



University of Pennsylvania
ScholarlyCommons

Finance Papers

Wharton Faculty Research


2011

What's Vol Got to Do With It

Itamar Drechsler

Amir Yaron
University of Pennsylvania

Follow this and additional works at: http://repository.upenn.edu/fnce_papers

 Part of the [Finance Commons](#), and the [Finance and Financial Management Commons](#)

Recommended Citation

Drechsler, I., & Yaron, A. (2011). What's Vol Got to Do With It. *Review of Financial Studies*, 24 (1), 1-45. <http://dx.doi.org/10.1093/rfs/hhq085>

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/fnce_papers/325
For more information, please contact repository@pobox.upenn.edu.

What's Vol Got to Do With It

Abstract

Uncertainty plays a key role in economics, finance, and decision sciences. Financial markets, in particular derivative markets, provide fertile ground for understanding how perceptions of economic uncertainty and cash-flow risk manifest themselves in asset prices. We demonstrate that the variance premium, defined as the difference between the squared VIX index and expected realized variance, captures attitudes toward uncertainty. We show conditions under which the variance premium displays significant time variation and return predictability. A calibrated, generalized long-run risks model generates a variance premium with time variation and return predictability that is consistent with the data, while simultaneously matching the levels and volatilities of the market return and risk-free rate. Our evidence indicates an important role for transient non-Gaussian shocks to fundamentals that affect agents' views of economic uncertainty and prices.

Disciplines

Finance | Finance and Financial Management

What's Vol Got to Do With It *

Itamar Drechsler[†]

Amir Yaron[‡]

First Draft: July 2007

Current Draft: December 2009

Abstract

Uncertainty plays a key role in economics, finance, and decision sciences. Financial markets, in particular derivative markets, provide fertile ground for understanding how perceptions of economic uncertainty and cashflow risk manifest themselves in asset prices. We demonstrate that the variance premium, defined as the difference between the squared VIX index and expected realized variance, captures attitudes toward uncertainty. We show conditions under which the variance premium displays significant time variation and return predictability. A calibrated, generalized Long-Run Risks model generates a variance premium with time variation and return predictability that is consistent with the data, while simultaneously matching the levels and volatilities of the market return and risk free rate. Our evidence indicates an important role for transient non-Gaussian shocks to fundamentals that affect agents' views of economic uncertainty and prices.

*We thank seminar participants at Imperial College, Michigan, USC, Wharton, Wisconsin, the CREATES workshop 'New Hope for the C-CAPM?', the 2008 Econometric Society Summer Meeting, the 2008 Meeting of the Society for Economic Dynamics, the 2008 NBER Summer Institute's Capital Markets and the Economy Workshop, the Bank of France 'Financial Markets and Real Activity', the NBER Fall Asset Pricing meeting, the AFA, the Utah Winter Finance Conference, WFA, our discussants Luca Benzoni, Imen Ghattassi, Lars Hansen, Jun Pan, George Tauchen, Victor Todorov, and Lu Zhang, an anonymous referee and the editor, Raman Uppal. The authors gratefully acknowledge the financial support of the Rodney White Center at the Wharton School.

[†]NYU Stern, Finance Department, Itamar.Drechsler@stern.nyu.edu.

[‡]The Wharton School, University of Pennsylvania and NBER, yaron@wharton.upenn.edu.

1 Introduction

The idea that volatility has a role in determining asset valuations has long been a cornerstone of finance. Volatility measures, broadly defined, are considered to be useful tools for capturing how perceptions of uncertainty about economic fundamentals are manifested in prices. Derivatives markets, where volatility plays a prominent role, are therefore especially relevant for unraveling the connections between uncertainty, the dynamics of the economy, preferences and prices. This paper focuses on a derivatives-related quantity called the variance premium, which is measured as the difference between (the square of) the CBOE's VIX index and the conditional expectation of realized variance. In this paper, we show theoretically that the variance premium is intimately linked to uncertainty about economic fundamentals and we derive conditions under which it predicts future stock returns.

We document the large and statistically significant predictive power of the variance premium for stock market returns. This finding is consistent with the work in Bollerslev, Tauchen, and Zhou (2009). The variance premium's predictive power is strong at short horizons (measured in months), in contrast to long-horizon predictors, such as the price-dividend ratio, that have been intensively studied in the finance literature. The variance premium is therefore interesting due to both its theoretical underpinnings as well as its empirical success above and beyond that of common return predictors. We analyze whether an extension of the Long Run Risks (LRR) model (as in Bansal and Yaron (2004)), that contains a rich set of transient dynamics, can quantitatively account for the time variation and return predictability of the variance premium while jointly matching 'standard' asset pricing moments, i.e. the level and volatility of the equity premium and risk free rate.

It has been shown that the variance premium equals the difference between the price and expected payoff of a trading strategy.¹ This strategy's payoff is exactly the realized variance of returns. The variance premium is essentially always positive, i.e. the strategy's price is higher than its expected payoff, which suggests it provides a hedge to macroeconomic risks. This mechanism underlies the model in this paper. In the model, market participants are willing to pay an insurance premium for an asset whose payoff is high when return variation is large. This is the case because large return variation is a result of big or important shocks to the economic state. Moreover, when investors perceive that the danger of big shocks to

¹See Demeterfi, Derman, Kamal, and Zou (1999), Britten-Jones and Neuberger (2000), Jiang and Tian (2005) and Carr and Wu (2007).

the state of the economy is high, the hedging premium increases, resulting in a large variance premium.

We model this mechanism in an extension of the Long Run Risks model of Bansal and Yaron (2004). As in their model, agents have a preference for early resolution of uncertainty and therefore dislike increases in economic uncertainty.² In particular, agents fear uncertainty about shocks to influential state variables, such as the persistent component in long-run consumption growth. Under these preferences, economic uncertainty is a priced risk-source that leads to time varying risk premia. We demonstrate that time variation in economic uncertainty and a preference for early resolution of uncertainty are required to generate a positive variance premium that is time-varying and predicts excess stock market returns.³

While our analysis shows that the LRR model captures some qualitative features of the variance premium, we demonstrate that it requires several important extensions in order to quantitatively capture the large size, volatility and high skewness of the variance premium, and importantly, its short-horizon predictive power for stock returns. Our extensions of the baseline LRR model focus on the stochastic volatility process that governs the level of uncertainty about shocks to immediate and long-run components of cashflows. Our specification adds infrequent but potentially large spikes in the level of uncertainty/volatility and infrequent jumps in the small, persistent component of consumption and dividend growth (i.e. we introduce some non-Gaussian shocks). We show that such an extended specification goes a long way towards quantitatively capturing moments of the variance premium and predictability data, while remaining consistent with consumption-dividend dynamics and standard asset pricing moments, such as the equity premium and risk free rate.⁴

²Bansal, Khatchatrian, and Yaron (2005) provide empirical evidence supporting the presence of conditional volatility in cashflows across several countries. Lettau, Ludvigson, and Wachter (2007) analyze whether the great moderation, the decline in aggregate volatility of macro aggregates can reconcile the run-up in valuation ratios during the late 90s. Bloom (2007) provides direct evidence linking spikes in market return uncertainty and subsequent declines in economic activity.

³Tauchen (2005) generalizes the volatility uncertainty in Bansal and Yaron (2004) to one in which the variance of volatility shocks is stochastic. Eraker (2007) adds jumps to the volatility specification. The focus on the variance premium is different from these papers.

⁴The inclusion of jump shocks is demanded by our desire to quantitatively jointly match the rich set of cashflow and asset price data moments we target. An early version of this paper considered a model without jumps but with large volatility in volatility, and for pedagogical reasons this model now appears in Appendix B. Bollerslev, Tauchen, and Zhou (2009) also utilize such a model to illustrate that variation in uncertainty can deliver return predictability by the variance premium. Evidence from reduced-form studies and our work with such a model strongly suggest that jump shocks have an important role in addressing the myriad data moments that we are interested in (see also Section 5).

There is a long-standing literature on option pricing, which typically formulates models with a reduced-form pricing kernel or directly within a risk-neutral framework. Our inclusion of non-Gaussian dynamics builds on some of the findings of this literature (e.g., Broadie, Chernov, and Johannes (2007), Chernov and Ghysels (2000), Eraker (2004), Pan (2002)). However, by construction, such models have limited scope for explicitly mapping macroeconomic fundamentals and preferences into risk prices. A contribution of this paper is to explicitly and quantitatively link information priced into a key derivatives index with a model of preferences and macroeconomic conditions. Understanding these connections is clearly an important challenge for macroeconomics and finance.⁵ Some recent papers linking prices of derivatives with recursive preferences and/or long-run risks fundamentals include Benzoni, Collin-Dufresne, and Goldstein (2005), Liu, Pan, and Wang (2005), Tauchen (2005), Bansal, Gallant, and Tauchen (2007), Bhamra, Kuehn, and Strebulaev (2007), Chen (2008), and Eraker and Shaliastovich (2008).

The paper continues as follows: Section 2 presents the data, defines the variance premium, discusses its statistical properties, and then proceeds to evaluate its role in predicting future returns. Section 3 presents a generalized LRR framework with jumps in volatility and cashflow growth, and discusses return premia. Section 4 derives the variance premium inside the model and provides the link between the variance premium and return predictability within the model. Section 5 provides results from calibrating several specifications of these models. Section 6 provides concluding remarks.

2 Definitions and Data

Our definitions of key terms are similar to those in Bollerslev, Gibson, and Zhou (2009) and Bollerslev, Tauchen, and Zhou (2009) and closely follow the related literature. We formally define the variance premium as the difference between the risk neutral and physical expectations of the market's total return variation. We will focus on a one month variance premium, so the expectations are of total return variation between the current time, t , and one month forward, $t + 1$. Thus, $vp_{t,t+1}$, the (one-month) variance premium at time

⁵It is by no means a foregone conclusion that a model that is able to capture the equity premium will also be consistent with the options data. The options data seem to require non-Gaussian features, and there is a substantial quantitative challenge in jointly matching their properties while remaining consistent with the cashflows and equity premium.

t , is defined as $E_t^Q[\text{Total Return Variation}(t, t + 1)] - E_t[\text{Total Return Variation}(t, t + 1)]$, where Q denotes the risk-neutral measure. Demeterfi, Derman, Kamal, and Zou (1999) and Britten-Jones and Neuberger (2000) show that, in the case that the underlying asset price is continuous, the risk neutral expectation of total return variance can be computed by calculating the value of a portfolio of European calls on the asset. Jiang and Tian (2005) and Carr and Wu (2007) show this result extends to the case where the asset is a general jump-diffusion. This approach is model-free since the calculations do not depend on any particular model of options prices. The VIX Index is calculated by the Chicago Board Options Exchange (CBOE) using this model-free approach to obtain the risk-neutral expectation of total variation over the subsequent 30 days. Therefore we obtain closing values of the VIX from the CBOE and use it as our measure of risk-neutral expected variance. Since the VIX index is reported in annualized “vol” terms, we square it to put it in “variance” space and divide by 12 to get a monthly quantity. Below we refer to the resulting series as squared VIX.

As the definition of $vp_{t,t+1}$ indicates, we also need conditional forecasts of total return variation under the true data generating process or physical measure. To obtain these forecasts we create measures of the total realized variation of the market, or realized variance, for the months in our sample. Our measure is created by summing the squared five-minute log returns over a whole month. For comparison, we do this for both the S&P 500 futures and S&P 500 cash index. We obtain the high frequency data used in the construction of our realized variance measures from TICKDATA. As discussed below, we project the realized variance measures on a set of predictor variables and construct forecasted series for realized variance.⁶ These forecast series are our proxy for the conditional expectation of total return variance under the physical measure. The difference between the risk neutral expectation, measured using the VIX, and the conditional forecasts from our projections, gives the series of one-month variance premium estimates.

Our data series for the VIX and realized variance measures covers the period January 1990 to March 2007. The main limitation on the length of our sample comes from the VIX, which is only published by the CBOE beginning in January of 1990. We obtain daily and monthly

⁶We treat returns overnight or over a weekend the same as one “five-minute” interval. Treating these longer periods as one interval does not bias the magnitude of the realized variance measure. For comparison, Table I below shows that the realized variance measure based on daily returns has a similar mean. Of course, the advantage of using a finer sampling frequency, where possible, is that provides more precise measures of variance.

returns on the value-weighted NYSE-AMEX-NASDAQ market index and the S&P 500 from CRSP. The monthly P/E ratio series for the S&P 500 is obtained from Global Financial Data. Our model calibrations will also require data on consumption and dividends. We use the longest sample available (1930:2006). Per-capita consumption of non-durables and services is taken from NIPA. The per-share dividend series for the stock market is constructed from CRSP by aggregating dividends paid by common shares on the NYSE, AMEX, and NASDAQ. Dividends are adjusted to account for repurchases as in Bansal, Dittmar, and Lundblad (2005).

Table I provides summary statistics for the monthly log excess returns on both the S&P 500 and the total value-weighted market return. The excess returns are constructed by subtracting the log 30-day T-Bill return, available from CRSP. The two series display very similar statistics. Both series have an approximately 0.53% mean monthly excess return with a volatility of about 4%. The other statistics are also quite close. Thus, although the availability of high-frequency data for the S&P 500 leads us to use it in our empirical analysis, our empirical inferences and theoretical model apply to the broader market.

