The Variance Risk Premium in Equilibrium Models*

Geert Bekaert, Columbia University and the Centre for Economic Policy Research,

Eric Engstrom, Board of Governors of the Federal Reserve System

Andrey Ermolov, Gabelli School of Business, Fordham University

March 16, 2022

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 only moderate persistence. High variance risk premiums coincide with the left tail of the consumption growth distribution shifting down. These facts, together with risk neutral skewness being substantially more negative than physical return skewness, 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, risk-neutral skewness, non-Gaussian dynamics, bad volatility, VIX, habit

JEL codes: E44, G12, G13

*Contact information: Geert Bekaert – ✉: 665 West 130$^{\text{th}}$ street, room 1123, 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. We thank an anonymous referee, Gurdip Bakshi (referee), Amit Goyal (editor), Daniel Andrei (discussant), Johnathan Loudis (discussant), conference participants at the 2020 Northern Finance Association Meeting, 2021 Financial Management Association Annual Meeting, and 2022 Midwest Finance Association Annual Meeting, and seminar participants at Syracuse University and the Virtual Derivatives Workshop for their constructive comments which greatly improved the manuscript. 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. 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 variance 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 (option-implied) 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).

Importantly, we also document a strong relation between tails in the variance risk premium and consumption growth distributions. Mimicking the general lack of correlation between asset returns and consumption growth, the correlation between the variance risk premium and consumption growth of non-durables and services is negligible. During 1990:M1-2017:M12, the monthly correlation between the variance premium at time t and monthly consumption growth 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. However, there is a strong correlation in the tails of the distribution. We construct two distributions for consumption growth, conditioning on low and high variance risk premiums, respectively. The consumption growth distribution conditioning on high variance risk premiums is more negatively skewed than the distribution conditioning on the low variance risk premiums. The quantile shifts for the consumption growth distribution when conditioning on low versus high variance risk premiums are statistically and economically significant for the left tail, but not for the right tail. In contrast to most of the literature focusing largely on either pure macro or pure financial moments, this novel statistic directly links macro and financial data. This asymmetric quantile shift in the consumption growth distribution persists in quarterly consumption data and is, if anything, stronger when using annual data going back to 1929. This empirical fact poses a challenge for most of the models we examine, and our results are therefore informative for guiding the literature towards successful macro-finance models.

The variance risk premium also shows surprisingly low persistence: around 50% at the monthly frequency. We show that this low persistence is a robust feature of the data. Finally, any option pricing model must replicate the risk neutral skewness being substantially more negative than estimates of the physical return skewness, which likely reflects the insurance-like features of some options.

We formally confront several prominent models in the literature with these stylized facts regarding the variance risk premium and find that they reject all models. These rejections are true at the parameters used in the original models, but we argue are also true under more general parameterizations as well. Therefore, we introduce a new model in the habit class, which can actually fit the data. The model uses a variant of the "Bad Environment Good Environment" (BEGE henceforth) framework of Bekaert and Engstrom (2017), but the preference specification is different. In particular, while Bekaert and Engstrom (2017) utilizes the highly non-linear price of risk specification of Campbell and Cochrane (1999), the new model features a constant price of risk, but a non-linear specification for the quantity of risk, which drives both consumption growth and the surplus ratio. We also do not impose that the shocks to the consumption and the surplus ratio are perfectly correlated. 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"). Importantly, the model can be solved in (quasi) closed-form, generating a variance risk premium that loads positively (negatively) on bad (good) volatility. Because the risk-neutral variance loads much more heavily on "bad" volatility, than does the physical variance, the model produces a sizable variance risk premium.

The low persistence of the variance premium is an extremely challenging moment for all asset pricing models, because they tend to require highly persistent state variables to generate sufficiently variable equity returns and price-dividend ratios. While different parameter configurations of our basic model can fit the variance risk premium persistence, the fit with respect to other moments then deteriorates slightly when we force the model to do so. The model can fit the variance risk premium persistence well without affecting its ability to fit other moments, if we allow for a small pure sentiment shock.

Our non-linear model for consumption growth fits into a long literature in macro on non-Gaussian models to fit asymmetries in real activity dynamics, including Hamilton (1989), Adrian, Boyarchenko and Giannone (2019), and Bekaert, Engstrom, and Ermolov (2021). Acemoglu, Ozdaglar, and Tahbaz-Salehi (2017) and Baqaee and Farhi (2019) provide models to endogenize asymmetries in aggregate real activity from microeconomic shocks.

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, including additional option pricing puzzles, such as the volatility smirk, negative straddle returns (see, e.g, Bakshi, Crosby, and Gao, 2021) and the flat term-structure of forward variance claims (Dew-Becker et al., 2017). Section 5 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, as published by the CBOE, the implied option volatility of the S&P500 index for contracts with a maturity of one month.[1] Mechanically, the VIX is computed as a weighted average of call and put option prices. Because the weights are proportional to the inverse of the

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

squared strike price (see Bakshi and Madan, 2000, and Britten-Jones and Neuberger, 2000), out of the money put prices receive relatively more weight. The risk-adjusted measure thus shifts weight to states with higher marginal utility (bad states) and this implies that in many realistic economic settings, the variance risk premium is increasing in aggregate risk aversion (see, e.g., Bakshi and Madan, 2006, for illuminating examples). 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 $2 \cdot (\ln[E(X)] - E[\ln(X)])$), whereas the actual risk neutral variance can be approximated using equally weighted call and option prices.[2] All of our results are robust to using the actual risk-neutral variance instead of the VIX^2, but that would postpone the start of our data sample. Our data start on January 02, 1990 (the start of the model-free VIX series)[3] and covers the period until the end of 2017. The panel of option prices that are required to calculate Martin's measure is available to us beginning in only 1996.

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). This computation is only correct under certain assumptions on the data generating process for returns (see Andersen et al., 2003, for a theoretical justification, and Barndorff-Nielsen and Shephard, 2002, for some exceptions). Equation (1) obviously ignores time-variation in conditional means during the month. Also, we use actual simple returns in these computations, whereas some articles suggest using logarithmic returns. We find that realized variances using the two methods are indistinguishable from each other, which is not surprising given the high frequency nature of the returns. Thus, under this definition the conditional variances of log returns and simple returns are actually identical.

The common approach to estimate the conditional variance in (1) uses empirical projections

[2] Martin (2017) denotes the square root of the risk-neutral variance as the "SVIX". While empirically SVIX and VIX are typically close to one another, the VIX is always higher than SVIX.

[3] 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.

of the realized variance onto variables in some 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 typically provides a very good fit. For robustness, we also consider an AR(1) model for realized variances and fit a GJR-GARCH(1,1) model (Glosten, Jagannathan, and Runkle, 1993) on stock returns to extract the conditional variance, with no meaningful differences in results. Note that the data strongly reject treating the realized variance as a martingale.

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 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%. 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

2.2.1 Monthly Data

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. The data are real and per capita. We aggregate real growth rates in non-durables and services at time t using averages of their nominal expenditure shares at time t and time $t-1$ (this is the so called Tornqvist adjustment, suggested by Whelan, 2002).[4] Our goal is to verify the shape of the consumption distribution as a function of variance risk premium realizations. To deal with the relative paucity of data, we contrast just two conditional distributions for consumption growth, depending on either "low" or "high" variance risk premiums, measured one period earlier.

[4]Note that using nominal expenditure shares at time $t-1$, leads to a consumption series that is 0.99 correlated with the series we use.

In Table 1, Panel A, we show the 10^{th} and 90^{th} percentiles of the consumption growth distribution (at $t+1$) together with the median, conditional on observing either a high or low variance risk premium (at t). 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. Given the scant number of available monthly observations, conditioning on more extreme tails is not feasible if we are to retain a reasonable amount of statistical power. Strikingly, the distribution of consumption growth is nearly identical at the median and 90^{th} percentile, but the left tail is 0.23% lower (representing a hefty 2.76 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 a block length of 60 months using 10,000 replications of historical length. These results are robust to the choice of block length and hold even for a block length of 1. The test proposed by Ibragimov and Mueller (2010) to identify differences in distributions also rejects the null of no downward shift of the 10% tail at the 1% level.

Figure 2 presents a graphical illustration of what is essentially a quantile shift of the left 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 suggests that a realistic data generating process for consumption growth must accommodate a shift in its distribution over time. Because this empirical fact is an important motivating factor for our analysis, we provide some robustness checks in Panels B-E of Table 1. In Panels B and C, we show robustness of the result to an alternative choice of the lower/upper percentile, using the $75/25^{\text{th}}$ and $85/15^{\text{th}}$ percentiles. In Panel D, we use the conditional variance of a GARCH model to compute the variance risk premium. The quantile shift is present in this robustness exercise as well, but is now only statistically significant at the 5% level.

One concern with results regarding tails of a distribution is that they may heavily depend on some outlier observations or a handful of episodes. Panel E analyzes in detail what observations drive our results. We have 336 monthly observations in our main sample, so conditional on the variance risk premium being above its 80^{th} percentile or below its 20^{th} percentile there are 7 observations below the 10^{th} percentile of the consumption growth distribution for each group. The observations conditional on the high variance risk premium are rather evenly spread across recessions in our sample. The observations conditional on the low variance risk premium also appear randomly spread across the sample. The only "unexpected" observation is perhaps October 2008 during the Great Recession, when the risk-neutral variance was high, but the estimated physical variance was also high resulting in a relatively low variance risk premium. However, removing this one data point from our sample would only strengthen the results, as the 10^{th} percentile of the consumption growth distribution conditional on the low variance risk premium would be higher than it is now.

2.2.2 Alternative Consumption Data

Schorfheide, Song, and Yaron (2018) discuss several sources of measurement error in monthly consumption growth data. We therefore consider several robustness checks to the use of monthly consumption data.

First, 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. In order to obtain estimates of the variance risk premium for the 1929-1989 period, for which it is not directly observable, we regress the variance risk premium during 1990-2017 on the sum of squared realized daily returns (including dividends) for the past week, month, and quarter and the price-to-earnings ratio[5]. We then use these projection coefficients to construct estimates of the variance premium for the early part of the sample. Note that high-frequency realized returns are not available then, which is why we use daily returns instead for the full period. 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 on consumption growth and the independent variables in the variance premium regression, using a block length of 5 years.

2. Inside the bootstrap, for observations falling in between 1929 and 1989, we input the fitted variance risk premium from the OLS regression using the independent variables at that point of time, 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}

[5]We prefer this measure to the price-dividend ratio, which exhibits a pronounced time trend during the sample.

percentile of the sampled variance risk premium distribution, and the low variance risk premium, defined as below the 20th 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 (Panel A) 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 10th and 20th percentiles. The 10th 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 60th percentile.

Second, Panels B and C of Figure 3 show analogous results for two alternative data sets. In Panel B, we confirm the results in a sample of quarterly consumption growth data, for a 1947:Q1-2017:Q4 sample and Panel C uses monthly data but over a longer 1959:M1-2017:M12 sample. The beginning of the samples is determined by data availability, and we use the block-bootstrap procedure described above to extend our variance risk premium estimates to the pre-1990 period. The downward quantile shift for the 10th percentile is again statistically significant at the 5% level for both the quarterly and monthly samples and for the 20th percentile it is significant at the 10% level for the quarterly sample.

2.2.3 Alternative Tail Measures

The quantile shift evidence suggests that the conditional skewness of consumption growth is lower when the variance risk premium is high. We start by investigating Kelly's skewness, which is also defined in terms of percentiles of a distribution, namely the 90th percentile+10th percentile-2$\times$50th percentile. An attractive feature of Kelly's skewness is that it has units equal to the original series. If we again define the high variance risk premium as being above its 80th unconditional percentile, the Kelly's skewness of the next month's consumption growth conditional on the high variance risk premium is -0.11% in the US 1990-2017 data, or -1.32 percent at an annual rate. The block-bootstrap standard error computed by re-sampling 10,000 time series of historical length with a block length of 5 years is 0.08%. If we define the low variance risk premium again as being below its 20th unconditional percentile, the Kelly's skewness of the next month's consumption growth conditional on the low variance risk premium is 0.08% with a block-bootstrap standard error of 0.05%. The difference between the two Kelly's skewness estimates is -0.18% (-2.1 percent at an annual rate) with a block-bootstrap standard error of 0.09%, which is statistically significant at the 5% level.

We also investigate the relationship between the variance risk premium and the standard scaled skewness coefficient of consumption growth. To infer a conditional scaled skewness series for monthly consumption growth, we estimate two empirical models which can capture conditional non-Gaussianities. The first model is a regime-switching model of the Hamilton (1989)-type (with regime dependent drifts, regime dependent volatilities and constant transition probabilities). We assume that the within-regime shock distributions are Gaussian, but because the model is essentially a time-varying mixture of normals, it can generate conditional non-Gaussianities. The second model is the BEGE-GARCH model proposed in Bekaert, Engstrom and Ermolov (2015).

Both models are estimated using the standard maximum likelihood procedure for the 1959:M1-2017:M12 sample for which monthly consumption growth data is available. For the regime switching model, the Bayesian information criterion (BIC) picks a 2-regime model specification. The standard scaled conditional consumption growth skewness series is available in closed form (see, e.g., Timmermann, 2000). For the BEGE GARCH model, the optimal specification in terms of BIC imposes a constant "good volatility" parameter, but a time-varying "bad volatility" process. Further details on the models and estimation can be obtained from the authors.

For the regime-switching model, the correlation between the variance risk premium and conditional consumption growth skewness (computed using ex-ante regime probabilities) in the US monthly 1990-2017 sample (where the variance risk premium is directly observed) is -0.27. The block-bootstrap standard error computed by re-sampling 10,000 time series of historical length (the variance risk premium and conditional scaled skewness at each point of time are resampled as pairs) with a block length of 60 months is 0.05. If smoothed regime probabilities are used to compute the conditional consumption growth series skewness, the correlation is -0.32 with a standard error of 0.04. For the 1959:M1-2017:M12 sample, where the variance risk premium values for the pre-1990 period are extrapolated as described above, the correlation with consumption growth skewness is -0.25 with a block-bootstrap standard error of 0.07.

The results are similar for the BEGE-GARCH model. The correlation between the variance risk premium and scaled skewness in the 1990:M1-2017:M12 sample is -0.33 with a block-bootstrap standard error of 0.06. For the 1959:M1-2017:M12 sample the correlation is -0.31 with a block-bootstrap standard error of 0.11. We conclude that there is a statistically strong and robust relationship between the negative tail of the consumption growth distribution and the variance risk premium in financial markets.

2.3 Other Moments

To further assess the fit of the various models, we consider a number of statistics reported in Table 2 (right hand side column). As indicated before, the variance risk premium is measured by the difference between VIX^2 and the expected physical return variance, where the physical variance is obtained using a projection model for variances. The variance premium has a mean of 0.0195, which is rather precisely estimated. For ease of economic interpretation, Panel B of Table 2 also reports the volatility premium fit, which simply replaces the variances by volatilities.

