14.1687	27.7271
	(4.3602)	(6.7220)
$\bar{n}$	0.0446	0.0459
	(0.0180)	(0.0198)
ρ_n	0.5796	0.5812
	(0.0062)	(0.0086)
σ_{nn}	0.0637	0.0646
	(0.0121)	(0.0145)
ϕ_g		-0.0003
		(0.0018)
	Dividend growth	
ϕ_d		-0.0005
		(0.0019)
γ_g		3.5660
		(0.4050)
γ_n		0.9339
		(0.4107)

Classical Minimum Distance Moments Fit. Values are monthly. The weighting matrix is the inverse of the covariance matrix estimated from re-sampling 10,000 time series of the historical length with the block length of 60 months. The diagonal weighting matrix is used. In data the risk-neutral log-return variance is proxied by VIX^2 and physical log-return variance by the value implied by a Corsi (2009)-type model using high-frequency data. *** corresponds to the rejection at the 1% significance level, respectively.

Moment	Base model	Full model	Data	Data standard error
		Consumption growth		
$E(g_t)$	0.0016	0.0016	0.0020	0.0002
$Std(g_t)$	0.0025	0.0025	0.0024	0.0002
$Skw(g_t)$	0.0283	-0.0536	0.1163	0.3141
$ExKur(g_t)$	2.3589	1.9518	2.0186	0.7741
		Risk-free rate		
$E(r_{f,t})$	0.0015	0.0015	0.0005	0.0004
$Std(r_{f,t})$	0.0021	0.0020	0.0016	0.0001
$Corr(r_{f,t}, r_{f,t+1})$	0.9713	0.9708	0.9735	0.0093
		Dividend growth		
$Std(d_t)$		0.0091	0.0075	0.0015
$Skw(g_t)$		-0.3866	-0.9717	0.6496
$ExKur(g_t)$		4.1096	3.0077	1.4094
		Equity		
$E(r_t - r_{f,t})$	0.0016	0.0016	0.0041	0.0023
$Std(r_t)$	0.0360	0.0371	0.0426	0.0039
$E(pd_t)$	6.4325	6.4286	6.3969	0.0644
$Std(pd_t)$	0.4230	0.4218	0.2796	0.0377
$Corr(pd_t, pd_{t+1})$	0.9971	0.9971	0.9917	0.0079
		Options		
$Var^Q(r_t) - Var(r_t)$	0.0010	0.0011	0.0016	0.0003
$Std(Var_t^Q(r_{t+1}) - Var_t(r_{t+1}))$	0.0004	0.0005	0.0019	0.0003
$AC_1(Var_t^Q(r_{t+1}) - Var_t(r_{t+1}))$	0.5796	0.5812	0.5197	0.0943
$VIX^2 - SVIX^2$	0.0005	0.0005	0.0006	0.0001
χ^2-statistic	74.9966***	74.0332***		
p-value	0.00%	0.00%		

Percentiles of Next Month Consumption Growth Conditional on the Current Variance Risk Premium: Model Implications. The model variance risk premium is defined as the difference between the conditional risk-neutral and physical variances of the next month's log-market return. In data the risk-neutral log-return variance is proxied by VIX^2 and physical log-return variance by the value implied by a Corsi (2009)-type model using high-frequency data. High variance risk premium is defined as the variance risk premium above 80^{th} unconditional percentile and low variance risk premium as the variance risk premium below 20^{th} unconditional percentile. Values are computed numerically by simulating time series of 100,000 months under each model. Data block-bootstrap standard errors in parentheses are obtained by re-sampling 10,000 time series of the historical length with a block length of 60 months. *** indicates the statistical significance of the data value at the 1% level.

	High variance risk premium	Low variance risk premium	High-Low difference
Panel A: US data 1990M1-2017M2			
10^{th} percentile	-0.23%	-0.02%	-0.21%***
			(0.07%)
50^{th} percentile	0.18%	0.20%	-0.02%
			(0.05%)
90^{th} percentile	0.50%	0.50%	0.00%
			(0.07%)
Panel B: Base model			
	High variance risk premium	Low variance risk premium	High-Low difference
10^{th} percentile	-0.23%	-0.12%	-0.11%
50^{th} percentile	0.10%	0.15%	-0.05%
90^{th} percentile	0.45%	0.48%	-0.03%
Panel C: Full model			
	High variance risk premium	Low variance risk premium	High-Low difference
10^{th} percentile	-0.24%	-0.12%	-0.12%
50^{th} percentile	0.11%	0.15%	-0.04%
90^{th} percentile	0.45%	0.48%	-0.03%

Variance Risk Premium and Excess Equity Return Predictability. The values are slope coefficients from monthly regressions. The data values are from regressing log aggregate excess equity returns on the variance risk premium, VP_t. Data standard errors in parentheses are obtained by re-sampling 10,000 time series of the historical length with a block length of 60 months. The model estimates are from regressing log aggregate equity excess returns on the difference between the variance risk premium of log-market returns. The model values are obtained by simulating 100,000 monthly samples. * and *** correspond to the statistical significance at the 10% and 1% levels, respectively.

Specification 1: $r_{t+1 \text{ month}} - r_{f,t} = \beta_0 + \beta_{VP} \cdot VP_t + \epsilon_{t+1}$			
	Base model	Full model	Data
β_{VP}	0.0127	0.0173	0.0378*
			(0.0217)
Specification 2: $r_{t+3 \text{ months}} - r_{f,t} = \beta_0 + \beta_{VP} \cdot VP_t + \epsilon_{t+1}$			
	Base	Full	Data
β_{VP}	0.0314	0.0359	0.0842***
			(0.0309)