The last four columns in Table I provide statistics for several measures of realized variance — potential inputs for our forecasts of realized variance: the squared VIX, the futures realized variance, cash index realized variance, and also the sum of squared daily returns over the month. The squared VIX value for a particular month is simply the value of the last observation for that month. The futures, cash, and daily realized variances are sums over the whole month. We will ultimately use the futures realized variance and we display the other two for comparison. Several issues are worth noting. First, all volatility measures display significant deviation from normality. The mean to median ratio is large, the skewness is positive and greater than 0, and the kurtosis is clearly much larger than 3. Bollerslev, Tauchen, and Zhou (2009) use the sum based on the cash index returns as their realized variance measure. This realized variance has a smaller mean than the futures and daily measures. This smaller mean is a result of a non-trivial autocorrelation in the five-minute returns on the cash index and is not present in the returns on the futures. We suspect that this autocorrelation is the effect of ‘stale’ prices at the five-minute intervals, since computation of the S&P 500 cash index involves 500 separate prices (see Campbell, Lo, and MacKinlay (1997) and references therein for a discussion of stale prices and return autocorrelation). As the S&P 500 futures involves only one price, and has long been one of the most liquid financial instruments available, we choose to use its realized variance measure

to proxy for the total return variation of the market.

Table II provides a comparison of conditional variance projections. Our approach is to find a parsimonious representation, yet one that delivers significant predictability. The last two regressions show our choice of projection for the S&P index and futures variance measures. For these dependent variables we find that a parsimonious projection on the lagged VIX and index realized variance achieves R^2 s of close to 60%. The addition of further lags or predictor variables adds very little predictive power. The first regression in the table provides the conditional volatility based on daily squared returns. We fit a GARCH(1,1) to provide a comparison with approaches used in early studies of variation, which used daily data. This regression achieves an R^2 of around 40%. It is the use of high-frequency returns and the VIX as predictor that accomplishes the increased predictive power of the first two regressions.

Table III provides summary statistics for various measures of the variance premium, constructed as differences of the squared VIX and various variance forecasts. For comparison, the first column also reports the measure that is the main focus of Bollerslev, Tauchen, and Zhou (2009). They calculate this measure of the variance premium by subtracting from the squared VIX the previous month's realized variance.⁷ It is apparent from the table that the mean of the variance premium is somewhat larger when based on the cash index measures as opposed to the futures or daily variance measures. Furthermore, the variance premium based on the futures measure is significantly less volatile than the other measures. Neither effects are surprising given the results in Table II and the discussion above regarding the cash index realized variance. The remaining statistics, in particular the skewness and kurtosis, seem to be quite similar across the variance premium proxies. In what follows, we use the variance premium based on the futures realized variance. As discussed above, the liquidity of the futures contract makes it an appropriate instrument for measuring realized variance. It is also the defacto instrument used by traders involved in related options trading. It is important to note however that our subsequent results are not materially effected by the use of this particular measure.

Table IV provides return predictability regressions. There are two sets of columns with

⁷Bollerslev, Tauchen, and Zhou (2009) conduct a robustness exercise where they also construct a measure of the variance premium using variance forecasts and show that their return predictability results are qualitatively unchanged.

regression estimates. The first set of columns shows OLS estimates and the second set provides estimates from robust regressions. Robust regression performs estimation using an iterative reweighted least squares algorithm that downweights the influence of outliers on estimates but is nearly as statistically efficient as OLS in the absence of outliers. It provides a check that the results are not driven by outliers.⁸ The first two regressions are one-month ahead forecasts using the variance premium as a univariate regressor, while the third forecasts one quarter ahead. The quarterly return series is overlapping. The last two specifications add the price-earnings ratio, which is a commonly used variable for predicting returns. As a univariate regressor, the variance premium can account for about 1.5-4.0% of the monthly return variation. The multivariate regressions lead to a substantial further increase in the R^2 – a feature highlighted in Bollerslev, Tauchen, and Zhou (2009). For example, in conjunction with the price-earnings ratio, the in-sample R^2 increases to as much as 13.4%.⁹ It is worth noting that the lagged variance premium seems to perform better than the immediate variance premium. Note that in both cases, as well as the multivariate specification, the variance premium enters with a significant positive coefficient. We will show that this sign and magnitude are consistent with theory. Finally, we note that the robust regression estimates agree both in magnitude and sign with the OLS estimates and in fact, some of the R-squares are even larger than their OLS counterparts.¹⁰

A natural question that arises is whether such R^2 s are economically significant. Cochrane (1999) uses a theorem of Hansen and Jagannathan (1991) to derive a relationship between the maximum unconditional Sharpe ratio attainable using a predictive regression and the regression R^2 . It says that $(s^*)^2 - s_0^2 = \frac{1+s_0^2}{1-R^2}R^2$, where s_0 is the unconditional buy-and-hold Sharpe Ratio and s^* is the maximum unconditional Sharpe ratio.¹¹ In our sample, s_0

⁸The robust regression R^2 s are pseudo R^2 s and they are calculated as the ratio of the variance of the regression forecast to the variance of the dependent variable, which corresponds to the usual R^2 calculation in the case of OLS.

⁹The in-sample R^2 of the price-earnings ratio alone is about 3.4%. The bivariate R^2 s are significantly higher than the sum of R^2 s from the univariate regressions. This is because of a positive correlation between the two regressors.

¹⁰Another robustness check we have done is to create the series of realized variance forecasts (used in the construction of the variance premium) using rolling projections estimated on only past data, instead of with the whole sample (as above). We use the first 24 months to initialize the rolling regression estimates. The results (not reported) are very similar to and actually slightly stronger than those reported in Table IV.

¹¹This formula corresponds to the case when the predictive regression's residual is homoskedastic. If the predictive regressor also forecasts increased residual variance, the improvement in unconditional Sharpe ratio will be less. This is clearly the case here since the predictors are closely related to volatility forecasts. Hence, we are *not* using the formula to draw any conclusions about attainable Sharpe ratios, but only to show that the R^2 sizes are economically meaningful.

is approximately 0.157 at a monthly frequency, or 0.543 annualized. Using the univariate regression with an R^2 of 4.07%, the maximal Sharpe ratio would rise to 0.904 annualized. With the bivariate R^2 of 8.30%, the maximal Sharpe Ratio would further increase to 1.19, more than double the unconditional ratio. In other words, the potential increases are quite large. It is important to keep in mind that these R^2 s are for a *monthly* horizon, and that Sharpe ratios increase roughly with the square root of the horizon. Hence an R^2 of 3% at the monthly horizon is potentially very useful. A comparison with “traditional” predictive variables found in the literature also shows this predictability is large. For example, Campbell, Lo, and MacKinlay (1997) examine the standard price-dividend ratio and stochastically detrended short-term interest rate, two of the more successful predictive variables, and show that in the more predictable second subsample, the predictive R^2 s are 1.5% and 1.9% respectively at the monthly horizon. Campbell and Thompson (2007) examine a large collection of predictive variables whose in-sample (monthly) R^2 s are much smaller than those reported in Table IV, but still conclude that these variables can be useful to investors. Finally, note that the variance related variables, i.e. the VIX^2 , realized variance measures, and variance premium, all have AR(1) coefficients of 0.79 or less, unlike the price-dividend ratio or short term interest rate, which have AR(1) coefficients much closer to 1. This means the variance related quantities will not suffer from the large predictive regression biases associated with extremely persistent predictive variables, such as the price-dividend ratio (e.g. Stambaugh (1999)), and will have much better finite sample properties.

3 Model Framework

The underlying environment is a discrete time endowment economy. The representative agent’s preferences on the consumption stream are of the Epstein and Zin (1989) form, allowing for the separation of risk aversion and the intertemporal elasticity of substitution (IES). Thus, the agent maximizes his life-time utility, which is defined recursively as

$$V_t = \left[(1 - \delta)C_t^{\frac{1-\gamma}{\theta}} + \delta \left(E_t[V_{t+1}^{1-\gamma}] \right)^{\frac{1}{\theta}} \right]^{\frac{\theta}{1-\gamma}} \quad (1)$$

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

(IES). Utility maximization is subject to the budget constraint,

$$W_{t+1} = (W_t - C_t)R_{c,t+1} , \quad (2)$$

where W_t is the wealth of the agent, and $R_{c,t}$ is the return on all invested wealth. As shown in Epstein and Zin (1989), for any asset j , the first order condition yields the following Euler condition,

$$E_t [\exp (m_{t+1} + r_{j,t+1})] = 1 \quad (3)$$

where $r_{j,t+1}$ is the log of the gross return on asset j , and m_{t+1} is the log of the intertemporal marginal rate of substitution, which is given by $\theta \ln \delta - \frac{\theta}{\psi} \Delta c_{t+1} + (\theta - 1)r_{c,t+1}$. Here $r_{c,t+1}$ is $\ln R_{c,t+1}$ and Δc_{t+1} is the change in $\ln C_t$.

3.1 Dynamics

For notational brevity and expositional ease, we specify the dynamics of the state vector in the model in a rather general framework. However, we then immediately provide the specific version of the dynamics that is our focus. Benzoni, Collin-Dufresne, and Goldstein (2005) is the first paper to model jumps within a long-run risks setup. The general framework in this paper most closely follows Eraker and Shaliastovich (2008), though in discrete time. The state vector of the economy is given by $Y_t \in \mathbb{R}^n$ and follows a VAR that is driven by both Gaussian and Poisson jump shocks:

$$Y_{t+1} = \mu + FY_t + G_t z_{t+1} + J_{t+1} \quad (4)$$

Here $z_{t+1} \sim \mathcal{N}(0, \mathcal{I})$ is the vector of Gaussian shocks and J_{t+1} is the vector of jump shocks. We let the jumps be compound-Poisson jumps. Therefore, the i -th component of J_{t+1} is given by $J_{t+1,i} = \sum_{j=1}^{N_{t+1}^i} \xi_i^j$, 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 . Thus, $J_{t+1,i}$ represents the total jump in $Y_{t+1,i}$ between time t and $t + 1$. We let the N_{t+1}^i be independent of each other conditional on time- t information and assume that the ξ_i^j are i.i.d.. The intensity process for N_{t+1}^i is given by the i -th component of the vector λ_t . In other words, λ_t is the vector of intensities for the Poisson counting processes.

To put the dynamics into the affine class (Duffie, Pan, and Singleton (2000)), we impose

an affine structure on G_t and λ_t :

$$\begin{aligned} G_t G_t' &= h + \sum_k H_k Y_{t,k} \\ \lambda_t &= l_0 + l_1 Y_t \end{aligned}$$

where $h \in \mathbb{R}^{n \times n}$, $H_k \in \mathbb{R}^{n \times n}$, $l_0 \in \mathbb{R}^n$, and $l_1 \in \mathbb{R}^{n \times n}$.

To handle the jumps we introduce some notation. Let $\psi_k(u_k) = E[\exp(u_k \xi_k)]$, i.e. ψ_k is the moment generating function (mgf) of the jump size ξ_k . The mgf for the k -th jump component, $E_t[\exp(u_k J_{t+1,k})]$, then equals $\exp(\Psi_{t,k}(u_k))$, where $\Psi_{t,k}(u_k) = \lambda_{t,k}(\psi_k(u_k) - 1)$. $\Psi_{t,k}$ is called the cumulant generating function (cgf) of $J_{t+1,k}$ and is a very helpful tool for calculating asset pricing moments. The reason is that its n -th derivative evaluated at 0 equals the n -th central moment of $J_{t+1,k}$. It is convenient to stack the mgf's into a vector function. Thus, for $u \in \mathbb{R}^n$ let $\psi(u)$ be the vector with k -th component $\psi_k(u_k)$ and let $\Psi_t(u)$ be defined analogously. It will also be necessary to evaluate the scalar quantity $E_t[\exp(u' J_{t+1})]$, $u \in \mathbb{R}^n$. Since the $J_{t+1,k}$ are (conditionally) independent of each other, this equals $\exp(\sum_k \lambda_{t,k}(\psi_k(u_k) - 1))$, or more compactly, $\exp(\lambda_t'(\psi(u) - 1))$.

3.2 Long Run Risks Model with Jumps

In the calibration section of the paper and also in some of the discussion that follows, we focus on a particular specification of (4). This specification is a generalized LRR model that incorporates jumps. Here we give an overview of this generalized LRR model and map it into the general framework in (4). Further details are also provided in the calibration section.

We specify:

$$Y_{t+1} = \begin{pmatrix} \Delta c_{t+1} \\ x_{t+1} \\ \bar{\sigma}_{t+1}^2 \\ \sigma_{t+1}^2 \\ \Delta d_{t+1} \end{pmatrix} \quad F = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & \rho_x & 0 & 0 & 0 \\ 0 & 0 & \rho_{\bar{\sigma}} & 0 & 0 \\ 0 & 0 & (1 - \tilde{\rho}_{\sigma}) & \rho_{\sigma} & 0 \\ 0 & \phi & 0 & 0 & 0 \end{pmatrix}$$

and the vector of Gaussian shocks is $z_{t+1} = (z_{c,t+1}, z_{x,t+1}, z_{\bar{\sigma},t+1}, z_{\sigma,t+1}, z_{d,t+1}) \sim \mathcal{N}(0, \mathcal{I})$ and $J_{t+1} = (0, J_{x,t+1}, 0, J_{\sigma,t+1}, 0)$ is the jump vector.

The first element of the state vector, Δc_{t+1} , is the growth rate of log consumption. As

in the long-run risks model, $\mu_c + x_t$ is the conditional expectation of consumption growth, where x_t is a small but persistent component that captures long run risks in consumption and dividend growth. The parameter ρ_x is the persistence of x_t . In the dividend growth specification, ϕ is the loading of Δd_{t+1} on the long-run component and will be greater than 1 in the calibrations, so that dividend growth is more sensitive to x_t than is consumption growth.