We also report the autocorrelation of the variance risk premium. The premium's autocorrelation is only 0.52 in the monthly data, with a block-bootstrap standard error of only 0.09 (see Panel A).[6] The autocorrelation of the volatility risk premium is even lower at 0.45 (see Panel B). This low value does not reflect measurement error, even though we require an empirical model to measure the physical variance. To justify this claim, we conducted the following exercise. The regression coefficients in this model are asymptotically normally distributed, with the covariance matrix estimated using a Newey-West (1987) procedure with 12 lags. We sample 10,000 possible regression coefficients and reconstruct the conditional physical variance using these resampled coefficients. With this, we recreate 10,000 different sample paths to the variance risk premium time series, keeping the risk-neutral component at its data value. We then construct the distribution of the variance risk premium persistence by computing the autocorrelation in each sample. In 97.5% (99.5%) of the samples the monthly variance risk premium autocorrelation is lower than 0.65 (0.69).

We conduct two further robustness checks for the veracity of our measure of the variance risk premium's persistence. First, Figure 1 illustrates that the variance risk premium is particularly

[6] Note that these autocorrelation block-bootstrap standard errors are computed without breaking up the time dependencies. For example, if the block-bootstrapped time series is observation_1, observation_2, observation_1, observation_2, we construct the series observation_2, observation_3, observation_2, observation_3 to compute the correlation with it in that bootstrap sample.

volatile during the Great Recession, when financial markets were in turmoil, potentially leading to imprecise measurement. Moreover, we see some negative variance risk premium values, which are implausible from the perspective of some, albeit not all, theoretical models. However, even during the pre-Great Recession period (January 1990-November 2007), the variance risk premium autocorrelation is only 0.49; whereas it is 0.57 over the full sample, excluding negative observations. Second, alternative models of the conditional physical variance, such as using an AR(1) realized variance process or a martingale model also deliver autocorrelations of around 0.50 or even lower.

A final powerful implication of options prices is that the risk neutral skewness of equity returns should be lower (or more negative) than the physical skewness, reflecting preferences for insuring against stock market crashes. We report the physical and risk neutral skewness in Panel C of Table 3. Our statistics are based on daily time series estimates for the conditional risk -neutral skewness of non-logged S&P5 returns provided by Bakshi, Crosby, Gao and Zhou (2020). Because they target the middle of the month as the option maturity date, their data reflect maturities between one month and just a few days. Therefore, we create two data series from these data, one using middle-of-the month data, reflecting a 30-day horizon (or the maturity closest to 30 days if unavailable, choosing the shortest maturity on ties), and one extracting the end-of-month data, reflecting a 14/15-day horizon. The two series are over 80% correlated, and their means are not too different. We use the first series for our empirical work, which has a mean of -1.31 (See Panel C of Table 3) for our sample (the mean of the end-of-month series is -1.46).[7] We also report the physical skewness of returns which is indeed much higher at -0.87.

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

We now survey various models in the consumption-based asset pricing literature, focusing on the mechanism they use to generate a meaningful volatility premium. We then show that no model can actually fit our stylized facts.

[7]It would be preferable to compute the risk neutral skewness from logged returns, and we did so, relying on the methodology in Bakshi, Kapadia and Madan (2003), but our data set here is only available from 1996 to 2014. Over this period, risk neutral skewness is somewhat more severe for logged versus non-logged returns.

3.1 The Variance Risk Premium in Existing Consumption-based Asset Pricing Models

Various results in the extant literature (see, among others, Bakshi and Madan, 2006, Chabi-Yo, 2008, 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.[8] Therefore, empirical results described above immediately refute the original formulations of Campbell and Cochrane's (1999) habit model and the first-order approximation to Bansal and Yaron's (2004) long-run risk model which are conditionally Gaussian.[9] An important caveat to this refutation is that it holds because, consistent with the bulk of the consumption-based pricing literature, these models are formulated in discrete time with the decision interval equal to the data frequency (e.g., monthly). In contrast, many option pricing models are formulated in continuous time, and in such a setting, variance risk premiums can arise in conditionally Gaussian models, as temporal aggregation can generate a positive variance risk premium.

We, thus, consider four different macro-finance models which have explicitly sought to explain option prices. Technical details are relegated to Online Appendix II. The first model is the "vol-of-vol" model of Bollerslev, Tauchen, and Zhou (2009), henceforth BTZ. It is important to mention that this paper is one of the first to show that, empirically, the variance risk premium predicts stock returns. 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"). It is the latter variable that drives variance risk premium variation in this article.

The second model, Drechsler and Yaron (2011), DY henceforth, is an extension of the long-run risk model meant to fit variance risk premium features. DY 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.

[8]Online Appendix I proves a related result using conditional variances taken from simple returns (not log-returns): log-normal models can not generate simultaneously a positive variance risk premium and a positive equity premium.

[9]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. 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 models below, we use Wachter's more accurate solution method to solve the model.

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.

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, which probability of occurrence varies through time. While the model was not designed to fit options data, Wachter and Seo (2019) show that this particular specification fits option prices rather well.[10] Importantly, the disaster parameters are inferred from international data, because the U.S. data feature no consumption disaster post the Great Depression.

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 (1999) setup. In this model, consumption growth features "bad" and "good" volatility, with shocks to bad (good) volatility decreasing (increasing) skewness in consumption growth. The modeling of the consumption surplus ratio is analogous to the original Campbell and Cochrane (1999) specification with its shock perfectly correlated with consumption growth, and a non-linear price of risk variable, adjusted for the presence of consumption heteroskedasticity.

The various models' fit of standard salient asset features, such as equity returns, and the risk-free rate, is in Online Appendix II. All the moments shown are annualized monthly values. The sample is 1990 to 2017 corresponding to the available data for the variance risk premium. 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 expected inflation for the corresponding quarter drawn from the Survey of Professional Forecasters (Federal Reserve Bank of Philadelphia); specifically, with π the quarterly inflation forecast, we use $(1+\pi)^{\frac{1}{3}}-1$. The risk-free rate is low on average (0.64%) 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% with a volatility of 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).

Because we use the calibrations/estimations provided in the various original papers, which did not use exactly our data sample, we do not expect the various models to fit all these moments

[10] Another popular rare disasters model is Gabaix (2012). However, in its main specification consumption growth is i.i.d. and, thus, cannot fit the link between consumption growth and variance risk premiums. Gourieroux et al. (2020) features a consumption growth process with gamma-distributed distress, induced by credit defaults.

exactly and we relegate a detailed discussion of the models' fit with respect to traditional moments to Online Appendix II. We note that the the fit of the BTZ model, however, is particularly poor, because their calibration focused on the levels of the unconditional equity premium and risk-free rate. In addition, the zero lower bound on both the consumption growth variance and the variance of consumption growth variance is hit in more than 10% of the simulated time points under the BTZ calibation, leading to significant deviations of simulated asset prices from their theoretical counterparts. Overall, statistically all models are strongly rejected by the data, but we argue that DY, Wachter (2013), and BEGE fit general asset prices reasonably well.

3.2 Fitting Variance Premium Statistics

We now turn to the fit of the models with respect to variance risk premium statistics, which we report in Table 2. To compute the variance risk premium in the models, our first set of statistics use the risk-neutral minus physical annualized variance of the log-return on the aggregate stock market. Following most of the macro-finance literature, we report the statistics for log instead of gross returns, but the differences between the two concepts are economically negligible. As mentioned before, the VIX^2 actually represents the risk-neutral entropy and always exceeds the risk neutral variance in the data (see Martin, 2017). We return to this issue separately below.

Table 2, Panel A, reveals that the variance risk premium is identically equal to zero in the BTZ model, as it is a log-normal model. Both the DY and BEGE models generate meaningful variance risk premiums, but they are still too small relative to the data. Wachter's model generates a much too high average level for the volatility risk premium of 15%.

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 $Var_t^Q(r_{t+1}) - Var_t(r_{t+1})$. The latter is the definition that we use; it is also what is actually priced by the VIX and variance swaps. This "skipping ahead" allows BTZ and DY to generate a positive variance risk premium instead of the zero value under log-normality, 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 volatility risk premium: 12.62% instead of 2.35% in Table 2 (which only takes into the account the "level difference", the $Var_t^Q(r_{t+1}) - Var_t(r_{t+1})$ term). These "drift terms" would naturally arise if the model were specified at a frequency higher than monthly, but the BTZ and DY articles use monthly specifications and data (see section 1.4 in BTZ and section 5.2 in DY), under which such terms do not arise naturally. Importantly, while our results for the DY model mimic the linear approximation of the equilibrium functions the authors employ, properly accommodating non-linearities does not salvage the model: Lorenz, Schmedders, and Schumacher (2020) show explicitly that when solved properly the variance risk premium in the DY model becomes much smaller and behaves procyclically, which is also counterfactual.

All the models also fail to generate a realistic autocorrelation for the variance premium, delivering autocorrelations of 0.90 or higher. The main reason for this is that the models' 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.

Panel C of Table 2 shows that matching skewness in return data is difficult for our 4 representative models. The BTZ model in fact generates slight positive skewness, due to the volatility-of-volatility state variable hitting the zero lower bound too often. The disaster model, perhaps not surprisingly, generates far too severe negative physical and risk neutral skewness. The BEGE model also overshoots, but more modestly so, whereas the DY model does not generate sufficiently negative skewness.

The shape of the risk-neutral return distribution is also reflected in the VIX/SVIX puzzle stressed by Martin (2017). As explained by 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). As a result, the disaster model generates a VIX much higher than the SVIX indicates that left-tail return outcomes are too severe compared to right-tail outcomes in the model. Nonetheless, we find that matching risk neutral skewness is more challenging than the VIX-SVIX moment stressed in Martin (2017).

3.3 Consumption Growth Quantile Shifts and the Variance Risk 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 in the data there is indeed 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 3, 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$^{\text{th}}$, 50$^{\text{th}}$, and 90$^{\text{th}}$ percentile of the consumption growth distribution, conditioning on either the variance premium being above its 80$^{\text{th}}$ unconditional percentile or below its 20$^{\text{th}}$ percentile. We also report the difference in the furthermost right column. Note that we do not include the BTZ model in this analysis, because its variance risk premium is always zero (see Table 2 and the corresponding discussion in section 3.1). Thus, it is not possible to condition on the variance risk premium.[11]

Panel B of Table 3 shows that the DY model 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 the consumption growth volatility and to the intensity of jumps in long-run consumption growth (the predictable part of consumption growth) and volatility (see equation (22) in DY). However, jumps in long-run consumption growth only affect the conditional distribution of consumption growth between time $t+1$ and $t+2$: consumption growth between t and $t+1$ is conditionally Gaussian and, thus, symmetric at time t, making asymmetric percentile shifts impossible. In principle, the DY model can 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

[11]Furthermore, even if we condition on the variance risk premium as defined by BTZ (that is, including the drift term), the model does not generate the time-varying left-tail distribution along with a constant right tail distribution, because its conditional consumption growth distribution is symmetric at all horizons.

probability of a left-skewed jump in long-run consumption growth. However, Panel C of Table 3 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 the predictable part of consumption growth contributes little to the total variation in consumption growth.

Panel D of Table 3 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 3 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 its "bad environment shock." A large bad environment shock increases the shape parameter of the bad consumption shock, shifting the left-tail percentiles of the 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 1).

If we compare directly the asymmetry of the quantile shifts in the left and right tails of the distribution (adding the 10^{th} and 90^{th} percentile shifts), the DY and Wachter models fail to generate asymmetry. The BEGE model does produce asymmetry, but still misses the data shift by a large margin. To investigate this formally, Panel F of Table 3 shows model-implied differences between Kelly's skewnesses of consumption growth conditional on the high and low variance risk premium. Computed as in section 2.2.3, the difference is -0.18% and statistically significantly different from zero in the data. With respect to this metric, the DY model is rejected at the 5% level, Wachter's model at the 10% level, while the BEGE model is not rejected, generating a shift of -0.08%.

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 the DY model features non-Gaussian jumps in the variance of *future* consumption growth and expected consumption growth). The Wachter (2013) model is not able to replicate the link quantitatively, 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 (although the magnitude of the shift is somewhat smaller than in the data). At the same time, Table 2 shows that the model faces other challenges such as a relatively low implied variance risk premium and risk-neutral return skewness. In Online Appendix II, we use formal estimation exercises to show that the disaster and BEGE models cannot quantitatively fit the variance risk premium statistics of Section 3.2 and the conditional consumption growth distribution dynamics as a function of the variance risk premium, under any parameterization.

4. A New Model

Given that extant models cannot fully fit the empirical facts, we develop an alternative model, which is related to but quite different from the BEGE-habit model in Bekaert and Engstrom (2017). The BEGE model comes closest to fitting the stylized facts, but it requires a time-consuming numerical solution procedure, greatly decreasing its practical appeal. We therefore propose a considerably more tractable variant of the habit-BEGE model, which allows for intuitive quasi closed-form asset pricing solutions, but nonetheless fits the data better. Economically, the main innovation of our model relative to Campbell and Cochrane (1999) and Bekaert and Engstrom (2017) is to have both consumption and risk aversion shocks depend on non-Gaussian (BEGE) shocks with time-varying higher order moments but constant prices of risk.

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. In section 4.3 we consider a variant of the model incorporating a preference shock. Section 4.4 verifies whether the model fits several additional stylized facts regarding options data.

4.1 A Tractable BEGE-habit Model

The utility function is standard and given by:

$$E_t \sum_{j=t}^{\infty} \beta^{j-t} \frac{(C_j - H_j)^{1-\gamma} - 1}{1-\gamma}, \tag{2}$$

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:

$$g_{t+1} = \bar{g} + \sigma_{cp}\omega_{p,t+1} - \sigma_{cn}\omega_{n,t+1}, \tag{3}$$

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

$$\begin{aligned} \omega_{p,t+1} &\sim \Gamma(p_t, 1) - p_t, \\ \omega_{n,t+1} &\sim \Gamma(n_t, 1) - n_t, \end{aligned} \tag{4}$$

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. Maintaining the Campbell and Cochrane (1999) assumption of a constant consumption growth mean is actually consistent with the data. The monthly consumption growth autocorrelation is -0.09 in the data with a standard error of 0.10. Schorfheide, Song, and Yaron (2018) argue that the quarterly and lower frequency autocorrelation is robustly positive, while the monthly autocorrelation is negative due to measurement errors in monthly consumption data. Therefore, maintaining a zero correlation is appropriate, and positive serial correlation for quarterly consumption growth automatically ensues from temporally aggregating consumption to lower frequencies.

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):

$$\begin{aligned} p_{t+1} &= \bar{p} + \rho_p(p_t - \bar{p}) + \sigma_{pp}\omega_{p,t+1}, \\ n_{t+1} &= \bar{n} + \rho_n(n_t - \bar{n}) + \sigma_{nn}\omega_{n,t+1}. \end{aligned} \tag{5}$$

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 (5). In particular, the probability density function in 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 in 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) unscaled skewness in consumption growth.[12] To ensure the positivity of p_t (n_t), it must be the case that $\sigma_{pp} < \rho_p$ ($\sigma_{nn} < \rho_n$). These conditions are always satisfied in our estimation.

Following much of the recent habit literature, 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:

$$q_{t+1} = \bar{q} + \rho_q(q_t - \bar{q}) + \sigma_{qp}\omega_{p,t+1} + \sigma_{qn}\omega_{n,t+1}. \tag{6}$$

The q_t-process drives time-variation in the effective risk aversion of the representative agent. While we assume the process to be 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.

Note that the economic mechanism through which this model operates is quite different from Campbell and Cochrane (1999) and Bekaert and Engstrom (2017). While q_t still measures risk aversion, the key drivers of time-variation in risk premiums are p_t and n_t which drive the

[12] Because an increase in n_t also increases the variance, the effect on the scaled skewness coefficient cannot be generally signed. Nonetheless, Bekaert and Engstrom (2017) show in their Appendix C that a risk-averse investor would dislike (prefer) exposure to bad (good) environment shocks relative to Gaussian shocks with the same variance.

correlation between returns and the pricing kernel. Compared to Bekaert and Engstrom (2017), there is no non-linearity emanating from a price of risk process, but the non-Gaussian features of kernel shocks are entirely driven by the BEGE shocks.

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

$$m_{t+1} = m_0 + m_q q_t + m_{\omega,p}\omega_{p,t+1} + m_{\omega,n}\omega_{n,t+1}, \tag{7}$$

where:

$$\begin{aligned} m_0 &= \ln\beta - \gamma\bar{g} + \gamma\bar{q}(1-\rho_q), \\ m_q &= -\gamma(1-\rho_q), \\ m_{\omega,p} &= \gamma(\sigma_{qp} - \sigma_{cp}), \\ m_{\omega,n} &= \gamma(\sigma_{qn} + \sigma_{cn}). \end{aligned} \tag{8}$$

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, and this is guaranteed when σ_{qp} (σ_{qn}) is negative (positive), as expected.

The key formula to derive most of the analytical 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 $g(x) = x + \ln(1-x)$. $g(\theta)$ is always negative assuming $\theta < 1$. Note that while our fundamental shocks $\omega_{p,t+1}$ and $\omega_{n,t+1}$ in (4) 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, rf_t, can be determined the usual way as the negative of the logarithm of the conditional expectation of the exponentiated pricing kernel in (7):

$$rf_t = f_0 + f_q q_t + f_p p_t + f_n n_t, \tag{9}$$

where coefficients are in the Online Appendix III. 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 $f_p < 0$).

Equity pricing

We assume a simple log-dividend growth to price equities:

$$d_{t+1} = \bar{g} + \gamma_g(\sigma_{cp}\omega_{p,t+1} - \sigma_{cn}\omega_{n,t+1}) \tag{10}$$

The price-dividend ratio can be computed as:

$$\frac{P_t}{D_t} = E_t \sum_{i=1}^{\infty} e^{\sum_{j=1}^{i} m_{t+j}+d_{t+j}}. \tag{11}$$

By recursively plugging the dividend dynamics from (10) and the stochastic discount factor from (7) into (11), we obtain an expression for the price-dividend ratio of the following form:

$$\frac{P_t}{D_t} = \sum_{i=1}^{\infty} e^{A_i+B_i p_t+C_i n_t+D_i q_t}, \tag{12}$$

where the A, B, C, and D coefficients follow difference equations described in Online Appendix III. With $pd_t = \ln(\frac{P_t}{D_t})$, we log-linearize equation (12) using a Taylor series approximation to find:

$$pd_t \approx K_0^1 + K_p^1 p_t + K_n^1 n_t + K_q^1 q_t, \tag{13}$$

where we again relegate the actual expressions for the K^1-coefficients to Online Appendix III. Note that we approximate the actual closed-form solution to the model. This differs greatly from Campbell-Shiller (1987) return linearizations or first-order approximations to solve the equilibrium in non-linear models.

The aggregate logged market return is defined as:

$$r_{t+1} = \ln(R_{t+1}) = \ln(\frac{P_{t+1}+D_{t+1}}{P_t}) = \ln(\frac{D_{t+1}}{D_t}\frac{1+\frac{P_{t+1}}{D_{t+1}}}{\frac{P_t}{D_t}}) = d_{t+1} + \ln(1+\frac{P_{t+1}}{D_{t+1}}) - pd_t. \tag{14}$$

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

$$\ln(1+\frac{P_{t+1}}{D_{t+1}}) \approx K_0^2 + K_p^2 p_{t+1} + K_n^2 n_{t+1} + K_q^2 q_{t+1}, \tag{15}$$

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

By plugging (10), (13), and (15) into (14) we obtain the following expression for the log-return:

$$r_{t+1} \approx r_0 + r_p p_t + r_n n_t + r_q q_t + r_{\omega,p}\omega_{p,t+1} + r_{\omega,n}\omega_{n,t+1}, \tag{16}$$

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

Combining (9) and (16), we get the following expression for the equity premium:

$$E_t(r_{t+1} - r_{f,t}) \approx (r_0 - f_0) + (r_p - f_p)p_t + (r_n - f_n)n_t + (r_q - f_q)q_t, \tag{17}$$

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 (7) and the aggregate equity return in (16) 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 (16) it follows that the conditional physical variance of the aggregate market return is:

$$Var_t(r_{t+1}) = r_{\omega,p}^2 p_t + r_{\omega,n}^2 n_t. \tag{18}$$

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:

$$
\begin{aligned}
E_t^Q(r_{t+1}) &= \frac{d}{d\nu}\left[\frac{E_t e^{m_{t+1}+\nu r_{t+1}}}{E_t e^{m_{t+1}}}\right]_{\nu=0}, \\
E_t^Q(r_{t+1}^2) &= \frac{d^2}{d\nu^2}\left[\frac{E_t e^{m_{t+1}+\nu r_{t+1}}}{E_t e^{m_{t+1}}}\right]_{\nu=0}.
\end{aligned} \tag{19}
$$

Plugging m_{t+1} from (7) and r_{t+1} from (16) into (19) results in:

$$
Var_t^Q(r_{t+1}) = (\frac{r_{\omega,p}}{1-m_{\omega,p}})^2 p_t + (\frac{r_{\omega,n}}{1-m_{\omega,n}})^2 n_t. \tag{20}
$$

Bringing (18) and (20) together, we obtain the expression for the variance risk premium:

$$
Var_t^Q(r_{t+1}) - Var_t(r_{t+1}) = r_{\omega,p}^2\left[\frac{1}{(1-m_{\omega,p})^2}-1\right]p_t + r_{\omega,n}^2\left[\frac{1}{(1-m_{\omega,n})^2}-1\right]n_t. \tag{21}
$$

The premium in (21) 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. Thus, theoretically, the model can accommodate negative variance risk premiums, as observed in Figure 1, although in practice the n_t component tends to dominate.[13]

VIX^2 and SVIX^2, as studied by Martin (2017) and which use simple return moments, are, in fact, also available in closed-form in the model. We relegate the derivations and expressions to Online Appendix III. It turns out that SVIX^2 delivers model-implied values very close to the risk-neutral variance for log returns computed as in (20).

4.2 Estimation and Empirical Fit

We estimate the model using the classical minimum distance methodology (Wooldridge, 2002). We match unconditional moments of monthly US real per-capita consumption growth

[13]Given that the equity premium in equation (17) loads positively on both p_t and n_t, the model can also potentially explain the negative coefficient of the "good"/right-tail part of the variance risk premium in excess equity return predictability regressions in Feunou et al. (2019), Kilic and Shaliastovich (2019), and Londono and Xu (2021).

of non-durables and services, annual dividend growth, 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, standard deviation, scaled skewness and excess kurtosis, the standard deviation of annual dividend growth, 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 and skewness 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 in the data), and the risk-neutral skewness of equity returns. To simplify the estimation, we set $\beta = 1.00$ and $p_t = \bar{p}$.[14] We also set $\bar{q} = 1$, because it always enters asset pricing equations multiplied by γ, rendering it unidentified. 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 historical length with a block length of 60 months.

The first column of Table 4 shows the parameter estimates. In terms of preferences, γ is close to 4.0, but, of course, this coefficient no longer represents risk aversion.[15] 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 11.1% 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 show only mild non-Gaussianities. In contrast, as $\bar{n}$ is less than 1, bad environment shocks are very skewed and non-Gaussian. The n_t process is highly persistent. The leverage coefficient for dividend growth (γ_g) is close to 11.00.

The first column of Table 5 shows how the model fits the data moments. On the right hand

[14] Figure 2 suggests that the right-tail of the conditional consumption growth distribution is largely constant. $\omega_{p,t+1}$ also enters the aggregate portfolio return expression in (16). Bakshi, Crosby, and Gao (2021) document that the right tail of the conditional S&P500 return distribution is relatively stable at the monthly frequency.

[15] The median of the local risk aversion process (γe^{q_t}) is 9.80 and its 95$^{\text{th}}$ percentile value is 16.82.

side, we show the data moments and the corresponding standard errors. The bottom lines show the test of the over-identifying restrictions.[16] It is well-known that Wald tests over-reject in small samples. We therefore run a bootstrap, simulating the model at the estimated parameters 10,000 times, re-estimating the model starting from the "true" parameters and recording the test values. There is indeed substantial finite sample bias, with the true p-value for a 5% test being 10.19%. What we report in Table 5 is the empirical p-value. The test does not reject the model at the 5% level. In fact, the model moments are always within one standard error band around the data moments.

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 both the mean of the variance risk premium, and average risk-neutral skewness. The model provides such a good fit while being consistent with the consumption growth moments. This includes generating moderately positive unconditional skewness for consumption growth (it is an insignificant -0.06 in the data). 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 6, Panel B, reveals that the model somewhat underestimates the downward shift in the 10^{th} consumption growth percentile going from low to high variance risk premiums, but it is within two standard errors of the data moment.[17]

The first column of Table 5 shows that the model implied persistence of the variance risk premium is 0.99, much higher than in the data. The difference is both statistically and economically large. The model's failure in fitting the low persistence of the variance risk premium

[16] Under the null, the objective function value follows the distribution $Z'(W^{-1} - W^{-1}J(J'W^{-1}J)^{-1}J'W^{-1})Z$, where Z is a random vector following a multivariate zero-mean Gaussian distribution with the covariance matrix representing the actual covariance matrix of orthogonality conditions, W is the weighting matrix, and J is the Jacobian of the orthogonality conditions (see Jagannathan and Wang, 1996, and Wooldridge, 2002). We estimate the actual covariance matrix by block-bootstrapping 10,000 samples of historical length using a block length of 60 months.

[17] We also verify that the shift is significant when we simulate 10,000 samples of historical length from the model, with the 99% confidence interval for the quantile shift not including zero.

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 (21) in fact only depends on n_t and thus inherits its persistence.

We next investigate if the model is flexible enough to match this moment by explicitly including it into the classical minimum distance estimation. In terms of the parameters, Table 4 shows that the autoregressive coefficient for n_t drops from over 0.99 to less than 0.64. Other parameters that change considerably in relative terms are n's innovation standard deviation that drops by more than 25%; $\bar{p}$, which increases more than five-fold rendering the good environment shocks essentially Gaussian, and the "leverage variable" for dividend growth which increases by more than 35%. The second column of Table 5 shows that the model now replicates the low persistence of the variance risk premium, delivering an autocorrelation of about 0.64 versus 0.52 in the data. While the model is 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 more than 2 standard deviations above the data counterpart, the model-implied value of 0.45 would be consistent with the data if the payout is computed as in Boudoukh et al. (2007) and would be lower than the data if the payout is computed as in Longstaff and Piazzesi (2004). Overall, of the 17 moments used in the estimation, only two moments are outside a two standard error band around the data moments, the other one being the standard deviation of dividend growth. Table 6 (Panel C) indicates that 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.

4.3 Model with a Preference Shock

To resolve the trade-off introduced by fitting the variance risk premium persistence, we now augment the model with a pure preference shock. 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 in equations (18) and (20). Fortunately, such rapidly mean-reverting dynamics is consistent with direct and indirect evidence in the recent asset pricing literature (e.g., Martin, 2017, Eraker and Yang, 2020, and Bekaert, Engstrom, and Xu, 2021).

We do not micro-found this preference shock and it could arise through various mechanisms. For example, Martin and Papadimitriou (2020) show how aggregate risk aversion varies with respect to news shocks in an economy with log-utility agents, which have heterogenous beliefs.

We keep the macroeconomic dynamics as in the benchmark model and introduce a new preference shock, $\omega_{q,t+1}$, into the log-inverse consumption surplus ratio equation (6):

$$q_{t+1} = \bar{q} + \rho_q(q_t - \bar{q}) + \sigma_{qp}\omega_{p,t+1} + \sigma_{qn}\omega_{n,t+1} + \sigma_{qq}\omega_{q,t+1}. \tag{22}$$

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

$$\begin{aligned} &\omega_{q,t+1} \sim \Gamma(s_t, 1) - s_t, \\ &s_{t+1} = \bar{s} + \rho_s(s_t - \bar{s}) + \sigma_{sq}\omega_{q,t+1}. \end{aligned} \tag{23}$$

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 via the classical minimum distance methodology, but also explicitly include the variance risk premium autocorrelation at lag 1 as a moment to match.[18]

Going back to Table 4, the estimated model parameters are in the last column. 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 impact on the variance risk premium. Its persistence is low ($\rho_s = 0.4713$), which helps to fit the low variance risk premium persistence. The parameters imply that unconditionally only 21.35% 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 $\omega_{n,t}$ and the n_t process is now estimated to be less non-Gaussian with its mean 10 times higher than in the benchmark model. 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.

Table 5 shows that, while the model is statistically rejected at the 5% significance level, it fits the various moments rather well: the vast majority of the model-implied moments is within

[18]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.

one standard deviation of the data counterparts. There is not a single moment outside a two standard deviation band around the data counterpart.

Table 6 (Panel D) indicates that this model also matches the left tail consumption growth percentile shifts conditional on the variance risk premium observed in the data. The model-implied shift is now slightly larger than it was in the benchmark model, and close to within one standard error of the data moment. In sum, the model with a preference shock fits all the salient features of the data. Note that while a preference shock could be introduced in the other models as well to help fit the variance risk premium persistence, none of the other models has the mechanisms in place to fit all the other salient stylized facts we study in this article.