The dynamics of volatility are driven by two factors, σ_t^2 and $\bar{\sigma}_t^2$. We let σ_t^2 control the conditional volatility and let $\bar{\sigma}_t^2$ drive variation in the long run mean of σ_t^2 (such a volatility structure is also utilized in, for example, Duffie, Pan, and Singleton (2000)). Hence, we set the conditional variance-covariance matrix of the Gaussian shocks to be $G_t G_t' = h + H_\sigma \sigma_t^2$. In addition, we focus attention on a jump intensity specification of the form $\lambda_t = l_0 + l_{1,\sigma} \sigma_t^2$. Thus, σ_t^2 also drives variation in the intensities of the jumps.¹² Since σ_t^2 is positive valued, positivity of the jump intensities is implied. The fact that $\bar{\sigma}_t^2$ controls the long run mean of σ_t^2 comes from the term $(1 - \tilde{\rho}_\sigma)$, the loading of σ_t^2 on $\bar{\sigma}_t^2$ in the matrix F . As will become clear in a later section, when there are no jumps in σ_t^2 , then $\tilde{\rho}_\sigma$ is simply equal to ρ_σ . When there are jumps, $\tilde{\rho}_\sigma - \rho_\sigma$ equals the compensation term for the conditional mean of the σ_t^2 jump shock (see section 5.1) and ensures that the unconditional mean of σ_t^2 remains the same when we include jump shocks.

The generalized LRR specification above is quite flexible and nests a number of related models. In particular, it nests the original Bansal and Yaron (2004) long-run risks model. To obtain the original long run risks model as a specific case, set $l_0 = l_1 = 0$, so there are no jumps, and parameterize the Gaussian variance-covariance matrix via $h = \text{diag}([0, 0, 0, \varphi_\sigma, 0])$ and $H_\sigma = \text{diag}([\varphi_c, \varphi_x, 0, 0, \varphi_d])$. In the Bansal and Yaron (2004) specification, the volatility of σ_t^2 shocks is constant, and the long run mean of volatility, $\bar{\sigma}_t^2$, also remains constant. Tauchen (2005) makes the volatility of σ_t^2 shocks stochastic via a square-root specification. To get this type of specification, set $H_\sigma = \text{diag}([\varphi_c, \varphi_x, 0, \varphi_\sigma, \varphi_d])$ and $h = 0$. Finally, as the specification above shows, we consider jumps in both σ_t^2 and x_t , but not in the immediate innovations to Δc_{t+1} , Δd_{t+1} , and $\bar{\sigma}_t^2$. As will be discussed below, the non-Gaussian (jump) shocks to these two state variables are important for establishing both the qualitative properties of the variance premium and for the quantitative model calibrations.

¹²Here $l_{1,\sigma}$ is the column multiplying σ_t^2 in the expression $l_1 Y_t$, which means it is just the fourth column of l_1 .

3.3 Model Solution

We now solve for the equilibrium price process of the model economy. The solution proceeds via the representative agent's Euler condition (3). To price assets we must first solve for the return on the wealth claim, $r_{c,t+1}$, as it appears in the pricing kernel itself. Denote the log of the wealth-to-consumption ratio at time t by v_t . Since the wealth claim pays the consumption stream as its dividend, this is simply the price-dividend ratio of the wealth claim. Next, we use the Campbell and Shiller (1988) log-linearization to linearize $r_{c,t+1}$ around the unconditional mean of v_t :

$$r_{c,t+1} = \kappa_0 + \kappa_1 v_{t+1} - v_t + \Delta d_{t+1} \quad (5)$$

This approach is also taken by Bansal and Yaron (2004), Eraker and Shaliastovich (2008), and Bansal, Kiku, and Yaron (2007b). We then conjecture that the no-bubbles solution for the log wealth-consumption ratio is affine in the state vector:

$$v_t = A_0 + A'Y_t$$

where $A = (A_c, A_x, A_{\bar{\sigma}}, A_{\sigma}, A_d)'$ is a vector of pricing coefficients. Substituting v_t into (5) and then substituting (5) into the Euler equation gives the equation in terms of A , A_0 and the state variables. The expectation on the left side of this Euler equation can be evaluated analytically, as shown in Appendix A.1. It is also shown there that the requirement that the Euler equation hold for any realization of Y_t implies that A_0 and A jointly satisfy a system of $n + 1$ equations which determine their values.

3.3.1 Pricing Kernel

Having solved for $r_{c,t+1}$, we can substitute it into m_{t+1} to obtain an expression for the log pricing kernel at time $t + 1$:

$$\begin{aligned} m_{t+1} &= \theta \ln \delta - \frac{\theta}{\psi} \Delta c_{t+1} + (\theta - 1)r_{c,t+1} \\ &= \theta \ln \delta + (\theta - 1)\kappa_0 + (\theta - 1)(\kappa_1 - 1)A_0 - (\theta - 1)A'Y_t - \Lambda'Y_{t+1} \end{aligned} \quad (6)$$

where $\Lambda = (\gamma e_c + (1 - \theta)\kappa_1 A)$ and e_c is $(1, 0, 0, 0, 0)'$ (the selector vector for Δc). The innovation to the pricing kernel, conditional on the time t information set, has the simple form:

$$m_{t+1} - E_t(m_{t+1}) = -\Lambda'(Y_{t+1} - E_t(Y_{t+1})) = -\Lambda'(G_t z_{t+1} + J_{t+1} - E_t(J_{t+1})) \quad (7)$$

Thus, Λ can be interpreted as the price of risk for Gaussian shocks and also the sensitivity of the IMRS to the jump shocks. From the expression for Λ one can see that the prices of risk are determined by the A coefficients. Since any predictive information in Δc_t and Δd_t is subsumed in x_t , they have no effect on v_t and therefore $A_c = A_d = 0$. Thus, $\Lambda = (\gamma, \kappa_1 A_x(1 - \theta), \kappa_1 A_{\bar{\sigma}}(1 - \theta), \kappa_1 A_{\sigma}(1 - \theta), 0)'$.

The expression for Λ shows that the signs of the risk prices depend on the signs of the A coefficients and $(1 - \theta)$. When $\gamma = \frac{1}{\psi}$ and $\theta = 1$ we are in the case of CRRA preferences, it is clear that only the transient shock to consumption $z_{c,t+1}$ is priced, and prices do not separately reflect the risk of shocks to x_t (“long-run risk”) or σ_t^2 (uncertainty/volatility related risk).

In the discussion below and in the calibrations we focus on the case where the agent’s risk aversion is greater than 1 and $\psi > 1$, which implies that $\Lambda_x > 0$ and $\Lambda_{\sigma} < 0$. Thus, positive shocks to long-run growth *decrease* the IMRS, while positive shocks to the level of uncertainty/volatility *increase* the IMRS. Note that in this case, since $(1 - \theta) > 0$, each of the A coefficients has the same sign as the corresponding price of risk. $A_x > 0$, so increases in long-run growth imply an increase in v_t , while $A_{\sigma} < 0$, so increases in uncertainty/volatility decrease v_t . Thus, an agent that has $\gamma > 1$ and $\psi > 1$ dislikes increases in the level of uncertainty/volatility (since the IMRS increases) and associates them with decreases in prices (the wealth-consumption ratio). This joint behavior of the IMRS and prices is important for our theoretical and quantitative results regarding the variance premium. We note that since $\gamma > \frac{1}{\theta}$, this parametrization of preferences is identified by Epstein and Zin (1989) as implying a preference for early resolution of uncertainty. It is also important to note that this configuration *endogenously* generates the ‘leverage effect’, the well documented negative correlation between innovations to returns and to volatility (see also Bansal and Yaron (2004) and Tauchen (2005)).

For comparison, when $\frac{1}{\psi} > \gamma > 1$ (preference for late resolution of uncertainty), $A_x < 0$ and $A_{\sigma} > 0$, and hence a positive shock to x_t (σ_t^2) *lowers* (*raises*) v_t . Moreover, $(1 - \theta) > 0$, so

the exactly the opposite is true for the IMRS. This type of configuration leads to qualitatively counterfactual results, such as a negative variance premium. When $\psi < 1$ and $\gamma > \frac{1}{\psi}$ (preference for early resolution of uncertainty), $A_x < 0$ and $A_\sigma > 0$. This configuration would cause the model to contradict the well known “leverage effect”, the empirical result that changes in prices and the level of volatility appear to be inversely related. Such a contradiction has further undesirable implications for quantitatively matching the variance premium and the shape of the option-implied volatility surface.

3.3.2 The Market Return

To study the variance premium, equity risk premium, and their relationship, we first need to solve for the market return. A share in the market is modeled as a claim to a dividend with growth process given by Δd_{t+1} . To solve for the price of a market share we proceed along the same lines as for the consumption claim and solve for $v_{m,t+1}$, the log price-dividend ratio of the market, by using the Euler equation (3). To do this, log-linearize the return on the market, $r_{m,t+1}$, around the unconditional mean of $v_{m,t+1}$:

$$r_{m,t+1} = \kappa_0 + \kappa_1 v_{m,t+1} - v_{m,t} + \Delta d_{t+1} \quad (8)$$

Then conjecture that $v_{m,t}$ is affine in the state variables:

$$v_{m,t} = A_{0,m} + A'_m Y_t$$

where $A_m = (A_{c,m}, A_{x,m}, A_{\bar{\sigma},m}, A_{\sigma,m}, A_{d,m})'$ is the vector of pricing coefficients for the market. Substituting the log-linearized return and conjecture for $v_{m,t}$ into the Euler equation and evaluating the left side leads also to a system of $n + 1$ equations, analogous to that described in section 3.3, which must hold for all values of Y_t . The equations for A_m are in terms of the solution of A , and since the A 's determine the nature of the pricing kernel, the A_m 's largely inherit their properties from the corresponding A 's. In particular, since our reference specification implies $A_c = A_d = 0$, it is also the case that $A_{c,m} = A_{d,m} = 0$. The solution method for A carries over almost directly for A_m . The derivation of A_m and further solution details are provided in Appendix A.3.

By substituting the expression for $v_{m,t}$ into the linearized return, we obtain an expression

for $r_{m,t+1}$ in terms of Y_t and its innovations:

$$r_{m,t+1} = r_0 + (B_r'F - A_m')Y_t + B_r'G_t z_{t+1} + B_r'J_{t+1} \quad (9)$$

where r_0 is a constant, $B_r = (\kappa_{1,m}A_m + e_d)$, and e_d is $(0, 0, 0, 0, 1)'$ (the selector vector for Δd).

Since, conditional on time t information, the components of z_{t+1} and J_{t+1} are all independent of each other, the conditional variance of the return is simply:

$$\text{var}_t(r_{m,t+1}) = B_r'G_tG_t'B_r + \sum_i B_r^2(i)\text{var}_t(J_{t+1,i})$$

where B_r^2 denotes elementwise squaring of B_r and $B_r^2(i)$ is its i -th element. Recall that the n -th central moment of $J_{t+1,i}$ is given by the n -th derivative of its cgf at 0, i.e. $\Psi_{t,i}^{(n)}(0)$. For the case of compound Poisson jumps, it was noted above that $\Psi_{t,i}(u) = \lambda_{t,i}(\psi_i(u) - 1)$, so the conditional variance can be rewritten concisely as:

$$\begin{aligned} \text{var}_t(r_{m,t+1}) &= B_r'G_tG_t'B_r + B_r^{2'}\Psi_t^{(2)}(0) \\ &= B_r'G_tG_t'B_r + B_r^{2'}\text{diag}(\psi^{(2)}(0))\lambda_t \end{aligned} \quad (10)$$

where $\text{diag}(\psi^{(2)}(0))$ denotes the matrix with $\psi^{(2)}(0)$ on the diagonal.

3.3.3 Risk Premia

Appendix A.4 derives the following expression for the conditional equity premium, which highlights the contribution of the compound Poisson shocks:

$$\ln E_t(R_{m,t+1}) - r_{f,t} = B_r'G_tG_t'\Lambda + \lambda_t'(\psi(B_r) - 1) - \lambda_t'(\psi(B_r - \Lambda) - \psi(-\Lambda)) \quad (11)$$

The first term, $B_r'G_tG_t'\Lambda$, represents the contributions of the Gaussian shocks to the risk premium. This is the standard and familiar expression for the equity premium in the absence of jump shocks — in this case, the left hand side simply equals $E_t(r_{m,t+1} - r_{f,t}) + 0.5\text{var}_t(r_{m,t+1})$. This term emanates from the covariance of the Gaussian shock in the pricing kernel (7) and the return equation (9). The next terms, $\lambda_t'(\psi(B_r) - 1) - \lambda_t'(\psi(B_r - \Lambda) - \psi(-\Lambda))$, represent the contributions from the jump processes. The derivation in Appendix A.4 indicates

that this term reflects the covariance of the jump component in the pricing kernel with the jump component in the return. This separation into Gaussian and jump contributions is due to the conditional independence of these two types of shocks. Note the presence of $\psi(\cdot)$, which encodes the jump distribution and is analogous to the presence of the covariance matrix in the Gaussian term. Furthermore, it is important to notice that the variation in the jump contribution is driven by the intensity of the jump shocks, λ_t . Under our reference parametrization, where $\gamma > 1$ and $\psi > 1$, the jump risk premia is positive – that is, the loadings on λ_t add up to a positive contribution to the risk premium. Thus, when λ_t increases, the market risk premium rises. Below, we discuss how the jump contribution to the risk premium can be interpreted in terms of the difference between risk neutral and physical measure quantities.