4.4 Additional Results on Option Prices

While the variance risk premium has received a lot of attention, the literature has generated a number of additional option pricing puzzles. We now verify whether our benchmark model (see the estimation in Table 4) can match these additional stylized facts, including the volatility smirk (Rubinstein, 1994), highly negative straddle returns (Coval and Shumway, 2001; Bakshi, Crosby, and Gao, 2021, for more recent and detailed evidence), and a very flat term structure of forward variance claims (Dew-Becker et al., 2017; Eraker and Wu, 2017). Our procedure to compute option prices is described in Online Appendix IV.

First, Panel A of Figure 5 plots Black-Scholes-implied volatilities for 1 month put options against their strike prices. The data values are from OptionMetrics.[19] Unfortunately, the coverage only starts in 1996. The model replicates the slope of the implied volatility well, but slightly underestimates the implied volatility level. This is likely due to the fact that OptionMetrics data covers 1996-2017, while the model is estimated for the 1990-2017 sample where the unconditional stock market volatility is lower.

Second, we compute the returns to the strategy of buying 1 month at-the-money puts and calls (a so called straddle). The annualized return for this strategy in the model is -160.48% with a corresponding data value of -187.30%.

[19]To construct the implied volatility pattern in the data, each month we look for the date closest to the end of month on which prices for options with maturities between 28 and 31 days are available (each month all options used to compute the implied volatility curve have the same maturity, but this maturity can vary between 28 and 31 days across the months). We then linearly interpolate the implied volatilities as a function of the strike to the current price ratio for the [0.9;1] strike-to-price ratio interval with a step size of 0.01. The [0.9;1] interval is due to data availability. Plotted data values are averages over monthly interpolated values.

Third, we study the term structure of forward variance claims.[20] An n-period time t forward variance claim is an asset which pays the realized variance between time $t+n$ and $t+n+1$. We express forward variance prices as annualized volatilities. Panel B of Figure 5 illustrates the term structure of 1 month forward variance claims. Data values are from Dew-Becker et al. (2017) (Eraker and Wu, 2017, also document a similar pattern). Unfortunately, the data is only available 1995:M12-2013:M9. The model replicates the shape of the forward variance claim term structure well, but again slightly underestimates its level. This is once more likely due to the fact that the unconditional stock market volatility is lower during 1990:M1-2017:M12 than during 1995:M12-2013:M9.

Lastly, the annual VIX-SVIX difference produced by the model is 1.29%, whereas Martin (2017) reports a 1.41% value. None of the extant models we analyzed before produces values that close (see Online Appendix II).

5. 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 fail to match 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 risk-neutral skewness. Importantly, they also cannot match the strong link between the left tail of the consumption growth distribution and the magnitude of the variance risk premium. 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 variance premium in this model intuitively loads positively on bad and negatively on good uncertainty. The model fits the data even better when a small sentiment shock is allowed. The model also fits volatility smirks in options, highly negative straddle returns, a slight upward slope in forward variances, and Martin's (2017) VIX-SVIX puzzle.

While we examine models representing the various main asset paradigms, research is continually evolving and alternative models may likewise be successful. For instance, Schreindorfer

[20] Dew-Becker et al. (2017) show that this term structure is informative about different mechanisms of risk premia, as many equilibrium models fail to generate it. Drechsler (2015) argues that the Dew-Becker et al. (2017) results are not robust in more recent sample periods with more active variance swap trading.

(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 (2021) show that extant representative agent models fare poorly in matching this evidence. Interestingly, Chabi-Yo and Loudis (2021) show that the new BEGE model we propose in this paper does well in this exercise reproducing the equity premium composition both unconditionally and conditionally.

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 (2021) 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 is potentially consistent with her findings.

References

Acemoglu, D., Ozdaglar, A., and Tahbaz-Salehi, A. (2017) Microeconomic origins of macroeconomic tail risks, *American Economic Review* 107, 54–108.

Adrian, T., Boyarchenko, N., and Giannone, D. (2019) Vulnerable growth, *American Economic Review* 109, 1263–1289.

Ait-Sahalia, Y., Karaman, M., and Mancini, L. (2018) The term structure of variance swaps and risk premia, *Working paper.*

Andersen, T., Bollerslev, T., Diebold, F., and Labys, P. (2003) Modeling and forecasting realized volatility, *Econometrica* 71, 579–625.

Backus, D., Chernov, M., and Martin, I. (2011) Disasters implied by equity index options, *Journal of Finance* 66, 1969–2012.

Bakshi, G., Crosby, J., and Gao, X. (2021) Dark matter in (volatility and) equity option risk premiums, *Working paper.*

Bakshi, G., Crosby, J., Gao, X., and Zhou, W. (2021) Implications of a general formula for the expected return on the market, *Working paper.*

Bakshi, G., Kapadia, N., and Madan, D. (2003) Stock return characteristics, skew laws, and the differential pricing of individual equity options, *Review of Financial Studies* 16, 101–143.

Bakshi, G., and Madan, D. (2000) Spanning and derivative-security valuation, *Journal of Financial Economics* 55, 205–238.

Bakshi, G., and Madan, D. (2006) A theory of volatility spreads, *Management Science* 52, iv–1982.

Bansal, R., and Yaron, A. (2004) Risks for the long run: a potential resolution of asset pricing puzzles, *Journal of Finance* 59, 1481–1509.

Baqaee, D., and Farhi, E. (2019) The macroeconomic impact of microeconomic shocks: beyond Hulten's theorem, *Econometrica* 87, 1155–1203.

Barndorff-Nielsen, O., and Shephard, N. (2002) Estimating quadratic variation using realized variance, *Journal of Applied Econometrics* 17, 457–477.

Barro, R., and Ursua, J. (2008) Consumption disasters in the twentieth century, *American Economic Review* 98, 58–63.

Beason, T., and Schreindorfer, D. (2021) On sources of risk premia in representative agent models, *Journal of Political Economy*, forthcoming.

Bekaert, G., and Engstrom, E. (2017) Asset return dynamics under habits and bad environment-good environment fundamentals, *Journal of Political Economy* 125, 713–760.

Bekaert, G., Engstrom, E., and Xu, N. (2021) The time variation in risk appetite and uncertainty, *Management Science*, forthcoming.

Bekaert, G., and Hoerova, M. (2014) The VIX, the variance premium and stock market volatility, *Journal of Econometrics* 183, 181–192.

Bekaert, G. , Engstrom, E., and Ermolov, A. (2015) Bad environments, good environments: a non-gaussian asymmetric volatility model, *Journal of Econometrics* 186, 258–275.

Bekaert, G., Engstrom, E., and Ermolov, A. (2021) Uncertainty and the economy: the evolving distributions of aggregate supply and demand shocks, *Working paper.*

Bollerslev, T., Tauchen, G., and Zhou, H. (2009) Expected stock returns and variance risk premia, *Review of Financial Studies* 22, 4463–4492.

Bollerslev, T., and Todorov, T. (2011) Tails, fears, and risk premia, *Journal of Finance* 66, 2165–2211.

Boudoukh, J., Michaely, R., Richardson, M., and Roberts, M. (2007) On the importance of measuring payout yield: implications for empirical asset pricing, *Journal of Finance* 62, 877–915.

Britten-Jones, M., and Neuberger, A. (2000) Option prices, implied price processes, and stochastic volatility, *Journal of Finance* 55, 839–866.

Campbell, J., and Cochrane, J. (1999) By force of habit: a consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* 107, 205–221.

Campbell, J., and Shiller, R. (1988) Stock prices, earnings, and expected dividends, *Journal of Finance* 43, 661–676.

Chabi-Yo, F. (2008) Conditioning information and variance bound on pricing kernels with higher-order moments: theory and evidence, *Review of Financial Studies* 21, 181–231.

Chabi-Yo, F., and J. Loudis (2021) A decomposition of conditional risk premia and implications for representative agent models, *Working paper.*

Corsi, F. (2009) A simple approximate long-memory model of realized volatility, *Journal of Financial Econometrics* 7, 174–196.

Coval, J., and Shumway, T. (2001) Expected option returns, *Journal of Finance* 56, 983–1009.

Dew-Becker, I., Giglio, S., Le, A., and Rodriguez, M. (2017) The price of variance risk, *Journal of Financial Economics* 123, 225–250.

Drechsler, I., and Yaron, A. (2011) What's vol got to do with it, *Review of Financial Studies* 24, 1–45.

Drechsler, I. (2015) Discussion of The price of variance risk by Dew-Becker, I., Giglio, S., Le, A., and Rodriguez, M. *NBER Asset Pricing Fall Meeting*, available at `https://sites.google.com/site/idrechsl/`.

Drechsler, I. (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, D. (2011) General equilibrium pricing of options with habit formation and event risks, *Journal of Financial Economics* 99, 400–426.

Epstein, L., and Zin, S. (1989) Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework, *Econometrica* 57, 937–969.

Eraker, B., and Yang, A. (2020) The price of higher order catastrophe insurance: the case of VIX options, *Working paper.*

Eraker, B., and Wu, Y. (2017) Explaining the negative returns to volatility claims: an equilibrium approach, *Journal of Financial Economics* 125, 72–98.

Feunou, B., Lopez Aliouchkin, R., Tedongap, R., and Lai, X. (2019) Loss uncertainty, gain uncertainty, and expected stock returns, *Working paper.*

Gabaix, X. (2012) Variable rare disasters: an exactly solved framework for ten puzzles in macro-finance, *Quarterly Journal of Economics* 127, 645–700.

Glosten, L., Jagannathan, R., and Runkle, D. (1993) On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* 48, 1779–1801.

Gourieroux, C., and Jasiak, J. (2006) Autoregressive gamma processes, *Journal of Forecasting* 25, 129–152.

Gourieroux, C., Monfort, A., Mouabbi, S., and Renne, J. (2021) Disastrous defaults, *Review of Finance* 25, 1727–1772.

Gürkaynak, R., Sack, B., and Wright, J. (2010) The TIPS yield curve and inflation compensation, *American Economic Journal: Macroeconomics* 2, 70–92.

Hamilton, J. (1989) A new approach to the economic analysis of non-stationary time series and the business cycle, *Econometrica* 57, 357–384.

Ibragimov, R., and Müller, U. (2010) t-statistic based correlation and heterogeneity robust inference, *Journal of Business and Economic Statistics* 28, 453–468.

Jagannathan, R., and Wang, Z. (1996) The conditional CAPM and the cross-section of expected returns, *Journal of Finance* 51, 3–53.

Jiang, G., and Tian, Y. (2005) The model-free implied volatility and its information content, *Review of Financial Studies* 18, 1305–1342.

Kilic, M., and Shaliastovich, I. (2019) Good and bad variance premia and expected returns, *Management Science* 65, 2445–2945.

Londono, J., and Xu, N. (2021) The global determinants of international equity risk premiums, *Working paper.*

Longstaff, F., and Piazzesi, M. (2004) Corporate earnings and the equity premium, *Journal of Financial Economics* 74, 401–421.

Lorenz, F., Schmedders, K., and Schumacher, M. (2020) Nonlinear dynamics in conditional volatility, *Working paper.*

Martin, I. (2017) What is the expected return on the market?, *Quarterly Journal of Economics* 132, 367–433.

Martin, I., and Papadimitriou, D. (2020) Sentiment and speculation in a market with heterogeneous beliefs, *Working paper*.

Pohl, W., Schmedders, K., and Wilms, O. (2018) Higher order effects in asset pricing models with long-run risks, *Journal of Finance* 73, 1061–1111.

Rietz, T. (1988) The equity risk premium: a solution, *Journal of Monetary Economics* 22, 117–131.

Routledge, B., and Zin, S. (2010) Generalized disappointment aversion and asset prices, *Journal of Finance* 65, 1303–1332.

Rubinstein, M. (1994) Implied binomial trees, *Journal of Finance* 49, 771–818.

Schorfheide, F., Dongho, S., and Yaron, A. (2018) Identifying long-run risks: a Bayesian mixed-frequency approach, *Econometrica* 86, 617–654.

Schreindorfer, D. (2020) Macroeconomic tail risks and asset prices, *Review of Financial Studies* 33, 3541–3582.

Timmermann, A. (2000) Moments of markov switching models, *Journal of Econometrics* 96, 75–111.

Wachter, J. (2005) Solving models with external habit, *Finance Research Letters* 2, 210–226.

Wachter, J. (2013) Can time-varying risk of rare disasters explain aggregate stock market volatility?, *Journal of Finance* 68, 987–1035.

Wachter, J., and Seo, S. (2019) Option prices in a model with stochastic disaster risk, *Management Science* 65, 3449–3469.

Whaley, R. (2000) The investor fear gauge, *The Journal of Portfolio Management* 26, 12–17.

Whelan, K. (2000) A guide to the use of chain aggregated NIPA data, *Review of Income and Wealth* 48, 217–233.

Wooldridge, J. (2002) Econometric analysis of cross section and panel data, MIT Press, 445–446.

Zviadadze, I. (2021) Term structure of risk in expected returns, *Review of Financial Studies* 34, 6032–6086.

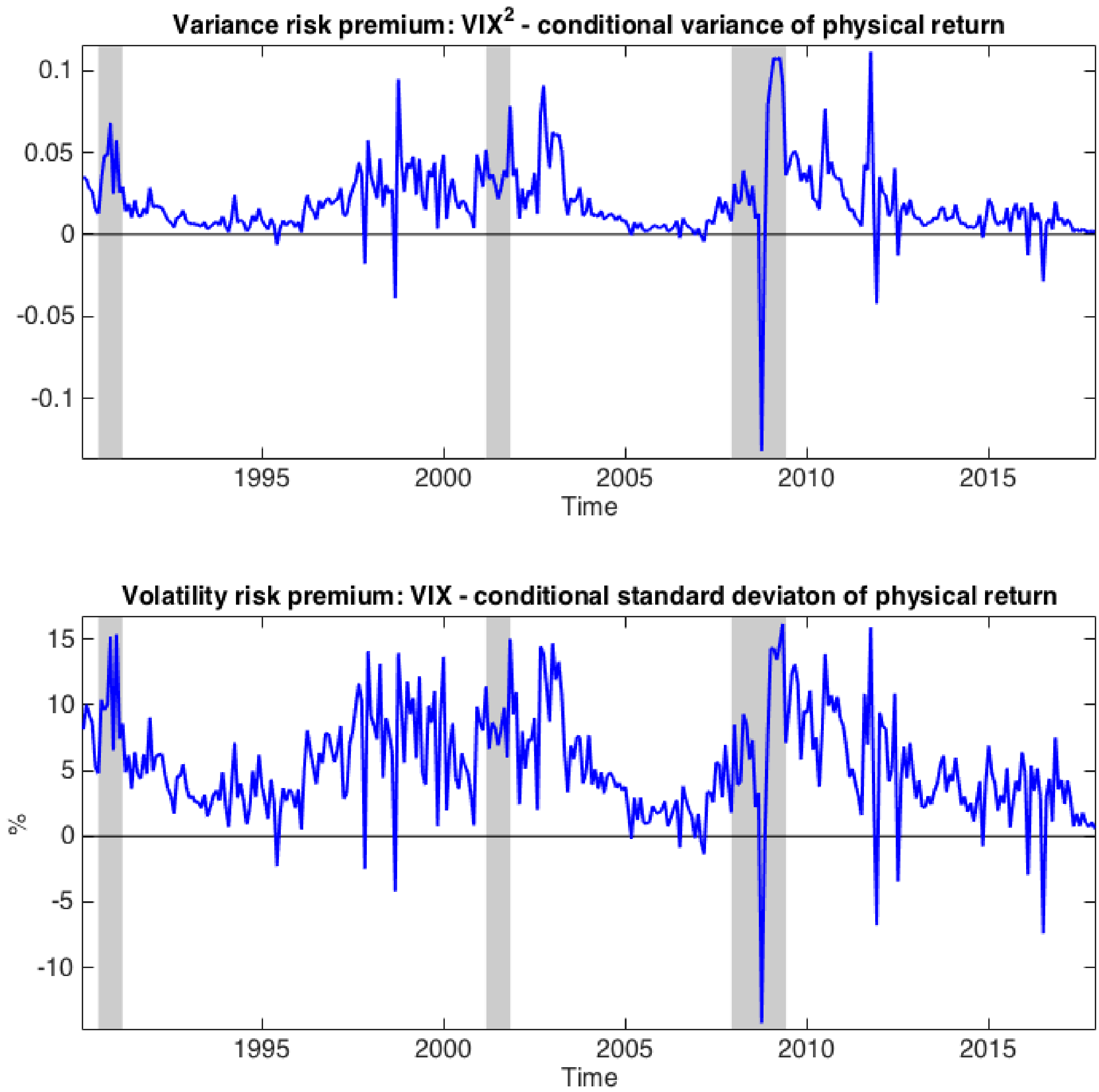

Figure 1: Variance and Volatility Risk Premia. Data is monthly from January 1990 to December 2017. The variance risk premium is the difference between annualized VIX^2 and expected physical variance of the $S\&P500$ return from Corsi (2009)-type model. The volatility risk premium is the difference between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

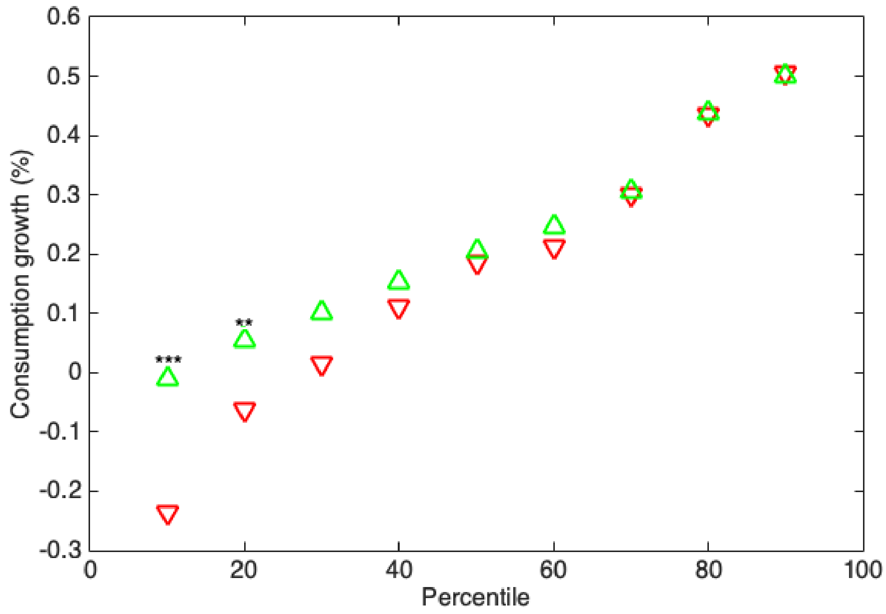

Figure 2: Percentiles of Next Month Consumption Growth Conditional on the Current Variance Risk Premium in the US data. Data is monthly from January 1990 to December 2017. Green up-pointing triangles correspond to quantiles conditional on the low variance risk premium, and red down-pointing triangles 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 the 80^{th} unconditional percentile in the data and low variance risk premium as the variance risk premium below the 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 a block length of 60 months.

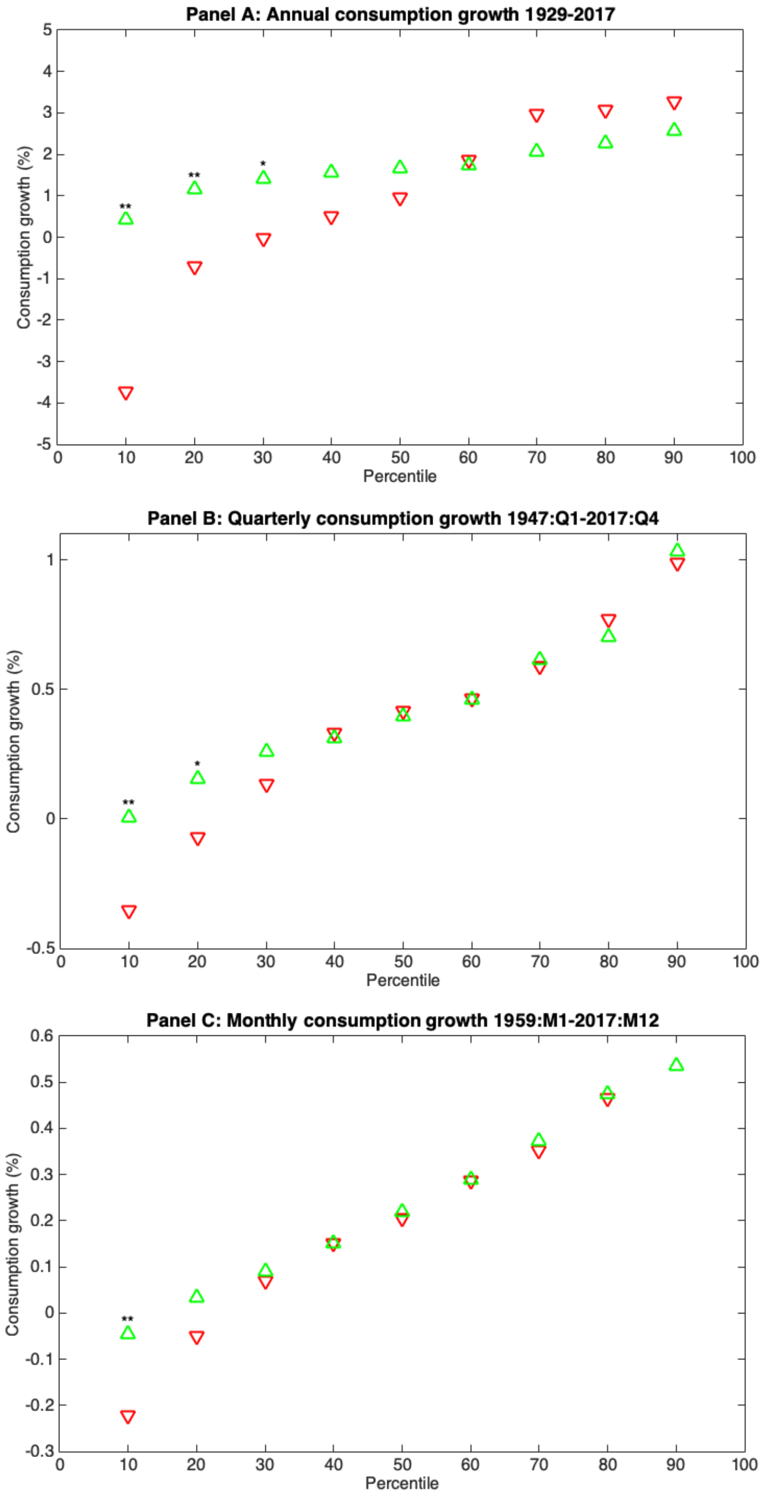

Figure 3: Percentiles of Next Period Consumption Growth Conditional on the Current Variance Risk Premium in the US Data: Alternative Samples. Green up-pointing triangles correspond to quantiles conditional on the low variance risk premium, and red down-pointing triangles correspond to quantiles conditional on the high variance risk premium. High variance risk premium is defined as the variance risk premium above the 80$^{\text{th}}$ unconditional percentile in the data and low variance risk premium as the variance risk premium below the 20$^{\text{th}}$ unconditional percentile in the data. ** and * correspond to statistical significance at the 5% and 10% level. Statistical significance is determined based on block-bootstrap computed by re-sampling 10,000 time series of historical length with a block length of 60 months.

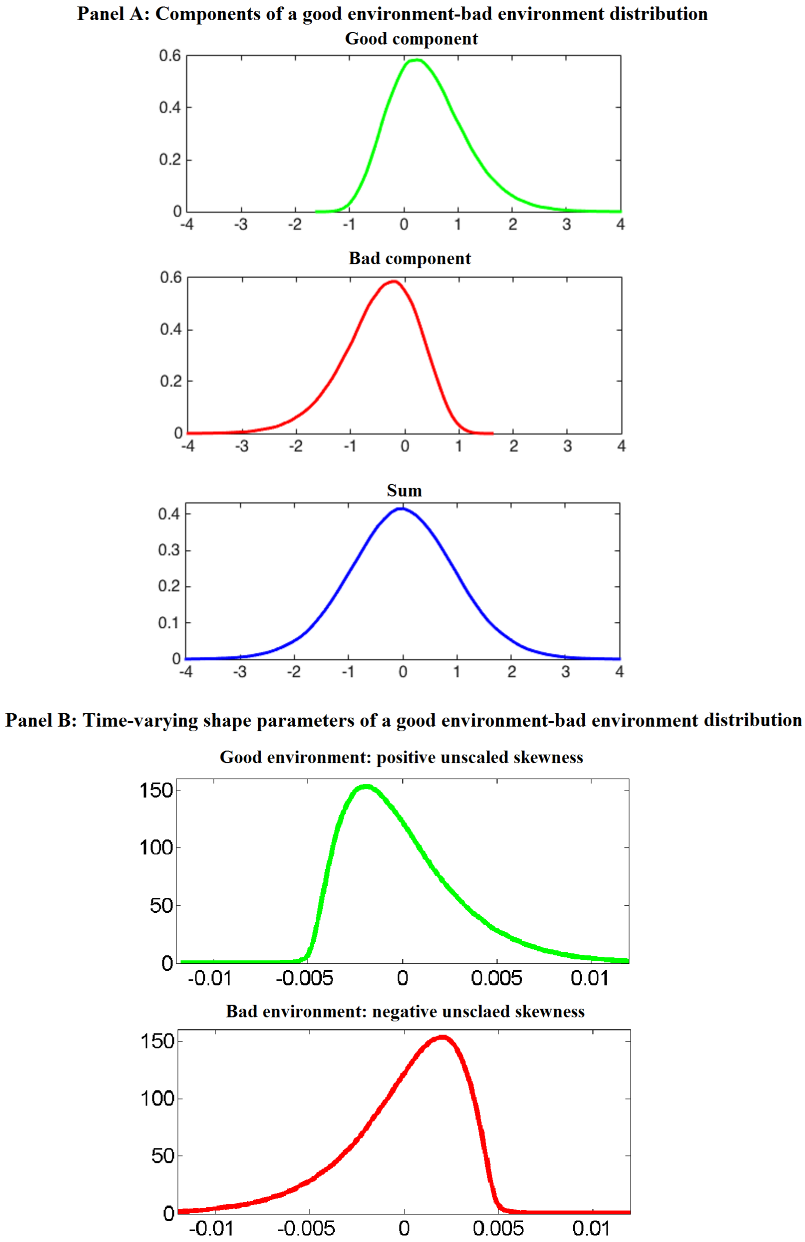

Figure 4: Bad Environment - Good Environment Distribution. Graphs are probability density functions.

Figure 5: Out-of-sample Option Moments Fit. Panel A plots Black-Scholes-implied volatilities for 1 month put options against the strike price. Volatilities are annualized. The model values are computed from 10,000 monthly observations simulated under the estimated model parameters. Data 1996:M1-2017:M12 averages are from OptionMetrics. Panel B plots 1 month forward realized variance claim prices expressed as annualized volatilities. Data 1995:M12-2013:M9 averages are from Dew-Becker et al. (2017).