4 The Variance Premium and Return Predictability

In this section we derive the variance premium and show that it effectively reveals the level of the (latent) jump intensity. When $\gamma > 1$ and $\psi > 1$, as in our reference parametrization, an increase in jump intensity causes an increase in both the variance premium and the market risk premium. As a result, the variance premium is able to capture time variation in the risk premium and is an effective predictor of market returns.

As defined in section 2 above, the one period variance premium at time t , $vp_{t,t+1}$, is the difference between the representative agent’s risk neutral and physical expectations of the market’s total return variation between time t and $t + 1$. In continuous-time models, total return variation is expressed as an integral of instantaneous return variation over infinitely many periods from t to $t + 1$. In a discrete-time model, where t to $t + 1$ represents one time period, strictly speaking the variance premium simply equals $\text{var}_t^Q(r_{m,t+1}) - \text{var}_t(r_{m,t+1})$. Here $\text{var}_t^Q(r_{m,t+1})$ denotes the conditional variance of market returns under the risk-neutral measure Q (we let P denote the physical measure, and where not explicitly specified, the measure is taken to be the physical measure). If we consider dividing t to $t + 1$ into n sub-periods, the variance premium would be defined as the following sum:

$$E_t^Q \left[\sum_{i=1}^{n-1} \text{var}_{t+\frac{i-1}{n}}^Q \left(r_{m,t+\frac{i-1}{n}, t+\frac{i}{n}} \right) \right] - E_t^P \left[\sum_{i=1}^{n-1} \text{var}_{t+\frac{i-1}{n}}^P \left(r_{m,t+\frac{i-1}{n}, t+\frac{i}{n}} \right) \right] \quad (12)$$

where $\text{var}_{t+\frac{i-1}{n}}(r_{m,t+\frac{i-1}{n},t+\frac{i}{n}})$ is notation for the time $t+\frac{i-1}{n}$ conditional variance of the market return between $t+\frac{i-1}{n}$ and $t+\frac{i}{n}$.

The variance premium is non-zero because of two effects discussed below. The first is that $\text{var}_t^Q(r_{m,t+1}) \neq \text{var}_t^P(r_{m,t+1})$. In other words, the levels of the conditional variances at time t are different under the physical and risk neutral measures. We refer to this difference in levels by the name:

$$\text{level difference} \equiv \text{var}_t^Q(r_{m,t+1}) - \text{var}_t^P(r_{m,t+1}) \quad (13)$$

The second effect is that the expected change, or drift, in the quantity $\text{var}_t(r_{m,t+1})$ is different under Q and P . In other words, $E_t^Q[\text{var}_{t+1}^Q(r_{m,t+2})] - \text{var}_t^Q(r_{m,t+1}) \neq E_t^P[\text{var}_{t+1}^P(r_{m,t+2})] - \text{var}_t^P(r_{m,t+1})$. This is a result of the fact that Y_t has different dynamics under Q and P . We refer to this difference in drifts as the:

$$\text{drift difference} \equiv \{E_t^Q[\text{var}_{t+1}^Q(r_{m,t+2})] - \text{var}_t^Q(r_{m,t+1})\} - \{E_t^P[\text{var}_{t+1}^P(r_{m,t+2})] - \text{var}_t^P(r_{m,t+1})\} \quad (14)$$

Equation (12) is effectively a sum of the level difference and differences in the drifts of conditional variance over the sub-periods. To capture both effects in our model, we define our $vp_{t,t+1}$ as the level difference plus the drift difference over the period t to $t+1$. Adding them together results in our definition of the variance premium:

$$vp_{t,t+1} \equiv E_t^Q[\text{var}_{t+1}^Q(r_{m,t+2})] - E_t^P[\text{var}_{t+1}^P(r_{m,t+2})] \quad (15)$$

Since the variance premium involves expectations under Q of functions of the state vector, to derive $vp_{t,t+1}$ we must solve for the model dynamics under the risk neutral measure.

4.1 Model Dynamics under the Risk Neutral Measure

Recall from (4) the state dynamics under the physical measure:

$$Y_{t+1} = \mu + FY_t + G_t z_{t+1} + J_{t+1}$$

The distributions of stochastic elements of the dynamics, z_{t+1} and J_{t+1} , are transformed by the change of probability measure. To change to the risk-neutral measure, we re-weight

probabilities according to the value of the pricing kernel. In other words we set the Radon-Nikodym derivative $\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{M_{t+1}}{E_t(M_{t+1})}$. From (7) we have $\frac{M_{t+1}}{E_t(M_{t+1})} \propto \exp(-\Lambda'(G_t z_{t+1} + J_{t+1}))$. Since z_{t+1} and J_{t+1} are independent, we can treat their measure transformations separately. The case of z_{t+1} is simple. Let $f_t(z_{t+1})$ denote the joint (time t conditional) density of z_{t+1} under P and let $f_t^Q(z_{t+1})$ be its Q counterpart. Then $f_t(z_{t+1}) \propto \exp(-\frac{1}{2}z'_{t+1}z_{t+1})$ and re-weighting it with the relevant part of the Radon-Nikodym derivative implies:

$$\begin{aligned} f_t^Q(z_{t+1}) &\propto \exp(-\frac{1}{2}z'_{t+1}z_{t+1}) \exp(-\Lambda'G_t z_{t+1}) \\ &\propto \exp(-\frac{1}{2}(z_{t+1} + G'_t\Lambda)'(z_{t+1} + G'_t\Lambda)) \end{aligned}$$

where the last line follows from a ‘‘complete-the-square’’ argument. This shows that

$$z_{t+1} \stackrel{Q}{\sim} \mathcal{N}(-G'_t\Lambda, I) \quad (16)$$

i.e. under Q , z_{t+1} is still a vector of independent normals with unit variances, but with a shift in the mean.

For the case of J_{t+1} we could also proceed by transforming the probability density function directly. A somewhat more general and easier way to proceed is by obtaining the cgf of J_{t+1} under Q . Proposition (9.6) in Cont and Tankov (2004) shows that under Q , the $J_{t+1,k}$ are still compound Poisson processes, but with cgf given by:

$$\Psi_{t,k}^Q(u_k) = \lambda_{t,k} \psi_k(-\Lambda_k) \left(\frac{\psi_k(u_k - \Lambda_k)}{\psi_k(-\Lambda_k)} - 1 \right) \quad (17)$$

A short discussion will help to interpret this result and see how it arises. First, under Q , the distribution of the jump size ξ_k is re-weighted by the probability density $\frac{\exp(-\Lambda_k \xi_k)}{E(\exp(-\Lambda_k \xi_k))}$. Thus, the mgf of ξ_k under Q is $E\left(\exp(u_k \xi_k) \frac{\exp(-\Lambda_k \xi_k)}{E(\exp(-\Lambda_k \xi_k))}\right) = \frac{\psi_k(u_k - \Lambda_k)}{\psi_k(-\Lambda_k)}$, which is in (17). There is some intuition behind this re-weighting. It ‘tilts’ the distribution of the jump size ξ_k in a direction depending only on the associated price of risk Λ_k . If $\Lambda_k < 0$, then $\exp(-\Lambda_k \xi_k)$ is larger for greater values of ξ_k . Hence, the distribution is transformed so that under Q more positive jumps have higher probability. Moreover, the extent of the tilting depends on the magnitude of the risk price. A larger risk price produces a greater transformation, while a zero risk price implies no alteration in the jump distribution under Q . One way to assess

this transformation is to compute the mean jump size under Q :

$$E^Q(\xi_k) = E^P\left(\xi_k \frac{\exp(-\Lambda_k \xi_k)}{E^P(\exp(-\Lambda_k \xi_k))}\right) = E^P(\xi_k) + \text{cov}\left(\xi_k, \frac{\exp(-\Lambda_k \xi_k)}{E^P(\exp(-\Lambda_k \xi_k))}\right)$$

This calculation shows that the covariation of the jump size with the tilting weight determines the difference in mean jump size between P and Q . The same computation on $E^Q(\xi_k^2)$ would indicate how the variance of the jump size changes under Q . The second implication of (17) is that, under Q , the jump intensity is $\lambda_{t,k}\psi_k(-\Lambda_k)$. The transformation of the jump intensity follows the same principle as for the jump distribution. The sign of the price of risk is important in determining whether the jump intensity is amplified or diminished, while the magnitude of the risk price controls the degree of the change.

Given (17), we can now easily compute the moments of J_{t+1} under Q by taking derivatives of the Q measure cgf:

$$E_t^Q(J_{t+1,k}) = \Psi_{t,k}^Q(1)(0) = \lambda_{t,k}\psi_k^{(1)}(-\Lambda_k) \quad (18)$$

$$\text{var}_t^Q(J_{t+1,k}) = \Psi_{t,k}^Q(2)(0) = \lambda_{t,k}\psi_k^{(2)}(-\Lambda_k) \quad (19)$$

Finally, we use these results to rewrite the state dynamics under Q . Let $\tilde{z}_{t+1} = z_{t+1} + G_t' \Lambda$. Then $\tilde{z}_{t+1} \stackrel{Q}{\sim} \mathcal{N}(0, \mathcal{I})$ and the state dynamics under Q can be rewritten as:

$$Y_{t+1} = \mu + FY_t - G_t G_t' \Lambda + G_t \tilde{z}_{t+1} + J_{t+1}^Q \quad (20)$$

where J_{t+1}^Q denotes the vector of independent compound Poisson processes with cgf given under Q by (17).

We can also use the above discussion to analyze the contribution of the jump terms to the equity premium in (11). To do so, it is helpful to rewrite their sum as $\lambda_t'(\psi(B_r) - 1) - (\lambda_t \cdot \psi(-\Lambda))'(\frac{\psi(B_r - \Lambda)}{\psi(-\Lambda)} - 1)$, where the division and multiplication by $\psi(-\Lambda)$ in the second term is componentwise. Comparing the second term to (17) shows that it equals $\Psi_t^Q(B_r)$, the cgf vector under Q evaluated at B_r . The first term is also a cgf vector evaluated at B_r , but under P . Thus, the jump contribution to the equity premium can be interpreted as the difference between the expectation of return jumps under P and Q , which is captured by the cgfs.

We now show why the jump contribution is positive. Recall that the intensity of jumps

is scaled under Q (componentwise) relative to P by the factor $\psi(-\Lambda)$. Our preference configuration leads to $\psi(-\Lambda) > 1$, an amplification in jump intensity for both σ_t^2 and x_t jumps. This is because $\Lambda_\sigma < 0$ and σ_t^2 jumps are positive, while $\Lambda_x < 0$ and x_t jumps are negatively skewed. Thus, to show that the total jump contribution is positive, it is sufficient to show that $\frac{\psi(B_r - \Lambda)}{\psi(-\Lambda)} < \psi(B_r)$ (componentwise). Since $\frac{\psi(B_r - \Lambda)}{\psi(-\Lambda)}$ is just $E^Q(\exp(B_r \xi))$, this amounts to showing that under Q the expectation of return jumps is less (i.e. more negative) than under P . To see this, recall that $E^Q(\exp(B_r \xi)) = E\left(\exp(B_r \xi) \frac{\exp(-\Lambda \xi)}{E(\exp(-\Lambda \xi))}\right)$ and rewrite it as $E(\exp(B_r \xi)) \cdot 1 + \text{cov}\left(\exp(B_r \xi), \frac{\exp(-\Lambda \xi)}{E(\exp(-\Lambda \xi))}\right)$. As $E(\exp(B_r \xi))$ is just $\psi(B_r)$, we just need to show that $\text{cov}\left(\exp(B_r \xi), \frac{\exp(-\Lambda \xi)}{E(\exp(-\Lambda \xi))}\right) < 0$. This is certainly the case for our configurations, and is also typically true, since shocks that carry a negative price of risk (so are disliked by the representative agent) usually also have a negative return loading (i.e. if $\Lambda_k < 0$ then $B_r(k) < 0$) and vice versa.

4.2 The Variance Premium and the Risk of Jumps

In this section we derive the level difference component of the variance premium and then analyze its relation to return predictability. Subsequently, we return to a discussion of the drift difference. We choose to focus first on the level difference as this allows us to highlight important qualitative points while simplifying the algebraic expressions and exposition.

It follows from (9), (16), and (19) that:

$$\begin{aligned} \text{var}_t^Q(r_{m,t+1}) &= B_r' G_t G_t' B_r + B_r^{2'} \Psi_t^{Q(2)}(0) \\ &= B_r' G_t G_t' B_r + B_r^{2'} (\text{diag}(\psi^{(2)}(-\Lambda)) \lambda_t \end{aligned} \quad (21)$$

Subtracting $\text{var}_t^P(r_{m,t+1})$ (equation (10)) from $\text{var}_t^Q(r_{m,t+1})$ then gives the level difference:

$$\text{var}_t^Q(r_{m,t+1}) - \text{var}_t^P(r_{m,t+1}) = B_r^{2'} \text{diag}(\psi^{(2)}(-\Lambda) - \psi^{(2)}(0)) \lambda_t \quad (22)$$

Some observations are now possible. First, note that the part of conditional variance coming from the Gaussian shocks, $B_r' G_t G_t' B_r$ cancels out in the level difference. The reason for this is that z_{t+1} has the same variance under P and Q . Thus, Gaussian-induced variance makes no contribution to the level difference since it is the same under the physical and

risk-neutral probabilities.¹³

Second, expression (22) shows that the level difference is simply proportional to the latent jump intensity and, so long as $\Lambda \neq 0$, can be used to reveal it. For example, suppose for simplicity that there are Poisson jumps in only one state variable, say x_t . If $\Lambda_x \neq 0$, i.e. x_t shocks are priced, then $\left(\psi_x^{(2)}(-\Lambda_x) - \psi_x^{(2)}(0)\right) \neq 0$. In this case, the level difference is just a multiple of the jump intensity λ_t and perfectly reveals its value. Since the variance premium includes the level difference, and tends in fact to be dominated by it, its value will also strongly reflect the latent jump intensity.¹⁴

Now consider how the level difference depends on the prices of risk and therefore indirectly on preferences. First, as discussed earlier, in the case of CRRA preferences ($\gamma = 1/\psi$) only the immediate shock to consumption is priced and $\Lambda_x = \Lambda_\sigma = 0$. Thus, equation (22) then clearly shows that the level difference is 0.