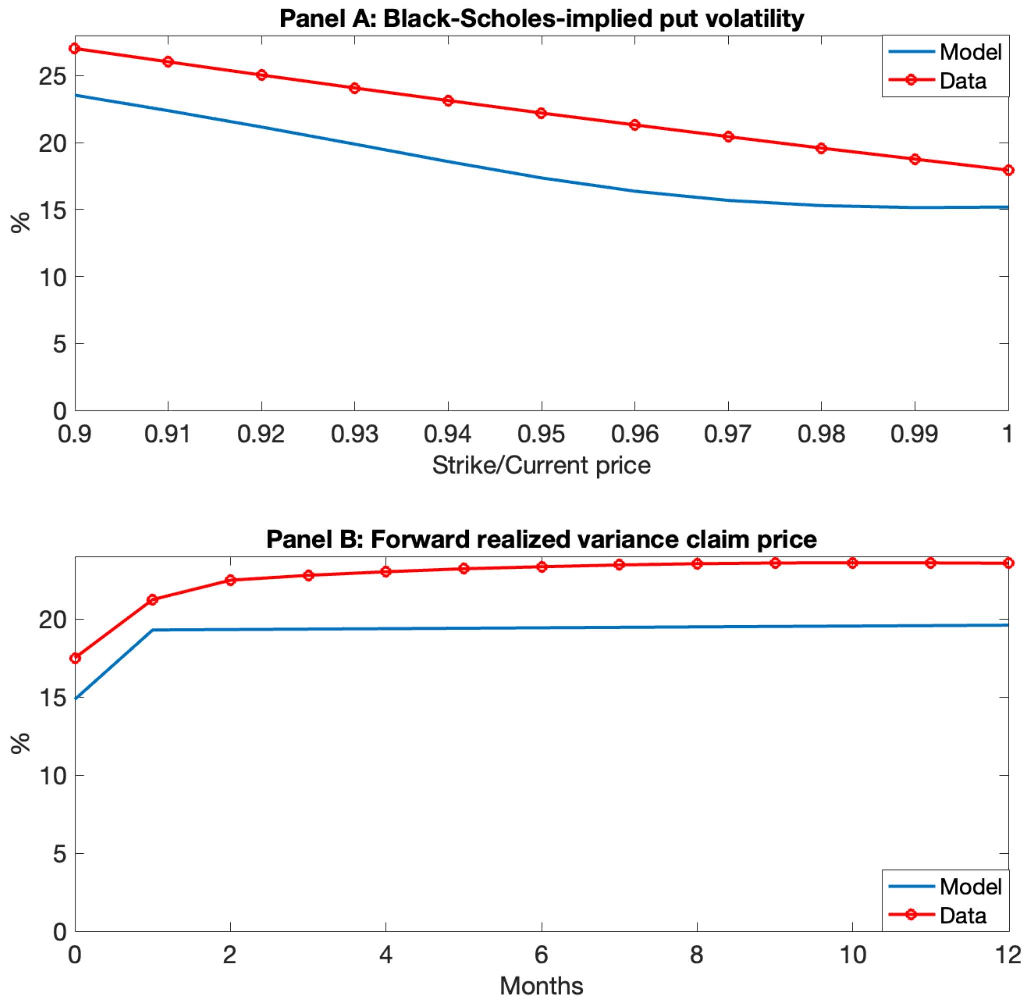

Table 1: Percentiles of Next Month Consumption Growth Conditional on the Current Variance Risk Premium in the US data. Data is monthly from January 1990 to December 2017. In Panels A, B, and C the variance risk premium is the difference between VIX^2 and the expected physical variance from a Corsi (2009)-type model. In Panel D the variance risk premium is computed using conditional physical variance from GJR-GARCH model of Glosten, Jagannathan and Runkle (1993). In Panels D and E, the high variance risk premium is defined as the variance risk premium above the 80^{th} unconditional percentile in the data and low variance risk premium as the variance risk premium below the 20^{th} unconditional percentile in the data. In Panel E, the consumption growth is for the month after the variance risk premium month. 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 80^{th} and 20^{th} unconditional percentiles, respectively			
	High variance risk premium	Low variance risk premium	High-low Difference
10^{th} percentile	-0.24%	-0.01%	-0.23%***
	(0.07%)	(0.04%)	(0.08%)
50^{th} percentile	0.19%	0.20%	-0.01%
	(0.04%)	(0.03%)	(0.05%)
90^{th} percentile	0.51%	0.50%	0.01%
	(0.05%)	(0.04%)	(0.06%)
Panel B: 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%)	(0.04%)
50^{th} percentile	0.18%	0.20%	-0.02%
	(0.03%)	(0.03%)	(0.04%)
90^{th} percentile	0.51%	0.50%	0.01%
	(0.06%)	(0.02%)	(0.06%)
Panel C: 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	-0.27%	-0.02%	-0.26%***
	(0.08%)	(0.05%)	(0.09)
50^{th} percentile	0.15%	0.21%	-0.06%
	(0.05%)	(0.04%)	(0.07)
90^{th} percentile	0.46%	0.51%	-0.05%
	(0.07%)	(0.04%)	(0.08%)
Panel D: 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	-0.23%	-0.06%	-0.18%**
	(0.07%)	(0.07%)	(0.09%)
50^{th} percentile	0.21%	0.18%	0.02%
	(0.03%)	(0.03%)	(0.04%)
90^{th} percentile	0.55%	0.49%	0.06%
	(0.09%)	(0.04%)	(0.09%)

Panel E: Below the 10^{th} percentile consumption growth observations			
Conditional on the high variance risk premium		Conditional on the low variance risk premium	
Variance risk premium month	Consumption growth	Variance risk premium month	Consumption growth
August 2001	-0.98%	September 2016	-0.12%
December 2000	-0.60%	October 2008	-0.11%
September 1990	-0.33%	August 2005	-0.10%
December 1990	-0.30%	March 2013	-0.07%
November 2008	-0.28%	June 1995	-0.05%
January 2009	-0.27%	November 2011	-0.03%
February 2009	-0.24%	August 2014	-0.01%

Table 2: 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 the 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 the physical log-return standard deviation by the value implied by a Corsi (2009)-type model using high-frequency data. Model values are obtained by simulating 100,000 monthly observations for each model. The data block-bootstrap standard error in parentheses are obtained by re-sampling 10,000 time series of 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	Data
Mean	0.0000	0.0093	0.1208	0.0064	0.0195
					(0.0026)
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	Data
Mean	0.00%	2.35%	15.08%	1.87%	5.36%
					(0.58%)
Lag 1 autocorrelation	1.00	0.89	0.99	0.97	0.45
					(0.10)
Panel C: Skewness					
	BTZ	DY	Wachter	BEGE	Data
$Skw(r_t)$	0.1698	-0.3998	-20.3167	-1.5060	-0.8680
					(0.2558)
$Skw^Q(R_t)$	0.1147	-0.6907	-26.4401	-1.7739	-1.3075
					(0.1213)

Table 3: Percentiles of Next Month Consumption Growth Conditional on the Current Variance Risk Premium Fit: Extant Model Implications. 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 the 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 the 80^{th} unconditional percentile and low variance risk premium as the variance risk premium below the 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). Data block-bootstrap standard errors in parentheses are obtained by re-sampling 10,000 time series of historical length with a block length of 60 months. *** indicates the statistical significance of the data value at the 1% level. For model-implied values, ***, **, and * indicate that the value is more than 2.58, 1.96, and 1.65 standard deviations away from the data counterpart, respectively.

Panel A: US 1990-2017			
	High variance risk premium	Low variance risk premium	High-Low difference
10^{th} percentile	-0.24%	-0.01%	-0.23%***
	(0.07%)	(0.04%)	(0.08%)
50^{th} percentile	0.19%	0.20%	-0.01%
	(0.04%)	(0.03%)	(0.05%)
90^{th} percentile	0.51%	0.50%	0.01%
	(0.05%)	(0.04%)	(0.06%)
Panel B: DY			
	High variance risk premium	Low variance risk premium	High-Low Difference
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			
	High variance risk premium	Low variance risk premium	High-Low Difference
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			
	High variance risk premium	Low variance risk premium	High-Low Difference
10^{th} percentile	-0.55%	-0.53%	$-0.01\%^{***}$
50^{th} percentile	0.20%	0.21%	-0.01%
90^{th} percentile	0.94%	0.94%	0.00%
Panel E: BEGE			
	High variance risk premium	Low variance risk premium	High-Low Difference
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: Difference between Kelly's skewnesses of consumption growth conditional on the high and low variance risk premium			
US: 1990-2017	DY	Wachter	BEGE
-0.18%**	0.00%**	-0.01%*	-0.08%
(0.09%)			

Table 4: 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 a block length of 60 months. A diagonal weighting matrix is used. Standard errors are in parentheses.

	Benchmark model	Fitting variance risk premium persistence	
		Benchmark model	Preference shock model
	Preferences		
β	1.0000	1.0000	1.0000
	(fixed)	(fixed)	(fixed)
γ	3.8746	3.1110	3.4381
	(0.8003)	(0.6958)	(0.7914)
$\bar{q}$	1.0000	1.0000	1.0000
	(fixed)	(fixed)	(fixed)
ρ_q	0.9940	0.9995	0.9991
	(0.0139)	(0.0131)	(0.0172)
σ_{qp}	-0.0013	-0.0014	-0.0029
	(0.0006)	(0.0006)	(0.0012)
σ_{qn}	0.1067	0.1129	0.0300
	(0.0350)	(0.0306)	(0.0129)
σ_{qq}			0.1984
			(0.0471)
$\bar{s}$			0.0042
			(0.0016)
ρ_s			0.4713
			(0.0214)
σ_{sq}			0.0116
			(0.0072)
	Macroeconomic dynamics		
$\bar{g}$	0.0021	0.0016	0.0020
	(0.0002)	(0.0002)	(0.0002)
σ_{cp}	0.0008	0.0003	0.0013
	(0.0003)	(0.0002)	(0.0003)
σ_{cn}	0.0033	0.0041	0.0019
	(0.0005)	(0.0005)	(0.0005)
$\bar{p}$	8.0410	44.5781	2.3726
	(3.2068)	(7.9504)	(0.8219)
$\bar{n}$	0.0600	0.0459	0.6450
	(0.0241)	(0.0206)	(0.1753)
ρ_n	0.9915	0.6405	0.9992
	(0.0183)	(0.0327)	(0.0196)
σ_{nn}	0.0271	0.0200	0.0236
	(0.0106)	(0.0085)	(0.0104)
γ_g	10.9229	14.9282	13.0802
	(2.9444)	(3.5109)	(3.3853)

Table 5: Classical Minimum Distance Moments Fit. Values are monthly. AC_1 is lag 1 autocorrelation. The weighting matrix is 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 diagonal weighting matrix is used. In 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. p-values for the overidentification test-statistics are computed by simulating 10,000 time series of historical length under the estimated parameters, re-estimating the model on each simulated time series and obtaining the corresponding overidentification statistic, and using critical values from this distribution. *, ** and *** correspond to a rejection at the 10%, 5%, and 1% significance levels, respectively.

Moment	Benchmark model	Fitting variance risk premium persistence		Data	Data standard error
		Benchmark model	Preference shock model		
Macroeconomic dynamics					
$E(g_t)$	0.0021	0.0016	0.0020	0.0020	0.0002
$Std(g_t)$	0.0024	0.0024	0.0025	0.0025	0.0002
$Skw(g_t)$	0.2920	-0.1914	0.0228	-0.0638	0.3654
$ExKur(g_t)$	1.8104	2.2923	2.3479	2.2779	1.0635
$Std(d_t)$ (annual)	0.0926	0.1278	0.1116	0.0771	0.0202
Risk-free rate					
$E(rf_t)$	0.0001	0.0004	0.0005	0.0005	0.0004
$Std(rf_t)$	0.0017	0.0014	0.0016	0.0016	0.0001
$AC_1[rf_t]$	0.9874	0.9604	0.9738	0.9735	0.0093
Equity					
$E(r_t - rf_t)$	0.0038	0.0031	0.0032	0.0057	0.0027
$Std(r_t)$	0.0429	0.0477	0.0391	0.0426	0.0039
$Skw(r_t)$	-1.0295	-1.2684	-0.9323	-0.8680	0.2258
$E(pd_t)$	6.3722	6.2982	6.3920	6.3969	0.0644
$Std(pd_t)$	0.2582	0.4497	0.3249	0.2796	0.0377
$AC_1[pd_t]$	0.9975	0.9995	0.9991	0.9917	0.0079
Options					
$Var^Q(r_t) - Var(r_t)$	0.0012	0.0010	0.0015	0.0016	0.0004
$Skw^Q(R_t)$	-1.3095	-1.3892	-1.3015	-1.3075	0.1213
$AC_1[Var^Q(r_t) - Var(r_t)]$	0.9915 (not in the estimation)	0.6405	0.5152	0.5197	0.0943
Overidentification test-statistic	11.5829*	55.5514***	7.1334**		
p-value	9.91%	0.00%	1.17%		

Table 6: 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 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 the 80^{th} unconditional percentile and low variance risk premium as the variance risk premium below the 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 1990M1-2017M12			
	High variance risk premium	Low variance risk premium	High-Low difference
10^{th} percentile	-0.24%	-0.01%	-0.23%***
			(0.08%)
50^{th} percentile	0.19%	0.20%	-0.01%
			(0.05%)
90^{th} percentile	0.51%	0.50%	0.01%
			(0.06%)
Panel B: Benchmark model			
	High variance risk premium	Low variance risk premium	High-Low difference
10^{th} percentile	-0.11%	-0.05%	-0.07%
50^{th} percentile	0.19%	0.19%	0.00%
90^{th} percentile	0.53%	0.52%	0.01%
Panel C: Benchmark model fitting the variance risk premium persistence			
	High variance risk premium	Low variance risk premium	High-Low difference
10^{th} percentile	-0.23%	-0.10%	-0.12%
50^{th} percentile	0.14%	0.17%	-0.03%
90^{th} percentile	0.47%	0.49%	-0.02%
Panel D: Preference shock model			
	High variance risk premium	Low variance risk premium	High-Low difference
10^{th} percentile	-0.17%	-0.04%	-0.14%
50^{th} percentile	0.20%	0.17%	0.03%
90^{th} percentile	0.53%	0.49%	0.04%

Online 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. *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:

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

the physical return variance can be computed as:

$$Var_t(R_{t+1}) = E_t(R^2_{t+1}) - E_t(R_{t+1})^2 = e^{2\bar{r}_t + 2\sigma^2_{r,t}} - e^{2\bar{r}_t + \sigma^2_{r,t}} = e^{2\bar{r}_t + \sigma^2_{r,t}}(e^{\sigma^2_{r,t}} - 1). \tag{24}$$

The risk-neutral expectation is:

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

The risk-neutral variance can now be computed as:

$$\begin{aligned}
&E_t(M_{t+1}) = E_t(e^{m_{t+1}}) = e^{\bar{m}_t + 0.5\sigma^2_{m,t}}, \\
&E_t(M_{t+1}R_{t+1}) = E_t(e^{m_{t+1}+r_{t+1}}) = e^{\bar{m}_t + \bar{r}_t + 0.5\sigma^2_{m,t} + 0.5\sigma^2_{r,t} + \sigma_{mr,t}}, \\
&E^Q_t(R_{t+1}) = e^{\bar{r}_t + 0.5\sigma^2_{r,t} + \sigma_{mr,t}}, \\
&E_t(M_{t+1}R^2_{t+1}) = E_t(e^{m_{t+1}+2r_{t+1}}) = e^{\bar{m}_t + 2\bar{r}_t + 0.5\sigma^2_{m,t} + 2\sigma_{mr,t} + 2\sigma^2_{r,t}}, \\
&E^Q_t(R^2_{t+1}) = e^{2\bar{r}_t + 2\sigma_{mr,t} + 2\sigma^2_{r,t}}, \\
&Var^Q_t(R_{t+1}) = E^Q_t(R^2_{t+1}) - E^Q_t(R_{t+1})^2 = e^{2\bar{r}_t + 2\sigma_{mr,t} + 2\sigma^2_{r,t}} - e^{2\bar{r}_t + \sigma^2_{r,t} + 2\sigma_{mr,t}} = \\
&e^{2\bar{r}_t + \sigma^2_{r,t} + 2\sigma_{mr,t}}(e^{\sigma^2_{r,t}} - 1).
\end{aligned} \tag{25}$$

Comparing the final line of (25) to the final line of (24), we can see that in order for the variance risk premium to be positive ($Var_t^Q(R_{t+1}) > Var_t(R_{t+1})$) 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:

$$\begin{aligned}
&E_t(M_{t+1}R_{t+1}) = 1, \\
&e^{\bar{m}_t+\bar{r}_t+0.5\sigma^2_{r,t}+0.5\sigma^2_{m,t}+\sigma_{mr,t}} = 1, \\
&e^{\bar{m}_t+\bar{r}_t+0.5\sigma^2_{r,t}+0.5\sigma^2_{m,t}} = e^{-\sigma_{mr,t}}, \\
&E_t(R_{t+1})E_t(M_{t+1}) = e^{-\sigma_{mr,t}}, \\
&\ln E_t(R_{t+1}) - r_{f,t} = -\sigma_{mr,t},
\end{aligned} \tag{26}$$

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

Online Appendix II: Extant Models

Bollerslev, Tauchen, and Zhou (2009)

The utility function is:

$$U_t = \left[(1-\delta)C_t^{\frac{1-\gamma}{\theta}} + \delta(E_t\left[U_{t+1}^{1-\gamma}\right])^{\frac{1}{\theta}}\right], \tag{27}$$

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:

$$\begin{aligned}
&g_{t+1} = d_{t+1} = \mu_g + \sigma_{g,t}z_{g,t+1}, \\
&\sigma^2_{g,t+1} = a_\sigma + \rho_\sigma\sigma^2_{g,t} + \sqrt{q_t}z_{\sigma,t+1}, \\
&q_{t+1} = a_q + \rho_q q_t + \phi_q\sqrt{q_t}z_{q,t+1}, \\
&z_{g,t+1} \sim \mathcal{N}(0,1), z_{\sigma,t+1} \sim \mathcal{N}(0,1), z_{q,t+1} \sim \mathcal{N}(0,1),
\end{aligned} \tag{28}$$

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.

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.

Drechsler and Yaron (2011)

The utility function is:

$$U_t = \left[(1-\delta)C_t^{\frac{1-\gamma}{\theta}} + \delta(E_t\left[U_{t+1}^{1-\gamma}\right])^{\frac{1}{\theta}}\right]. \tag{29}$$

There are 5 macroeconomic variables, which dynamics follow:

$$\begin{bmatrix} g_{t+1} \\ x_{t+1} \\ \bar{\sigma}^2_{t+1} \\ \sigma^2_{t+1} \\ d_{t+1} \end{bmatrix} = \begin{pmatrix} g_0 \\ x_0 \\ \bar{\sigma}^2_0 \\ \sigma^2_0 \\ d_0 \end{pmatrix} + \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & \rho_x & 0 & 0 & 0 \\ 0 & 0 & \rho_{\bar{\sigma}^2} & 0 & 0 \\ 0 & 0 & (1-\rho_{\sigma^2}) & \rho_{\sigma^2} & 0 \\ 0 & \phi & 0 & 0 & 0 \end{pmatrix} \begin{bmatrix} g_t \\ x_t \\ \bar{\sigma}^2_t \\ \sigma^2_t \\ d_t \end{bmatrix} + G_t z_{t+1} + \tilde{J}_{t+1}, \tag{30}$$

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^2_t represents the conditional volatility and $\bar{\sigma}^2_t$ is the long-run mean component of σ^2_t.

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

$$z_{t+1} \sim \mathcal{N}(\mathbf{0}_{5\times 1}, \mathcal{I}_{5\times 5}), \tag{31}$$

$$G_t G_t' = diag(\varphi \bullet \sqrt{1-\omega})\,\Omega\,diag(\varphi \bullet \sqrt{1-\omega})' + diag(\varphi \bullet \sqrt{\omega})\,\Omega\,diag(\varphi \bullet \sqrt{\omega})'\sigma^2_t,$$

$$\varphi = \begin{pmatrix} \varphi_g \\ \varphi_x \\ \varphi_{\bar{\sigma}^2} \\ \varphi_{\sigma^2} \\ \varphi_d \end{pmatrix}, \omega = \begin{pmatrix} \omega_g \\ \omega_x \\ \omega_{\bar{\sigma}^2} \\ \omega_{\sigma^2} \\ \omega_d \end{pmatrix}, \Omega = \begin{pmatrix} 1 & 0 & 0 & 0 & \Omega_{cd} \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ \Omega_{cd} & 0 & 0 & 0 & 1 \end{pmatrix},$$

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.[21] 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.

The model is calibrated monthly to match a wide set of unconditional macroeconomic and financial moments of quarterly US data 1930-2006. A related model is Drechsler (2013), who adds model uncertainty to the framework of DY. The additional layer of uncertainty allows the model to have smaller and less frequent jumps. Because the model analyzed in the paper features qualitatively exactly the same fundamentals dynamics as DY with jumps to the volatility and expected consumption growth, we do not consider it further.

Wachter (2013)

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

$$U_t = E_t \int_t^\infty \beta(1-\gamma)U_s(\ln C_s - \frac{1}{1-\gamma}\ln((1-\gamma)U_s)ds. \tag{32}$$

Consumption (C_t) and dividends (D_t) follow:

$$\begin{aligned} dC_t &= \mu C_{t^-}dt + \sigma C_{t^-}dB_t + (e^Z - 1)C_{t^-}dN_t, \\ dD_t &= (\phi\mu + \frac{1}{2}\phi(\phi-1)\sigma^2)D_{t^-}dt + \phi\sigma D_{t^-}dB_t + (e^{\phi Z} - 1)D_{t^-}dN_t, \end{aligned} \tag{33}$$

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. Z is a random variable with a time-invariant distribution, which determines the jump (disaster) size. B_t, $B_{\lambda,t}$, N_t, and Z are independent. 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 $s > t$).

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

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.

Bekaert and Engstrom (2017)

The utility function falls into the external habit class:

$$E_t \sum_{j=t}^{\infty} \delta^{j-t} \frac{(C_j - H_j)^{1-\gamma} - 1}{1-\gamma}, \tag{34}$$

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:

$$\begin{aligned}
g_{t+1} &= \bar{g} + \sigma_{cp}\omega_{p,t+1} - \sigma_{cn}\omega_{n,t+1}, \\
d_{t+1} &= \bar{g} + \sigma_{dp}\omega_{p,t+1} - \sigma_{dn}\omega_{n,t+1}, \\
\omega_{p,t+1} &\sim \Gamma(\bar{p}, 1) - \bar{p}, \\
\omega_{n,t+1} &\sim \Gamma(n_t, 1) - n_t, \\
n_{t+1} &= \bar{n} + \rho_n(n_t - \bar{n}) + \sigma_{nn}\omega_{n,t+1},
\end{aligned} \tag{35}$$

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:

$$\begin{aligned}
s_{t+1} &= \bar{s} + \phi(s_t - \bar{s}) + \lambda_t(g_{t+1} - \bar{g}), \\
\lambda_t &= \begin{cases} \frac{1}{s_t}\sqrt{1 - 2(s_t - \bar{s})} - 1, & \text{if } s_t < s_{t,max} \\ 0, & \text{otherwise} \end{cases}, \\
s_{t,max} &= \bar{s} + \frac{1}{2}(1 - \bar{S}_t^2),
\end{aligned} \tag{36}$$

$$\bar{S}_t = \sqrt{(\sigma_{cp}^2 \bar{p} + \sigma_{cn}^2 n_t)\frac{\gamma}{1 - \phi - \frac{b}{\gamma}}}.$$

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.

Model Parameters

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 (jump size)				
	0.0252	0.02	Barro and Ursua (2008) - OECD countries				
D_t	ϕ						
	2.6						
λ_t	$\bar{\lambda}$	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					

General Asset Pricing Fit. Values are unconditional moments. The data is 1990:M1-2017:M12. 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 a 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	Data
$E(rf_{,t})$	0.69%	0.95%	0.47%	1.24%	0.64%
					(0.60%)
$Std(rf_t)$	9.86%	2.12%	2.73%	1.46%	1.92%
					(0.18%)
$E(r_t - rf_t)$	7.79%	6.04%	7.53%	5.99%	4.92%
					(2.69%)
$Std(r_t)$	5.70%	18.10%	21.31%	16.81%	14.64%
					(1.33%)
$E(\frac{P_t}{D_t})$	2.04	19.96	93.77	19.53	51.95
					(4.69)
$Std(\frac{P_t}{D_t})$	0.09	2.96	26.72	5.35	14.30
					(2.61)
χ^2-test	3468***	190***	130***	144***	
χ^2-test: no $\frac{P}{D}$ stats	2645***	12**	43***	20***	

We now discuss models' general asset pricing fit in the table above in more details. Recall that the fit of the BTZ model is poor, because their calibration focused on the unconditional equity premium and risk-free levels, and the zero lower bound on both the consumption growth variance and the variance of consumption growth variance is hit in more than 10% of the simulated time points, leading to significant deviations from their theoretical counterparts. 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, compared to its data counterpart of over 40. The model parameters also 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 **??** 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%. 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 very 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. The table above 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 the table above. Wachter's model generally fits the salient asset return features very 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.

Martin (2017) Bound Fit. Values are computed from monthly data and annualized. Model values are obtained by simulating 100,000 monthly observations for each model. The data 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).

	BTZ	DY	Wachter	BEGE	Data
$\overline{VIX}$	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%)

The Models Fit under Alternative Parameterizations

Sections 3.1-3.3 in the paper show that the extant models do not match the stylized facts under the authors' parameterizations. Here, we argue they cannot do so under any parameterization.

First, given that BTZ is a conditionally Gaussian model, it cannot generate a positive variance risk premium under any parameterization. Second, DY is a conditionally non-Gaussian model that does generate a positive variance risk premium. However, the consumption growth process is conditionally Gaussian preventing the model to match the left-tail consumption growth quantile shifts conditional on the variance risk premium under any parameterization. Finally, the rare disaster and BEGE models cannot be as easily dismissed and we conduct several estimation exercises to demonstrate their poor fit.

To check the Wachter's model's ability to match the moments of interest we perform a formal estimation, using the classical minimum distance (CMD) methodology of Wooldridge (2002). Importantly, the difference between the 10^{th} percentiles of the consumption growth conditional on the high and low variance risk premium (the first row in Table 1) is one of the moments we try to match. However, the volatility risk premium is still too high, the risk-neutral skewness too low, and the consumption growth quantile shift conditional on the variance risk premium is of much smaller magnitude than in the data.

For the BEGE model, an accurate evaluation of the fit in economically interesting parameter regions requires iterating on a high precision multi-dimensional grid, making a formal estimation infeasible. After exploring a combination of calibrations and informal small-scale optimization, we fail to substantially improve the overall asset pricing fit for the BEGE model with alternative parameter choices.

Online Appendix III: Model Solution

The continuously compounded log-risk-free rate, rf_t, is:

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

The price-dividend ratio is (see *"Equity pricing"* subsection of section 4.1 for the derivation algorithm):

$$
\begin{aligned}
&\frac{P_t}{D_t} = \sum_{i=1}^{\infty} e^{A_i + B_i p_t + C_i n_t + D_i q_t},\\
&A_1 = \ln\beta + (1-\gamma)\bar{g} + \gamma(1-\rho_q)\bar{q},\\
&B_1 = -g(m_{\omega,p} + \gamma_g\sigma_{cp}),\\
&C_1 = -g(m_{\omega,n} - \gamma_g\sigma_{cn}),\\
&D_1 = -\gamma(1-\rho_q),\\
&A_n = A_{n-1} + A_1 + B_{n-1}\bar{p}(1-\rho_p) + C_{n-1}\bar{n}(1-\rho_n) + D_{n-1}\bar{q}(1-\rho_q),\\
&B_n = B_{n-1}\rho_p - g(m_{\omega,p} + \gamma_g\sigma_{cp} + B_{n-1}\sigma_{pp} + D_{n-1}\sigma_{qp}),\\
&C_n = C_{n-1}\rho_n - g(m_{\omega,n} - \gamma_g\sigma_{cn} + C_{n-1}\sigma_{nn} + D_{n-1}\sigma_{qn}),\\
&D_n = D_{n-1}\rho_q + D_1.
\end{aligned}
\tag{37}
$$

Formulas in (37) can be derived starting from equation (11) in the paper. The first term can be computed by plugging in (7) and (10):

$$
\begin{aligned}
&E_t e^{m_{t+1}+d_{t+1}} = E_t e^{m_0 + m_q q_t + m_{\omega,p}\omega_{p,t+1} + m_{\omega,n}\omega_{n,t+1} + \bar{g} + \gamma_g(\sigma_{cp}\omega_{p,t+1} - \sigma_{cn}\omega_{n,t+1})} =\\
&E_t e^{m_0 + \bar{g} + m_q q_t + (m_{\omega,p} + \gamma_g\sigma_{cp})\omega_{p,t+1} + (m_{\omega,n} - \gamma_g\sigma_{cn})\omega_{n,t+1}} = e^{m_0 + \bar{g} + m_q q_t - g(m_{\omega,p} + \gamma_g\sigma_{cp}) - g(m_{\omega,n} - \gamma_g\sigma_{cn})},
\end{aligned}
$$

and plugging in m_0, m_q, $m_{\omega,p}$, and $m_{\omega,n}$ from (8) yields coefficients A_1, B_1, C_1, and D_1 in (37).

All other terms in (11) can now be computed recursively using the law of iterated expectations:

$$E_t e^{\sum_{i=1}^{n} m_{t+i}+d_{t+i}} = E_t[e^{m_{t+1}+d_{t+1}} E_{t+1} e^{\sum_{i=2}^{n} m_{t+i}+d_{t+i}}] = E_t e^{m_{t+1}+d_{t+1}+A_{n-1}+B_{n-1}p_{t+1}+C_{n-1}n_{t+1}+D_{n-1}q_{t+1}} =$$

$$E_t e^{m_0+m_q q_t+m_{\omega,p}\omega_{p,t+1}+m_{\omega,n}\omega_{n,t+1}+\bar{g}+\gamma_g(\sigma_{cp}\omega_{p,t+1}-\sigma_{cn}\omega_{n,t+1})+A_{n-1}+B_{n-1}(\bar{p}+\rho_p(p_t-\bar{p})+\sigma_{pp}\omega_{p,t+1})+}$$

$$^{C_{n-1}(\bar{n}+\rho_n(n_t-\bar{n})+\sigma_{nn}\omega_{n,t+1})+D_{n-1}(\bar{q}+\rho_q(q_t-\bar{q})+\sigma_{qp}\omega_{p,t+1}+\sigma_{qn}\omega_{n,t+1})} =$$

$$e^{m_0+\bar{g}+A_{n-1}+B_{n-1}\bar{p}(1-\rho_p)+C_{n-1}\bar{n}(1-\rho_n)+D_{n-1}\bar{q}(1-\rho_q)+(m_q+D_{n-1}\rho_q)q_t+}$$

$$^{(B_{n-1}\rho_p-g(m_{\omega,p}+\gamma_g\sigma_{cp}+B_{n-1}\sigma_{pp}+D_{n-1}\sigma_{qp}))p_t+(C_{n-1}\rho_n-g(m_{\omega,n}-\gamma_g\sigma_{cn}+C_{n-1}\sigma_{nn}+D_{n-1}\sigma_{qn}))n_t},$$

and plugging in m_0, m_q, $m_{\omega,p}$, and $m_{\omega,n}$ from (8) yields recursive expressions for coefficients A_n, B_n, C_n, and D_n in (37).