Next, consider the jump in σ_t^2 in our reference configuration. To determine the sign of the corresponding contribution to the level difference, we need to sign the term $\psi_\sigma^{(2)}(-\Lambda_\sigma) - \psi_\sigma^{(2)}(0)$, and based on the mgfs this term equals $E_t(\xi_\sigma^2 [\exp(-\Lambda_\sigma \xi_\sigma) - 1])$. In the model calibrations, ξ_σ has a gamma distribution, which means all jump sizes are positive. It is therefore the case that $[\exp(-\Lambda_\sigma \xi_\sigma) - 1]$ is either always positive or always negative depending on the sign of Λ_σ . As discussed above, for $\gamma > 1, \psi > 1$, we get $\Lambda_\sigma < 0$, and so the term's contribution to the level difference is positive. This is a direct outcome of the representative agent's aversion to increases in uncertainty/volatility. As discussed earlier, for this preference configuration the representative agent dislikes increases in uncertainty, and so his risk-neutral measure puts greater weight on states where there is a large, positive shock to σ_t^2 . Thus, large shocks are more probable under Q , which implies a higher variance, so that the level difference is positive. By comparison, if $1 < \gamma < \frac{1}{\psi}$, then $\Lambda_\sigma > 0$ and the representative agent downweights the probability of large shocks. The resulting level difference is then *negative*

¹³This conclusion is the discrete-time analog to what is typically the case in continuous-time diffusion models of option pricing, though it is perhaps less obvious under the continuous-time formulations. For example, in the well-known Heston (1993) model, the variance premium for the “dt” interval $[t, t+dt)$ is actually 0. It is non-zero for any finite interval $[t, t+\delta t)$ because of what we are calling here the drift difference between Q and P . Later we show that in our calibration the level difference dominates quantitatively the drift difference.

¹⁴The level difference can also reveal the jump intensity when there are Poisson jumps in multiple state variables, but the λ_t vector is driven by a single state variable (for example λ_t may be a state variable itself). Then, if $\Lambda \neq 0$, the level difference is simply a multiple of the jump state variable and therefore makes it observable. This is the case for our reference calibration configuration in which λ_t is driven by σ_t^2 .

and also leads, counterfactually, to $vp_{t,t+1} < 0$. Though the reasoning here is for a gamma jump specification, it applies much more generally.

Consider also the contribution of the jumps in x_t to the level difference. In the calibrations we consider two distributions for x_t jumps, a symmetric and an asymmetric one. The symmetric distribution is just a mean-zero normal distribution. Let $\xi_x \sim \mathcal{N}(0, \sigma_x^2)$. Then an easy calculation gives:

$$\psi_x^{(2)}(-\Lambda_x) - \psi_x^{(2)}(0) = \exp\left(\frac{1}{2}\Lambda_x^2\sigma_x^2\right)\Lambda_x^2\sigma_x^4 + \exp\left(\frac{1}{2}\Lambda_x^2\sigma_x^2\right)\sigma_x^2 - \sigma_x^2 \quad (23)$$

which is clearly positive so long as $\Lambda_x \neq 0$, regardless of its sign. This happens because the pricing kernel is convex in shocks to x_t (or in fact any priced state variable), so that it increases more quickly with the size of a ‘bad’ shock than it decreases with the size of a ‘good’ shock. As a result, under Q the agent places a higher probability, on average, on states with large magnitude shocks. This implies that variance is higher under the risk neutral measure and that the level difference is positive.

The above discussion refers to the case for a symmetric distribution. Now consider negatively skewed shocks to x_t , i.e. negative jumps in x_t are larger (but relatively rare) while positive jumps are smaller (but more frequent). As discussed above, for $\gamma > 1$, $\psi > 1$, we get $\Lambda_x > 0$, and the pricing factor $\exp(-\Lambda_x\xi_x)$ will tilt the risk neutral probabilities towards the negative shocks. Since negative shocks are predominantly also large shocks, this will increase risk neutral variance even more than in the symmetric case (holding constant the price of risk) and lead to an even more positive level difference.

4.2.1 Return Predictability

Why should the variance premium have predictive power for future returns? Formally, it is now easy to see why the level difference, and therefore the variance premium, should predict returns. Substituting in $G_t G'_t = h + H_\sigma \sigma_t^2$ in equation (11), one can see that variation in the market risk premia is driven by the processes σ_t^2 and λ_t . For our configuration, time variation in λ_t is driven only by σ_t^2 , and hence for notational simplicity we denote the loading of the market risk premia on σ_t^2 by $\beta_{r,\sigma}$. As discussed in section 3.3.3, when $\gamma > 1$, $\psi > 1$,

$\beta_{r,\sigma}$ is positive.¹⁵

According to the level difference equation (22), and since $\lambda_t = l_{1,\sigma}\sigma_t^2$, we can rewrite $\text{var}_t^Q(r_{m,t+1}) - \text{var}_t(r_{m,t+1})$ as $\beta_{\text{lev},\sigma}\sigma_t^2$. As discussed above, $\gamma > 1$, $\psi > 1$ implies that the level difference is positive, so $\beta_{\text{lev},\sigma} > 0$.

Now, consider the predictive regression for excess market returns:

$$r_{m,t+1} - r_{f,t} = \alpha + \beta_{\text{pred}} \left(\text{var}_t^Q(r_{m,t+1}) - \text{var}_t(r_{m,t+1}) \right) + \epsilon_{t+1}$$

Substituting in the expressions gives

$$\begin{aligned} \beta_{\text{pred}} &= \frac{\text{cov} \left(E_t(r_{m,t+1} - r_{f,t}) + \epsilon_{t+1}, \text{var}_t^Q(r_{m,t+1}) - \text{var}_t(r_{m,t+1}) \right)}{\text{var} \left(\text{var}_t^Q(r_{m,t+1}) - \text{var}_t(r_{m,t+1}) \right)} \\ &= \frac{\text{cov}(\beta_{r,\sigma}\sigma_t^2, \beta_{\text{lev},\sigma}\sigma_t^2)}{\beta_{\text{lev},\sigma}^2 \text{var}(\sigma_t^2)} = \frac{\beta_{r,\sigma}}{\beta_{\text{lev},\sigma}} > 0 \end{aligned}$$

Therefore, $\text{var}_t^Q(r_{m,t+1}) - \text{var}_t(r_{m,t+1})$ predicts excess returns on the market. The predictive coefficient is positive, as in the data. The intuition is as follows. As it controls the intensity of jumps through λ_t , the state variable σ_t^2 is important in determining expected excess returns. When jump intensity is high, there is a relatively high possibility of a large negative shock to x_t (the long run growth component) or a large positive shock to σ_t^2 (the level of uncertainty/volatility). An agent whose preferences are characterized by $\gamma > 1$, $\psi > 1$ is averse to both such shocks. Therefore, the agent considers times of high jump intensity as very risky, and they are therefore characterized by a high conditional equity risk premium. Second, as discussed earlier, the agent's aversion to the large shocks makes the risk-neutral conditional variance higher than the physical one. This difference in the variances rises with the jump intensity, leading to the positive covariation between the variance premium and conditional equity premium.

¹⁵An interesting extension of our reference configuration would create a wedge between λ_t and σ_t^2 . For example, a minor extension of the model could add an additional innovation to our specification of λ_t , i.e. $\lambda_t = l_0 + l_{1,\sigma}\sigma_t^2 + \varphi_\lambda z_{\lambda,t}$. This would reduce the perfect correlation between σ_t^2 , λ_t and the resulting variance premium. In general, the inclusion of an additional state variable to the model to drive λ_t is potentially desirable. Though such a state variable would not materially change the underlying mechanisms at work in the model, it will increase the complexity of the model and its calibration. We believe the reference configuration strikes a good balance between parsimony and achievement of the main objectives of the model.

4.3 Drift Difference

We now examine the contribution to $vp_{t,t+1}$ from the drift difference, which is the difference in the “drift” of the conditional variance between the two measures (see equation (14)). It is simplest to look at this in the case of purely Gaussian shocks, but the principle carries through when there are also Poisson shocks. If all shocks are Gaussian then, as mentioned earlier, we have $\text{var}_t^Q(r_{m,t+1}) = \text{var}_t^P(r_{m,t+1})$. Hence, the drift difference is simply $E_t^Q[\text{var}_{t+1}(r_{m,t+2})] - E_t^P[\text{var}_{t+1}(r_{m,t+2})]$. From (10) we have that, in the pure Gaussian case, $\text{var}_t(r_{m,t+1}) = B_r' G_t G_t' B_r$. In our reference configuration, $G_t G_t' = h + H_\sigma \sigma_t^2$, so that $\text{var}_t(r_{m,t+1}) = B_r' h B_r + B_r' H_\sigma B_r \sigma_t^2$ and the drift difference is just $B_r' H_\sigma B_r [E_t^Q(\sigma_{t+1}^2) - E_t^P(\sigma_{t+1}^2)]$, i.e. this quantity arises from the different drift of σ_t^2 between Q and P . Moreover, since $B_r' H_\sigma B_r \geq 0$ (H_σ is positive semi-definite), the drift difference is just a positive multiple of $E_t^Q(\sigma_{t+1}^2) - E_t^P(\sigma_{t+1}^2)$.

Recall that the dynamics of the state vector are different under Q and P . We are now interested specifically in the dynamics of σ_t^2 under the two measures. From (20) we see that for the reference configuration the pure Gaussian case gives:

$$E^Q(Y_{t+1}) - E^P(Y_{t+1}) = -G_t G_t' \Lambda = -(h + H_\sigma \sigma_t^2) \Lambda$$

Let σ_t^2 correspond to row i of Y_t (in our reference model $i = 4$). Assume for simplicity, that shocks to σ_t^2 are uncorrelated with the other shocks. In this case, within the i -th row of $h + H_\sigma \sigma_t^2$ only the i -th element is non-zero, and the drift difference is simply:

$$\text{drift difference} = -B_r' H_\sigma B_r [h(i, i) + H_\sigma(i, i) \sigma_t^2] \Lambda_\sigma$$

A few observations are worth making about this expression. First, the sign of the drift difference depends on the sign of Λ_σ . When $\Lambda_\sigma < 0$, so the agent is averse to increases in σ_t^2 , then the drift difference is *positive*. As discussed earlier, this is the case for $\gamma > 1$, $\psi > 1$. However, for $1 < \gamma < \frac{1}{\psi}$, ($\gamma > 1$ and preference for late resolution of uncertainty), the opposite is the case and the drift difference is *negative*. Lastly, in the CRRA case, $\Lambda_\sigma = 0$ and the drift difference is 0.

A second important observation is that the size of the wedge in expectations increases with the expected magnitude of shocks to σ_t^2 , i.e. with the conditional volatility of the shocks. Thus, time variation in the size of the drift difference is determined by whatever

variables drive variation in the conditional volatility of shocks to σ_t^2 . In the reference model this is σ_t^2 itself (so long as $H_\sigma(i, i) \neq 0$) and therefore the drift difference reveals the value of σ_t^2 . However, this idea is true more broadly. If, for example, a separate state variable drives the magnitude of σ_t^2 shocks, then it will determine variation in the drift difference. Appendix B gives a simple (pure Gaussian) example of such a model, where a new variable, denoted q_t , determines the volatility of σ_t^2 shocks (leading to a variation on a CIR process).

Finally, consider an economy where $H_\sigma(i, i) = 0$, i.e. the volatility of σ_t^2 shocks is constant. This is the case in the Bansal and Yaron (2004) model. In this case, the drift difference is constant. Moreover, in Bansal and Yaron (2004) all shocks are Gaussian, so the level difference is zero. The sum of these two parts, which is the total variance premium $vp_{t,t+1}$, is the constant drift difference. Since the variance premium is constant, it cannot have predictive power for returns in that model.

4.3.1 Predictability

Since the drift difference is directly related to the expected size of shocks to σ_t^2 , it will have predictive power for returns under Epstein-Zin preferences. In our reference model, the drift difference reflects the value of σ_t^2 . As σ_t^2 also drives time variation in risk premia, a projection of excess returns on the drift difference captures this time variation. Moreover, when $\gamma > 1$, $\psi > 1$, the projection coefficient is positive as both the drift difference and risk premium increase with σ_t^2 .

This predictability by the drift difference holds more generally than in just the reference model. For example, in the model of Appendix B, the state variable q_t controls the expected magnitude of shocks to σ_t^2 . Hence, q_t is a distinct, priced risk factor. A projection of excess returns on the drift difference captures the component of the risk premium attributable to q_t . For $\gamma > 1$, $\psi > 1$, the drift difference and projection coefficient are both positive. Thus, a similar mechanism again implies that the drift difference is related to a (latent) variable that is associated with the level of uncertainty, imparting it with predictive power for returns.

4.3.2 Adding Up the Parts

The final part to discuss is the drift difference when the effect of the Poisson jumps on it is included. We relegate this discussion to Appendix C and note that it largely parallels

the discussion above about the drift difference in the Gaussian case. Finally, the Appendix discusses the total variance premium resulting from adding up the level and drift differences. To learn more about the properties of the model and investigate whether the model is able to capture quantitative properties of the data, we now turn to the model calibration.

5 Calibration & Results

In this section we calibrate the model with the goal of matching a broad set of cashflow and asset pricing moments that, among others, include salient features of the variance premium. We first describe our particular parametrization of the model and then proceed to discuss our calibration criteria and the empirical results.

5.1 Parametrization

We specify a gamma distribution for the sizes of the jumps in σ_t^2 : $\xi_\sigma \sim \Gamma(\nu_\sigma, \frac{\mu_\sigma}{\nu_\sigma})$. This parametrization of the gamma jump follows Eraker and Shaliastovich (2008). It is convenient since it implies that $E[\xi_\sigma] = \mu_\sigma$. The parameter ν_σ is called the shape parameter of the gamma distribution (the other parameter is the ‘scale’ parameter). As ν_σ decreases, the right tail of the distribution becomes thicker and the distribution becomes more asymmetric. When $\nu_\sigma = 1$, the gamma distribution reduces to an exponential distribution.

For the jumps in x_t (the long run component in cash flows), we consider one symmetric and one asymmetric jump distribution. Our benchmark specification involves an asymmetric demeaned gamma distribution: $\Gamma(\nu_x, \frac{\mu_x}{\nu_x}) - \mu_x$. Demeaning the jump size prevents λ_t , the jump intensity, from entering into the expected change in x_t . Otherwise, it would become a factor in the expected change in x_t , since it drives the expected number of jumps during the following period. We choose to make x_t jumps *negatively* skewed, i.e. larger shocks tend to be negative (but relatively infrequent), whereas smaller shocks tend to be positive (and relatively more common). Therefore we take the negative of the demeaned gamma distribution, i.e. $\xi_x \sim -\Gamma(\nu_x, \frac{\mu_x}{\nu_x}) + \mu_x$. For completeness, we will also present results for a symmetric distribution that is a zero-mean normal distribution: $\xi_x \sim \mathcal{N}(0, \sigma_x^2)$.

It is easiest to choose parameters for the calibration after the state dynamics (4) are

rewritten in terms of compensated (i.e. conditionally demeaned) jump shocks, so that these shocks become true ‘innovations’. Let $\tilde{J}_{t+1} = J_{t+1} - E_t(J_{t+1})$ denote the compensated jump processes. Then the model dynamics can be rewritten in this ‘innovations’ form by using $E_t(J_{t+1}) = \text{diag}(\psi^{(1)}(0)) \lambda_t$, and the identity $\lambda_t = l_0 + l_1 Y_t$, to obtain:

$$Y_{t+1} = \tilde{\mu} + \tilde{F}Y_t + G_t z_{t+1} + \tilde{J}_{t+1} \quad (24)$$

where $\tilde{F} = (F + \text{diag}[\psi^{(1)}(0)] l_1)$ and $\tilde{\mu} = \mu + \text{diag}(\psi^{(1)}(0)) l_0$. Since the diagonal of \tilde{F} (rather than F) represents the true autoregressive parameters for the state variables, it is clearer from (24) whether the parameters imply stationary dynamics. Therefore, we directly parametrize \tilde{F} in (24) rather than F . For our specifications, the difference between them is in the equation for σ_t^2 . This is the result of the jumps in σ_t^2 , which have a non-zero mean. Since σ_t^2 itself drives the intensity of the jumps, \tilde{F} implies that the true autoregressive parameter for σ_t^2 is larger than the parameter ρ_σ in (4). We label the true autoregressive parameter $\tilde{\rho}_\sigma$ and write:

$$\tilde{F} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & \rho_x & 0 & 0 & 0 \\ 0 & 0 & \rho_\sigma & 0 & 0 \\ 0 & 0 & (1 - \tilde{\rho}_\sigma) & \tilde{\rho}_\sigma & 0 \\ 0 & \phi & 0 & 0 & 0 \end{pmatrix}$$

Furthermore, rather than parameterizing the VAR constant term $\tilde{\mu}$ directly, we specify the unconditional mean, $E(Y_t)$, since this is more intuitive. The mapping between the two is simply $(I - \tilde{F})E(Y_t) = \tilde{\mu}$, where I is the identity matrix. Without loss of generality, we adopt the following normalization, $E[\sigma_t^2] = E[\tilde{\sigma}_t^2] = 1$. This normalization makes many parameters easier to interpret. For example, the unconditional mean of the jump intensity is then just $l_0 + l_{1,\sigma}$. By a property of the Poisson process, this then equals the average number of jumps in a single period.

Finally, we parameterize the variance-covariance matrix of the Gaussian shocks by specifying h and H_σ . The specification is motivated by two requirements: (i) allow the conditional volatility of the state variable shocks to have potentially different sensitivities to time variation in σ_t^2 (ii) allow for correlations between the shocks.

To gain intuition about our ultimate specification, we first discuss requirement (i) in the absence of any cross-shock correlations. In this case, requirement (i) can be achieved by

specifying that for shock i : $h(i, i) + H_\sigma(i, i)\sigma_t^2 = \varphi_i^2(1 - w_i)E(\sigma_t^2) + \varphi_i^2 w_i \sigma_t^2$ and by setting the off-diagonal elements of H to zero. Variable i 's conditional shock variance is then a weighted average of its unconditional mean and a time-varying part driven by σ_t^2 . The parameter w_i is the weighting that controls the conditional shock variance's sensitivity to changes in σ_t^2 . Note that the mean of the conditional shock variance is simply $\varphi_i^2 E(\sigma_t^2) = \varphi_i^2$. Now consider the second requirement, allowing for correlations between any of the shocks. Let Ω be a correlation matrix (with diagonal elements equal to 1, i.e. $\Omega_{ii} = 1$ for all i) and let φ be the vector of φ_i and w be the vector of w_i . Then we set:

$$h + H\sigma_t^2 = \text{diag}(\varphi\sqrt{1-w}) \Omega \text{diag}(\varphi\sqrt{1-w})' + \text{diag}(\varphi\sqrt{w}) \Omega \text{diag}(\varphi\sqrt{w})' \sigma_t^2$$

On the diagonal this is the same as $h(i, i) + H_\sigma(i, i)\sigma_t^2 = \varphi_i^2(1 - w_i)E(\sigma_t^2) + \varphi_i^2 w_i \sigma_t^2$. For off-diagonal terms, it implies that the unconditional correlation of shocks i and j is approximately Ω_{ij} , with the approximation becoming exact when $w_i = w_j$. The conditional correlation is also approximately Ω_{ij} , with the approximation becoming precise as σ_t^2 moves to extreme values.¹⁶ We highlight that, although the specification above is quite general, for parsimony, in the calibrations below we only introduce correlation between the immediate shocks to dividends and consumption and leave the shocks to x_t and σ_t^2 orthogonal to all the others. Furthermore, we only adjust w_i to fractional values for the consumption and dividend processes. For x_t , we keep the shock structure the same as in Bansal and Yaron (2004) by letting $w_x = 1$. We make σ_t^2 a square-root (CIR) process, by setting $w_\sigma = 1$. In parallel again with Bansal and Yaron (2004), we make $\bar{\sigma}_t^2$ homoscedastic, by setting $w_{\bar{\sigma}} = 0$.

5.2 Results

In what follows we confront the model with a broad set of cashflow and asset pricing targets. Specifically, we calibrate the model with the following objectives: we would like to find a specification for the long run, volatility, and jump shocks such that (i) the model's consumption and dividend growth statistics are consistent with salient features of the consumption and dividends data (ii) the model generates consistent unconditional moments of asset prices, such as the equity premium and the risk free rate (iii) the model generates a

¹⁶To be precise, the unconditional correlation is $\Omega_{ij} \left(\sqrt{(1-w_i)(1-w_j)} + \sqrt{w_i w_j} \right)$. This is very nearly Ω_{ij} so long as $|w_i - w_j|$ is not close to 1. For the calibrations, we use $w_c = 0.5$, $w_d = 0.125$, for which this quantity equals $0.91 \times \Omega_{cd}$.

large and volatile variance premium and features of its return predictability (iv) the model is consistent with consumption and return predictability by the price-dividend ratio. It is also important, particularly regarding (iii), that the model remain consistent with the dynamics of conditional return volatility. We show that the model is successful at matching many of these data features. In addition to highlighting the specific results, we provide further discussion of the roles various model parameters have in generating the results.

In Table V we provide the parameter specification for the model economy described above with gamma-distributed jump shocks to x_t . Tables VI–X provide the data and corresponding model based statistics. In comparing the model fit to the data we provide model based finite sample statistics. Specifically, we present the model based 5%, 50% and 95% percentiles for the statistics of interest generated from 1000 simulations, with each statistic calculated from a sample of size equal to its data counterpart. The decision interval in the model is assumed to be monthly. For the consumption and dividend dynamics we utilize the longest sample available; hence, the simulations are based on 924 monthly observations which are time-averaged to an annual sample of length 77 as in the annual data (1930:2006). We provide similar statistics for the ‘standard’ asset pricing moments, such as the mean and volatility of the market and risk free rate. To obtain the real risk-free rate, we used the monthly return on the 3-month T-Bill minus the realized CPI return for that month, and multiplied by 12 and square root of 12 to get an annualized mean and standard deviation respectively. Recall that for the variance premium-related statistics the data is monthly and available only from the latter part of the sample (1990.1-2007.3). Thus, the model’s variance premium-related statistics are based on the last 207 monthly observations in each of the 1000 simulations. Under the view that the model is the appropriate data generating process, the data point estimates should be within the 90% confidence interval generated by the model. Nonetheless, for completeness we also provide HAC robust standard errors for the data statistics.

Consumption and Dividend Dynamics

Before examining asset pricing moments, and in particular the variance premium properties, it is important to establish that the model’s cashflow (consumption and dividend) dynamics are consistent with the data. The calibrated parameters governing the persistence of x_t , the long run prospects of the economy, as well as the leverage parameter, are generally close to those in Bansal and Yaron (2004). The top panel in Table VI shows that the model captures quite well key moments of annualized consumption and dividend growth.

The data-based mean and volatility of dividends and consumption growth fall well within the 90% confidence interval generated by the model, and are in fact very close to the median estimates from the model. Our configuration has allocated some of the source of variation in persistent consumption growth to non-Gaussian shocks. Specifically, we calibrate the jump intensity so jumps to the x_t component arrive at an average rate of 0.8 per year. To investigate whether our calibration of this non-Gaussian component generates some undesirable effects at higher frequencies, Table VI also reports the kurtosis of consumption growth at the quarterly frequency (note this data corresponds to 1947.2-2006.4). The model matches this feature well, showing that at the higher quarterly frequency the jump effects do not cause inconsistency between the model and data in terms of higher moments of consumption growth.

Equity Returns and Risk Free Rate

Table VII presents a number of the model's asset pricing implications. This table pertains to annual data on the market, risk free rate and price-dividend ratio. As discussed earlier, the corresponding model statistics are also time averaged annual figures. Again, the model does a good job in capturing the equity premium, the volatility of the market return and the low mean and volatility of the risk free rate. Hence, the results in this table indicate that the jump component does not alter the ability of the long run risk model to generate cashflow and asset pricing dynamics consistent with the data.

The bottom of Table VII looks at higher moments of the market return. The non-normality of equity returns, especially at the monthly frequency, is well known. In particular, the skewness of returns provides information on the shape of the return shock distribution, while the kurtosis is informative about the role of stochastic volatility and jumps. These features of the shock distribution clearly have direct implications for option prices and the variance premium. The bottom four entries in Table VII show that the model captures these data statistics very well. As the table shows, monthly returns are negatively skewed. The skewness inherent in both the shocks to x_t , and volatility σ_t^2 , help the model capture this feature of the data. Moreover, there is significant excess kurtosis in monthly returns, which the model also captures.

Variance Premium

The top panel in Table VIII provides several statistics pertaining to the dynamics of conditional return variance and properties of the variance premium, all at the monthly

frequency. As is well known in the literature, conditional return variance is highly stochastic and is significantly autocorrelated. Clearly, a model that wants to confront the variance premium should also confront these aspects of the data. The first three rows of Table VIII provide the volatility and the first two autocorrelations of the conditional variance of the market return, based on the proxy constructed in section 2. The table shows that the model's median values are right in line with the data estimates. Matching these is important, since trying to capture the large variance premium may lead one to blindly introduce large volatility shocks within the model. These moments restrict the amount of underlying volatility shocks and the persistence of their impact. While remaining consistent along these lines, the model is able to generate the large and volatile variance premium. Table VIII shows that the median of the model generated mean variance premium is slightly smaller than its data counterpart. However, the model's 90% confidence interval easily includes the data point estimate of the mean variance premium.¹⁷ The median of the model's variance premium volatility is somewhat larger than the data point estimate, though again the data point estimate is well within the model's 90% confidence interval. The next two entries in Table VIII show that the model also captures well the degree to which the variance premium distribution is fat-tailed, in particular, its large skewness and kurtosis.