To compute the first order Taylor approximation of the log-price dividend ratio, we first use the chain rule to compute the partial derivatives of $\ln(\sum_{i=1}^{\infty} e^{A_i+B_i p_t+C_i n_t+D_i q_t})$ with respect to p_t ($\frac{\sum_{i=1}^{\infty} B_i e^{A_i+B_i p_t+C_i n_t+D_i q_t}}{\sum_{i=1}^{\infty} e^{A_i+B_i p_t+C_i n_t+D_i q_t}}$), n_t ($\frac{\sum_{i=1}^{\infty} C_i e^{A_i+B_i p_t+C_i n_t+D_i q_t}}{\sum_{i=1}^{\infty} e^{A_i+B_i p_t+C_i n_t+D_i q_t}}$), and q_t ($\frac{\sum_{i=1}^{\infty} D_i e^{A_i+B_i p_t+C_i n_t+D_i q_t}}{\sum_{i=1}^{\infty} e^{A_i+B_i p_t+C_i n_t+D_i q_t}}$). Thus, the Taylor series log-linearization of (37) around the state variables mean values is:

$$\begin{aligned}
&pd_t \approx \ln[\sum_{i=1}^{\infty} e^{A_i+B_i\bar{p}+C_i\bar{n}+D_i\bar{q}}] + \frac{\sum_{i=1}^{\infty} B_i e^{A_i+B_i\bar{p}+C_i\bar{n}+D_i\bar{q}}}{\sum_{i=1}^{\infty} e^{A_i+B_i\bar{p}+C_i\bar{n}+D_i\bar{q}}}(p_t-\bar{p})+\\
&\frac{\sum_{i=1}^{\infty} C_i e^{A_i+B_i\bar{p}+C_i\bar{n}+D_i\bar{q}}}{\sum_{i=1}^{\infty} e^{A_i+B_i\bar{p}+C_i\bar{n}+D_i\bar{q}}}(n_t-\bar{n}) + \frac{\sum_{i=1}^{\infty} D_i e^{A_i+B_i\bar{p}+C_i\bar{n}+D_i\bar{q}}}{\sum_{i=1}^{\infty} e^{A_i+B_i\bar{p}+C_i\bar{n}+D_i\bar{q}}}(q_t-\bar{q}) =\\
&K_0^1 + K_p^1 p_t + K_n^1 n_t + K_q^1 q_t,\\
&K_p^1 = \frac{\sum_{i=1}^{\infty} B_i e^{A_i+B_i\bar{p}+C_i\bar{n}+D_i\bar{q}}}{\sum_{i=1}^{\infty} e^{A_i+B_i\bar{p}+C_i\bar{n}+D_i\bar{q}}},\\
&K_n^1 = \frac{\sum_{i=1}^{\infty} C_i e^{A_i+B_i\bar{p}+C_i\bar{n}+D_i\bar{q}}}{\sum_{i=1}^{\infty} e^{A_i+B_i\bar{p}+C_i\bar{n}+D_i\bar{q}}},\\
&K_q^1 = \frac{\sum_{i=1}^{\infty} D_i e^{A_i+B_i\bar{p}+C_i\bar{n}+D_i\bar{q}}}{\sum_{i=1}^{\infty} e^{A_i+B_i\bar{p}+C_i\bar{n}+D_i\bar{q}}},\\
&K_0^1 = \ln[\sum_{i=1}^{\infty} e^{A_i+B_i\bar{p}+C_i\bar{n}+D_i\bar{q}}] - K_p^1\bar{p} - K_n^1\bar{n} - K_q^1\bar{q}.
\end{aligned} \tag{38}$$

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

$$
\begin{aligned}
&\ln(1+\frac{P_{t+1}}{D_{t+1}}) \approx K_0^2 + K_p^2 p_{t+1} + K_n^2 n_{t+1} + K_q^2 q_{t+1}, \\
&K_p^2 = \frac{\sum_{i=1}^{\infty} B_i \exp(A_i + B_i\bar{p} + C_i\bar{n} + D_i\bar{q})}{1+\sum_{i=1}^{\infty}\exp(A_i + B_i\bar{p} + C_i\bar{n} + D_i\bar{q})}, \\
&K_n^2 = \frac{\sum_{i=1}^{\infty} C_i \exp(A_i + B_i\bar{p} + C_i\bar{n} + D_i\bar{q})}{1+\sum_{i=1}^{\infty}\exp(A_i + B_i\bar{p} + C_i\bar{n} + D_i\bar{q})}, \\
&K_q^2 = \frac{\sum_{i=1}^{\infty} D_i \exp(A_i + B_i\bar{p} + C_i\bar{n} + D_i\bar{q})}{1+\sum_{i=1}^{\infty}\exp(A_i + B_i\bar{p} + C_i\bar{n} + D_i\bar{q})}, \\
&K_2^0 = \ln[1+\sum_{i=1}^{\infty}\exp(A_i + B_i\bar{p} + C_i\bar{n} + D_i\bar{q})] - K_p^2\bar{p} - K_n^2\bar{n} - K_q^2\bar{q}.
\end{aligned} \tag{39}
$$

The expression for the log-return is:

$$
\begin{aligned}
&r_{t+1} \approx r_0 + r_p p_t + r_n n_t + r_q q_t + r_{\omega,p}\omega_{p,t+1} + r_{\omega,n}\omega_{n,t+1}, \\
&r_0 = \bar{g} + K_0^2 - K_0^1 + K_p^2\bar{p}(1-\rho_p) + K_n^2\bar{n}(1-\rho_n) + K_q^2\bar{q}(1-\rho_q), \\
&r_p = K_p^2\rho_p - K_p^1, \\
&r_n = K_n^2\rho_n - K_n^1, \\
&r_q = K_q^2\rho_q - K_q^1, \\
&r_{\omega,p} = \gamma_g\sigma_{cp} + K_p^2\sigma_{pp} + K_q^2\sigma_{qp}, \\
&r_{\omega,n} = -\gamma_g\sigma_{cn} + K_n^2\sigma_{nn} + K_q^2\sigma_{qn}.
\end{aligned} \tag{40}
$$

SVIX^2 can be computed as:

$$
\begin{aligned}
&SVIX^2 = \sigma^2_{Q,t}(\frac{R_{t+1}}{R_{f,t}}) = \frac{1}{R_{f,t}^2}\sigma^2_{Q,t}(R_{t+1}) = \frac{1}{R_{f,t}^2}(E_t^Q(R_{t+1}^2) - (E_t^Q R_{t+1})^2) = \\
&\frac{1}{R_{f,t}^2}(R_{f,t}E_t(M_{t+1}R_{t+1}^2) - (R_{f,t}E_t(M_{t+1}R_{t+1}))^2) = \frac{1}{R_{f,t}^2}(R_{f,t}E_t(M_{t+1}R_{t+1}^2) - R_{f,t}^2) = \\
&\frac{E_t(M_{t+1}R_{t+1}^2)}{R_{f,t}} - 1 = E_t[exp(m_{t+1} + 2r_{t+1} - r_{f,t})] - 1.
\end{aligned} \tag{41}
$$

By plugging (7), (9) and (16) into (41) we obtain:

$$
\begin{aligned}
&SVIX_t^2 = exp(m_0 + 2r_0 - f_0 + [m_q + 2r_q - f_q]q_t + \\
&[2r_p - g(m_{\omega,p} + 2r_{\omega,p}) - f_p]p_t + [2r_n - g(m_{\omega,n} + 2r_{\omega,n}) - f_n]n_t) - 1.
\end{aligned} \tag{42}
$$

Next we compute VIX^2 as the risk neutral entropy:

$$VIX_t^2 = 2 \cdot [\ln E_t^Q(\frac{R_{t+1}}{R_{f,t}}) - E_t^Q(\ln \frac{R_{t+1}}{R_{f,t}})] = 2 \cdot [\ln E_t^Q(R_{t+1}) - E_t^Q(\ln R_{t+1})] =$$
$$2 \cdot [r_{f,t} - R_{f,t} E_t(M_{t+1} \ln R_{t+1})] = 2 \cdot [r_{f,t} - exp(r_{f,t}) E_t(exp(m_{t+1}) r_{t+1})]. \tag{43}$$

Obtaining a closed-form expression for (43), requires computing the expectation of $X \cdot \exp(X)$, where $X \sim \Gamma(k, \theta)$ with k being the shape and θ the scale parameter, respectively:

$$E[X \cdot \exp(X)] = \frac{1}{\Gamma(k)\theta^k} \int_0^\infty x^{k-1} e^{-\frac{x}{\theta}} x e^x dx = \frac{1}{\Gamma(k)\theta^k} \int_0^\infty x^{(k+1)-1} e^{-\frac{x}{\frac{\theta}{1-\theta}}} dx =$$
$$\frac{1}{\Gamma(k)\theta^k} \cdot \Gamma(k+1) \cdot (\frac{\theta}{1-\theta})^{k+1} \cdot \underbrace{\frac{1}{\Gamma(k+1) \cdot (\frac{\theta}{1-\theta})^{k+1}} \cdot \int_0^\infty x^{(k+1)-1} e^{-\frac{x}{\frac{\theta}{1-\theta}}} dx}_{=1, \text{ as pdf of gamma distribution with shape } k+1 \text{ and scale } \frac{\theta}{1-\theta}} = \tag{44}$$
$$\frac{\Gamma(k+1) \cdot (\frac{\theta}{1-\theta})^{k+1}}{\Gamma(k)\theta^k} = \frac{k\theta}{(1-\theta)^{k+1}},$$

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^2 can be computed by plugging (7), (9), and (16) into (43):

$$VIX_t^2 = 2 \cdot [r_{f,t} - e^{r_{f,t}} \cdot$$
$$E_t\{e^{m_0 + m_q q_t + m_{\omega,p}\omega_{p,t+1} + m_{\omega,n}\omega_{n,t+1}} \cdot (r_0 + r_p p_t + r_n n_t + r_q q_t + r_{\omega,p}\omega_{p,t+1} + r_{\omega,n}\omega_{n,t+1})\}] =$$
$$2 \cdot [r_{f,t} - e^{r_{f,t} + m_0 + m_q q_t} \cdot E_t\{e^{m_{\omega,p}\omega_{p,t+1} + m_{\omega,n}\omega_{n,t+1}} \cdot (r_0 + r_p p_t + r_n n_t + r_q q_t) +$$
$$e^{m_{\omega,p}\omega_{p,t+1} + m_{\omega,n}\omega_{n,t+1}} \cdot (r_{\omega,p}\omega_{p,t+1} + r_{\omega,n}\omega_{n,t+1})\}] = 2 \cdot [r_{f,t} - e^{r_{f,t} + m_0 + m_q q_t} \cdot$$
$$\{e^{-g(m_{\omega,p})p_t - g(m_{\omega,n})n_t} \cdot (r_0 + r_p p_t + r_n n_t + r_q q_t) + E_t[e^{m_{\omega,p}\omega_{p,t+1}} e^{m_{\omega,n}\omega_{n,t+1}} r_{\omega,p}\omega_{p,t+1}] +$$
$$E_t[e^{m_{\omega,p}\omega_{p,t+1}} e^{m_{\omega,n}\omega_{n,t+1}} r_{\omega,n}\omega_{n,t+1}]\}] = 2 \cdot [r_{f,t} - e^{r_{f,t} + m_0 + m_q q_t} \cdot \{e^{-g(m_{\omega,p})p_t - g(m_{\omega,n})n_t} \cdot$$
$$(r_0 + r_p p_t + r_n n_t + r_q q_t) + e^{-g(m_{\omega,n})n_t} r_{\omega,p} E_t[e^{M_{\omega,p}\omega_{p,t+1}}\omega_{p,t+1}] + e^{-g(m_{\omega,p})p_t} r_{\omega,n} E_t[e^{M_{\omega,n}\omega_{n,t+1}}\omega_{n,t+1}]\}] =$$
$$2 \cdot [r_{f,t} - e^{r_{f,t} + m_0 + m_q q_t} \cdot \{e^{-g(m_{\omega,p})p_t - g(m_{\omega,n})n_t} \cdot (r_0 + r_p p_t + r_n n_t + r_q q_t) +$$
$$e^{-g(m_{\omega,n})n_t} r_{\omega,p} \cdot (-p_t e^{-g(m_{\omega,p})p_t} + e^{-m_{\omega,p}p_t} \frac{p_t}{(1-m_{\omega,p})^{p_t+1}}) +$$
$$e^{-g(m_{\omega,p})p_t} r_{\omega,n} \cdot (-n_t e^{-g(m_{\omega,n})n_t} + e^{-m_{\omega,n}n_t} \frac{n_t}{(1-m_{\omega,n})^{n_t+1}})\}], \tag{45}$$

where the last step follows from (44).

Online Appendix IV: Option Pricing

The option pricing in our model is illustrated below using a put option as an example. The pricing of call options is similar. The price of a put option with maturity at time T at time t is:

$$X_t(P_t, K) = E_t[M_{t+T} \cdot (K - P_T)^+], \tag{46}$$

where P is the price of an underlying, K is the strike price, M is the stochastic discount factor, and the + sign indicates that the expression in the parentheses is 0 if it is negative. For simplicity, we operate with current price-adjusted option prices:

$$\frac{X_t(P_t, K)}{P_t} = E_t[M_{t+T} \cdot (\frac{K}{P_t} - \frac{P_T}{P_t})^+]. \tag{47}$$

The expression for M_{t+T} can be obtained as a product of one period ahead stochastic discount factors: $M_{t+T} = \prod_{i=0}^{T-1} M_{t+i \to t+i+1}$. $\frac{P_T}{P_t}$ can be computed as a function of price-dividend ratios and dividend growth: $\frac{\frac{P_T}{D_T}}{\frac{P_t}{D_t}} \cdot \frac{D_T}{D_t}$. Although all components on the right hand-side of (47) are available in closed-form, the expectation has to be evaluated numerically, because there is no formula to assess a positive part of a component model of gamma distributions. To compute 1 month option prices, we evaluate the expectation numerically by sampling 5 million $\omega_{p,t+1}/\omega_{n,t+1}$ pairs. To compute unconditional expectations, we sample a path of 10,000 monthly observations under the estimated model parameters.