The inclusion of jump shocks is important for the model's ability to jointly reconcile the many moments of the conditional return variance and variance premium that are discussed above. In comparison, to match the large magnitude of the variance premium, for example, a model with only Gaussian shocks needs to compensate for the lack of jumps through other channels. We present such a model in Appendix B (see also Bollerslev, Tauchen, and Zhou (2009)). Such a model generates no level difference and must therefore attempt to match the variance premium by greatly amplifying the drift difference. This can be problematic, as based on our experience, it is likely to require a volatility of conditional variance that is much too high relative to the data. A related point has been made by some reduced-form studies. Broadie, Chernov, and Johannes (2007) argue that in order to match the sharp changes that occur periodically in return volatility, a CIR process would necessarily imply too much volatility of the conditional variance. They demonstrate that their estimate of a CIR model generates far too little kurtosis in changes in the VIX and therefore argue for a

¹⁷In relatively short finite samples, the 5%-95% interval for moments of heavy tailed distributions may be difficult to compare to the HAC standard errors of these moments' estimates. For comparison, we also computed the HAC standard errors inside the model for each of our simulations. We note that the median HAC standard error for the mean of the variance premium is 1.63, whereas this number is 0.93 in Table VIII, so that along this dimension the model and data seem to correspond reasonably well.

model with volatility jumps. As the last line of the top panel in Table VIII shows, the model performs well along this dimension.

At the outset of this paper, we highlighted the ability of the variance premium to predict future returns. The middle panel of Table VIII provides the corresponding results for univariate projections of excess returns on the variance premium. The model is able to replicate the return predictability of the variance premium as found in the data. It is interesting to note that the projection coefficients have the right sign and are well within one standard error of the data. Moreover, the R^2 of these predictability regressions are quite large for these short horizons. The model median R^2 for the one-month ahead projection is about 2% and the 90% finite sample confidence band for R^2 clearly includes the 1.5% R^2 from the data (as well as the robust regression and lagged variance premium predictive R^2 s given in Table IV). Furthermore, the 5.9% R^2 for the 3-month ahead projection is quite close to the model's median R^2 estimate. In order to assess whether the price-dividend ratio crowds out the variance premium's ability to predict returns, the bottom panel of Table VIII provides multivariate projections of excess returns on the variance premium and price-dividend ratio for the data and the model. As in the data, the model's coefficient for the variance premium remains positive and of similar magnitude. Furthermore, in the model the price-dividend ratio coefficient has the familiar negative sign and similar magnitude to that in the data. Finally, note that the predictive R^2 are larger than their univariate counterpart and increase with horizon as in the data. Overall, the results of this table indicate that this augmented long run risk model can capture quite well the cashflow, asset pricing, and variance premium moments in the data.

Long-Horizon Predictability

In this section we investigate a few additional model implications that focus on aspects of long-horizon predictability. It is in these moments that the impact of $\bar{\sigma}_t^2$, the highly persistent and smoothly moving process that drives the mean of long run volatility, becomes apparent. The persistence of this process is similar to that used in Bansal, Kiku, and Yaron (2007b) and Bansal and Yaron (2004). We focus on four projections that often appear in the literature. These use the price-dividend ratio to predict (i) consumption growth (ii) excess returns (iii) consumption volatility and (iv) return volatility.

Table IX investigates the predictability of consumption growth and excess returns by the price-dividend ratio. The top panel shows that in the data there is significant (moderate)

consumption predictability at the 1-year (5 year horizon) respectively. The model based median R^2 s indicate moderate consumption predictability at all horizons. Moreover, the data-based R^2 s are easily within the model based R^2 confidence bands. We also note (not reported) that the model captures the positive sign of the projection of consumption growth onto the price-dividend ratio. The next panels in Table IX provide the predictive R^2 s from projecting future excess returns on the price-dividend ratio. The second panel refers to univariate OLS projections. The model's median predictive R^2 is very close to the data for the 1-year horizon (and, not reported, for all horizons up to one year). The median predictive R^2 s in the model are somewhat smaller than their data counterparts for the 3 and 5 year horizons, though the three year R^2 is well within the model based 90% confidence interval. It is important to note that in the data the predictive R^2 s for 3 and 5 years rise at a rate that is faster than linear in the horizon. It is well known in the literature that there are many factors potentially plaguing the estimates and inference of this projection (see Hodrick (1992), Stambaugh (1999), Boudoukh, Richardson, and Whitelaw (2009), Bansal, Kiku, and Yaron (2007a)). One method of addressing the concerns of small sample with overlapping observations and persistent regressors raised by these studies is to calculate long-horizon statistics implied by a VAR estimated at a monthly horizon. The third panel in Table IX provides predictive R^2 s implied by a first-order VAR that includes the price-divided ratio and the market excess return. The data and model VAR based predictive R^2 s are closely aligned and are both around 10% at the 3 and 5-year horizons.¹⁸

Table X investigates the predictability of future consumption and return volatility by the price-dividend ratio. The consumption volatility measure is calculated by summing up for the relevant horizon the absolute values of the residuals from an AR(1) model of consumption growth, and then taking the log of this quantity. The return volatility is calculated as the annualized standard deviation of monthly excess returns over the relevant horizon. The slope and predictive R^2 s for both returns and consumption growth are well within the the model based confidence bands. Furthermore, both the model and data generate negative slope coefficients. This negative relationship between the price-dividend ratio and consumption and return volatility is an implication of the model when agents have preference for early resolution of uncertainty (recall $A_\sigma < 0$).

¹⁸In the data the univariate and VAR based R^2 s are very close to each other for all horizons from 1-month to 1 year but diverge at 3 years. In conjunction with the model's performance on the VAR, this evidence highlights the fragility of inferences on long horizon return predictability based on univariate price-dividend projections.

As a final reflection, we note that though we do not formally estimate the model, the number of reported statistics greatly exceeds the number of parameters in the model, so that capturing the long list of moments in Tables VI-X is by no means an obvious outcome. Furthermore, it is important to recognize that the preference parameters used here (e.g., risk aversion of about ten and IES greater than one) are similar in magnitude to those used and estimated successfully in other applications of the Long Run Risks model (e.g. Bansal, Kiku, and Yaron (2007b)). This provides some cross-validation of these type of preferences.

Normally Distributed Jumps

Table XI provides the cashflow output for a model in which the jump sizes of x_t are drawn from a normal distribution. We provide this experiment to demonstrate that our main quantitative results are not particularly sensitive to the choice of gamma shocks. In order to allow the model to generate an equity premium that is broadly in line with the data, risk aversion is set to 10.¹⁹ Comparing these two configurations also allows one to evaluate the relative merits of the non-symmetric jumps present in the gamma distribution. One can easily observe from Table XI that this model also produces cashflow statistics that are consistent with their data counterparts. Table XII provides the equity premium and risk-free rate for this model configuration, which are also consistent with their data counterparts. In essence, it is quite difficult to distinguish this configuration from the one given in Tables VI and VII purely along these cashflow and return dimensions. The main fit deterioration of this model relative to the one with gamma shocks is in its lower level of the variance premium and higher variance premium kurtosis. Furthermore, the skewed shock structure emanating from the gamma specification leads to a better behaved equity return skewness.

Given the earlier discussion of the level and drift difference, it is interesting to note the quantitative contribution of these two parts to the variance premium under our calibrations. For the results in Table VIII, the corresponding level difference component has a median size and standard deviation that are approximately 78 and 88 percent of the total variance premium's size and standard deviation, respectively. For Table XIII, the corresponding percentages are 74 and 83. Hence, under both calibrations, the level difference accounts for the bulk of the variance premium's size and volatility, though the drift difference also makes a nontrivial contribution.

¹⁹Other than this parameter change, we simply convert the demeaned gamma distribution to a zero mean normal distribution with the same standard deviation. In our notation we set $\sigma_x = \mu_x$.

Shutting Off Jump Shocks

In Table XIV we conduct a three part comparative statistics exercise on the model of Table V by shutting off the Poisson jump shocks. The first panel, labeled Model 1-A, is for a model that shuts off only the Poisson component of σ_t^2 ($l_{1,\sigma} = 0$). The second panel, Model 1-B, turns off only the Poisson component of x_t ($l_{1,x} = 0$). Finally, the third panel shuts off both Poisson processes. We do these comparative statics in order to provide some quantitative assessment of the role of these jump shocks, which are relatively large but infrequent.²⁰ What is interesting is that the cashflow dynamics still match quite well the consumption and dividend data statistics. However, now the three panels' median estimates for the market return drop significantly to a range of about 2.9%-4.6% (from almost 7.0% in the case of Table VII). This happens in spite of the fact that the median volatility of the market return drops by only 1-2% in each case. It is also the case that in these situations the unconditional level of the price-dividend ratio is too large. Nonetheless, one could argue that in each panel these moments are still reasonable asset pricing moments, which many other models fail to match. Where the largest discrepancy appears is in the variance premium related moments. The mean and the volatility of the variance premium are quite small in the first panel and almost zero in the last two.²¹ Moreover, when the jump in σ_t^2 is shut off in the first and third panels, the volatility of the conditional variance of the market return is quite far below its data counterpart. In both these cases the 90% confidence interval does not come close to its corresponding data statistic. Finally, and almost by construction, the predictability regressions in all three panels yield median R^2 s that are far below their data counterparts and the predictive regression coefficients are very unstable.

As a final thought, it is worth commenting on the difference between ex-ante risk and ex-post realizations in the presence of somewhat infrequent and influential jump shocks, such as those in the model. As evidenced by our analysis, risk premia levels and time-variation clearly show the effect (ex-ante) of these risks. However, actual (ex-post) return shock realizations materialize less frequently. Thus, there may be relatively extended periods where large, negative shocks are not realized, though the risk of them is real and varies through

²⁰In providing these results it should be clear that the overall variance of x_t and σ_t is lower than their counterparts in Table VI since we shut off the jump shocks. While other permutations are possible, this allows one to evaluate the contributions of the jumps within our current specification.

²¹There is no requirement that the means of the variance premium under the first two panels sum to the mean variance premium under the benchmark model. The inclusion of x_t jumps raises the importance of the jump intensity variable σ_t^2 and increases the magnitude of A_σ . This reinforcing of risk channels is an interesting feature of structural models utilizing recursive preferences (see also Bansal and Yaron (2004)).

time. This is reflected in the finite-sample R^2 distributions, which correspond to a sample of the same length as the corresponding data. Note that in Tables VIII and XIII, the right tails of the R^2 distributions include periods where predictability by the variance premium is *very* high. These right-tail samples did not experience any significant negative realizations following spikes in the variance premium. On the other hand, the median statistics show that the negative realizations that eventually occur greatly diminish the estimated return predictability. The population R^2 s implied by the model are close to these median values.

6 Conclusion

During the recent macroeconomic and financial turmoil the VIX index has been a focal point of attention among a broad swath of market participants, serving as the ‘fear gauge’ for the market’s concerns of surprise economic shocks. As uncertainty increased and worries about a prolonged slowdown in growth heightened, the level and volatility of the VIX reached unprecedented levels. The prominence of the VIX during this period highlights the importance of understanding its fluctuations and its embedded risk premium, the variance premium. This paper shows that the variance premium is useful for measuring agents’ perceptions of uncertainty and the risk of influential shocks to the economic state vector. In addition, it provides a useful vehicle for understanding what preferences are able to map this risk into observed asset prices. We demonstrate that a risk aversion greater than one and a preference for early resolution of uncertainty correctly signs the variance premium and the coefficient from a predictive regression of returns on the variance premium. In addition, we show that time variation in economic uncertainty is a minimal requirement for qualitatively generating a positive, time varying variance premium that predicts excess stock returns. Finally, we show that an extended Long Run Risks model, with jumps in uncertainty and the long-run component of cashflows, can generate many of the quantitative features of the variance premium while remaining consistent with observed aggregate dynamics for dividends and consumption as well as standard asset pricing data, such as the equity premium and risk free rate. Utilizing a persistent volatility component, the model further matches key evidence on long-horizon predictability. We find that the jump shocks are helpful in matching the standard asset pricing data, and that they are particularly important for our ‘nonstandard’ moments related to conditional volatility, the variance premium, and its predictive regression for market returns.

Whereas this paper offers a risk-centric explanation for the size and time-variation of the variance premium and its return predictability, there may be alternative and potentially complimentary avenues for generating these. For example, one may consider mechanisms that directly generate variation in risk prices, e.g. habit-formation or variation in investor ambiguity/desire for robustness. For instance, a high level of investor ambiguity can effectively increase risk prices, thereby generating an increased variance premium and equity premium. Hansen and Sargent (2008) demonstrate that a desire for robustness can lead to interesting time-varying misspecification risk premia components. It also seems reasonable that periods with elevated feelings of ambiguity be correlated with periods of elevated risk. Endogenously generating such dynamics could be interesting, and the variance premium and other options-based information could provide valuable empirical underpinnings for assessing a model that does this.

More generally, risk attitudes toward uncertainty play an important role in interpreting asset markets. The Long Run Risks model has channels for several priced risk factors, including the level of uncertainty and its rate of change. An interesting direction for future research is determining the extent to which these risks are also important in the cross-section of returns. Bansal, Kiku, and Yaron (2007b) utilize an uncertainty factor in the cross-section of returns within the long-run risks framework, but are constrained to identify it based solely on cashflows. The evidence in this paper suggests that derivative markets and high frequency measures of variation should be very useful at identifying these risk factors. Interesting implications could therefore arise from jointly using cashflows and derivative markets to understand the influence of uncertainty on the cross-section.

Appendix

A Solving the Model

A.1 Solving for A and A_0

To solve for asset prices we first solve for the return on the wealth claim, $r_{c,t+1}$. As discussed in Section 3.3, we conjecture that the log wealth-consumption ratio follows $v_t = A_0 + A'Y_t$. We use the Euler equation (3) to determine A and A_0 . This equation must hold for the returns on all assets, including the return on the aggregate consumption claim. Thus, set $r_{j,t+1} = r_{c,t+1}$ in (3), and substitute in $m_{t+1} = \theta \ln \delta - \frac{\theta}{\psi} \Delta c_{t+1} + (\theta - 1)r_{c,t+1}$. Then replace $r_{c,t+1}$ with its log-linearization (5) to obtain:

$$E_t \left[\exp \left(\theta \ln \delta - \theta \left(\frac{1}{\psi} - 1 \right) \Delta c_{t+1} + \theta \kappa_0 + \theta \kappa_1 v_{t+1} - \theta v_t \right) \right] = 1$$

Now substitute in the conjecture for v_t to get the equation in terms of A_0 and A . Also, replace Δc_{t+1} with $e_c' Y_{t+1}$, where e_c denotes the vector that selects Δc_{t+1} from Y_{t+1} . Collecting the constants and the terms in Y_t and Y_{t+1} yields the following:

$$E_t \left[\exp \left(\theta \ln \delta + \theta (\kappa - 1) A_0 + \theta \kappa_0 - \theta A' Y_t + \left(\theta \left(1 - \frac{1}{\psi} \right) e_c + \theta \kappa_1 A \right)' Y_{t+1} \right) \right] = 1 \quad (\text{A.1.1})$$

In order to compute the left-hand side expectation it is useful to establish the following functional relationship:

For $u \in \mathbb{R}^n$:

$$E [\exp(u' Y_{t+1} | Y_t)] = \exp(\mathbf{f}(u) + \mathbf{g}(u)' Y_t) \quad (\text{A.1.2})$$

$$\mathbf{f}(u) = \mu' u + \frac{1}{2} u' h u + l'_0 (\psi(u) - 1) \quad (\text{A.1.3})$$

$$\mathbf{g}(u) = F' u + \frac{1}{2} [u' H_i u]_{i \in \{1 \dots n\}} + l'_1 (\psi(u) - 1) \quad (\text{A.1.4})$$

and $[u' H_i u]_{i \in \{1 \dots n\}}$ denotes the $n \times 1$ vector with i -th component equal to $u' H_i u$.

Proof. Substitute for Y_{t+1} in the left-hand side expectation and break the resulting expres-

sion into three terms:

$$\begin{aligned} E_t [\exp(u'Y_{t+1})] &= E_t [\exp(u'(\mu + FY_t + G_tZ_{t+1} + J_{t+1}))] \\ &= \exp(u'\mu + u'FY_t)E_t(\exp(u'G_tZ_{t+1}))E_t(\exp(u'J_{t+1})) \end{aligned}$$

where the second line follows from the conditional independence of the Gaussian and jump shocks. Evaluating the two conditional expectations gives:

$$\begin{aligned} E_t(\exp(u'G_tZ_{t+1})) &= \exp\left(\frac{1}{2}u'G_tG_t'u\right) = \exp\left(\frac{1}{2}u'h u + \frac{1}{2}\sum_i u'H_iu'Y_t(i)\right) \\ E_t(\exp(u'J_{t+1})) &= \exp(\lambda_t'(\psi(u) - 1)) = \exp(l_0'(\psi(u) - 1) + (l_1Y_t)'(\psi(u) - 1)) \end{aligned}$$

Multiplying the three terms together and collecting the constants and Y_t terms into the functions $\mathbf{f}(u)$ and $\mathbf{g}(u)$, respectively, gives the result. ■

Continuing with the derivation, use (A.1.2) to evaluate the expectation in (A.1.1). Then, taking logs of both sides results in the following equation:

$$\begin{aligned} 0 = \theta \ln \delta + \theta \kappa_0 + \theta(\kappa_1 - 1)A_0 + \mathbf{f}\left(\theta\left(1 - \frac{1}{\psi}\right)e_c + \theta\kappa_1A\right) \\ + \left[\mathbf{g}\left(\theta\left(1 - \frac{1}{\psi}\right)e_c + \theta\kappa_1A\right) - A\theta\right]' Y_t \quad (\text{A.1.5}) \end{aligned}$$

This equation is a restriction that must hold for all values of Y_t . This implies that the term multiplying Y_t must be identically 0 and therefore that the constant is 0 as well. The result is the following system of $n + 1$ equations in A_0 and A :

$$0 = \theta \ln \delta + \theta \kappa_0 + \theta(\kappa_1 - 1)A_0 + \mathbf{f}\left(\theta\left(1 - \frac{1}{\psi}\right)e_c + \theta\kappa_1A\right) \quad (\text{A.1.6})$$

$$0 = \mathbf{g}\left(\theta\left(1 - \frac{1}{\psi}\right)e_c + \theta\kappa_1A\right) - A\theta \quad (\text{A.1.7})$$

Closed-form expressions for the components of A are attainable for a number of specifications. Bansal and Yaron (2004) provide expressions for their specification, while Tauchen (2005) shows how to solve for A_σ when the volatility process is of the square-root form. Quasi closed-form expressions are even possible in some specifications that have both jumps and square-root volatility. However, in general, closed-form expressions for A and A_0 are

unavailable and the solutions must be found numerically. As the the linearization constants κ_0 and κ_1 are endogenous, they are solved for jointly with A and A_0 by adding equations to the system in (A.1.6)-(A.1.7). Further details regarding the numerical solution of this system are given in the next subsection.

A.2 Numerical Solution

The log-linearization constants are given by $\kappa_1 = \frac{e^{E(v)}}{1+e^{E(v)}}$ and $\kappa_0 = \ln(1 + e^{E(v)}) - \kappa_1 E(v)$. Inverting the definition of κ_1 gives the useful identity:

$$\ln \kappa_1 - \ln(1 - \kappa_1) = E(v_t) = A_0 + A'E(Y_t) \quad (\text{A.2.1})$$

Substituting this in for $E(v_t)$ in the definition of κ_0 gives an expression for κ_0 purely in terms of κ_1 :

$$\kappa_0 = -\kappa_1 \ln \kappa_1 - (1 - \kappa_1) \ln(1 - \kappa_1) \quad (\text{A.2.2})$$

As (A.2.1) shows, the value of κ_1 depends directly on the values of A and A_0 and is therefore endogenous to the model. Moreover, from (A.1.6) and (A.1.7) we have that the values of the A coefficients themselves depend on the log-linearization constants. Therefore, (A.2.1) and (A.2.2) must be solved jointly with (A.1.6) and (A.1.7). One way to do this is to simply augment the system of equations. Instead, we keep the numerically solved system the same size using the following identity, which is easily derived from (A.2.1) and (A.2.2):

$$\kappa_0 + (\kappa_1 - 1)A_0 = -\ln \kappa_1 + (1 - \kappa_1)A'E(Y_t)$$

We eliminate κ_0 and A_0 from the numerically solved system by substituting this identity into (A.1.6) to get

$$0 = \theta \ln \delta + \theta (-\ln \kappa_1 + (1 - \kappa_1)A'E(Y_t)) + \mathbf{f} \left(\theta \left(1 - \frac{1}{\psi}\right) e_c + \theta \kappa_1 A \right) \quad (\text{A.2.3})$$

and solving (A.2.3) together with (A.1.7) to obtain κ_1 and A . Using the identities above, one can then solve directly for A_0 and κ_0 in terms of the values of κ_1 and A .

A.3 Solving for the Market Return

The procedure for solving for $A_{0,m}$ and A_m is similar to the one used to for determining A_0 and A_1 . The Euler equation is again used to derive a system of equations whose solution determines $A_{0,m}$ and A_m . To this end, apply the Euler equation to the market return by setting $r_{j,t+1} = r_{m,t+1}$ in (3). Then making the follow substitutions into the Euler equation to get it in terms of the A_m coefficients and model primitives: (1) replace m_{t+1} with (6) (2) substitute in (8) for $r_{m,t+1}$ and (3) replace $v_{m,t}$ with the conjectured form $A_{0,m} + A'_m Y_t$. After collecting terms in Y_t and Y_{t+1} and simplifying the resulting equation is:

$$E_t[\exp(\theta \ln \delta - (1 - \theta)(\kappa_1 - 1)A_0 - (1 - \theta)\kappa_0 + \kappa_{0,m} + (\kappa_{1,m} - 1)A_{0,m} + ((1 - \theta)A - A_m)' Y_t + (e_d + \kappa_{1,m} - \Lambda)' Y_{t+1})] = 1 \quad (\text{A.3.1})$$

where e_d is the vector that selects Δd_{t+1} from Y_{t+1} . Evaluating the expectation using the result in (A.1.2), taking logs, and setting the constant and the term multiplying Y_t to 0, results in the following system of equations in $A_{0,m}$ and A_m :

$$\begin{aligned} 0 &= \theta \ln \delta - (1 - \theta)(\kappa_1 - 1)A_0 - (1 - \theta)\kappa_0 + \kappa_{0,m} + (\kappa_{1,m} - 1)A_{0,m} + \mathbf{f}(e_d + \kappa_{1,m} - \Lambda) \\ 0 &= \mathbf{g}(e_d + \kappa_{1,m} - \Lambda) + (1 - \theta)A - A_m \end{aligned}$$

A.4 Equity Premium

This appendix derives a concise expression for the equity premium in the presence of compound Poisson shocks that generalizes the familiar expression that arises when there are only Gaussian shocks. To derive the expression, it is useful to separate expectations involving the Gaussian shocks from the ones involving the compound Poisson (jump) shocks when taking expectations over components of the pricing kernel and returns. For notational convenience in what follows, we use a superscript J over a variable to denote the jump shock component of that variable and a superscript g over a variable to denote its Gaussian component plus any constants. We can then decompose the variable as a sum of its two types of components: for example, $m_{t+1} = m_{t+1}^g + m_{t+1}^J$.

To derive the market risk premium, evaluate the Euler equation (3) as follows:

$$\begin{aligned} 1 &= E_t[\exp(m_{t+1} + r_{m,t+1})] \\ &= e^{E_t(m_{t+1}^g) + E_t(r_{m,t+1}^g) + \frac{1}{2}\text{var}_t(r_{m,t+1}^g) + \frac{1}{2}\text{var}_t(m_{t+1}^g) + \text{cov}_t(m_{t+1}^g, r_{m,t+1}^g)} E_t[e^{m_{t+1}^J + r_{m,t+1}^J}] \quad (\text{A.4.1}) \end{aligned}$$

On the second line, the multiplicative separation of the expectation for the Gaussian terms and Poisson terms is due to the conditional independence of the Gaussian and Poisson shocks. Letting $R_{m,t+1}$ denote the gross (simply-compounded) return on the market and M_{t+1} the pricing kernel, we note that:

$$\begin{aligned} E_t(R_{m,t+1}) &= E_t[e^{r_{m,t+1}^J}] e^{E_t(r_{m,t+1}^g) + \frac{1}{2}\text{var}_t(r_{m,t+1}^g)} \\ E_t(M_{t+1}) &= E_t[e^{m_{t+1}^J}] e^{E_t(m_{t+1}^g) + \frac{1}{2}\text{var}_t(m_{t+1}^g)} \end{aligned}$$

Dividing both sides of (A.4.1) by $e^{\text{cov}_t(m_{t+1}^g, r_{m,t+1}^g)} E_t[e^{m_{t+1}^J + r_{m,t+1}^J}]$, then multiplying both sides by $E_t[e^{r_{m,t+1}^J}] E_t[e^{m_{t+1}^J}]$, and substituting in $E_t(R_{m,t+1})$ and $E_t(M_{t+1})$ gives:

$$E_t(R_{m,t+1}) E_t(M_{t+1}) = e^{-\text{cov}_t(m_{t+1}^g, r_{m,t+1}^g)} E_t[e^{r_{m,t+1}^J}] E_t[e^{m_{t+1}^J}] / E_t[e^{m_{t+1}^J + r_{m,t+1}^J}]$$

Let $R_{f,t}$ denote the gross (simply-compounded) risk-free rate. Then, substituting in $E_t(M_{t+1}) = R_{f,t}^{-1}$ and taking logs of both sides, we get an expression for the (log) equity premium:

$$\ln E_t(R_{m,t+1}) - r_{f,t} = -\text{cov}_t(m_{t+1}^g, r_{m,t+1}^g) + \ln E_t[e^{r_{m,t+1}^J}] + \ln E_t[e^{m_{t+1}^J}] - \ln E_t[e^{m_{t+1}^J + r_{m,t+1}^J}] \quad (\text{A.4.2})$$

When there are no compound Poisson shocks, this reduces to the familiar expression that has only the covariance term on the right side. We can now substitute in for the terms on the right side of (A.4.2). Using (7) for the pricing kernel and (9) for the market return gives $\ln E_t[e^{m_{t+1}^J}] = \lambda'_t(\psi(-\Lambda) - 1)$ and $\ln E_t[e^{r_{m,t+1}^J}] = \lambda'_t(\psi(B_r) - 1)$. Moreover, $\ln E_t[e^{m_{t+1}^J + r_{m,t+1}^J}] = \lambda'_t(\psi(B_r - \Lambda) - 1)$. Finally, $-\text{cov}_t(m_{t+1}^g, r_{m,t+1}^g) = B'_r G_t G'_t \Lambda$. Substituting into (A.4.2) and rearranging, the (log) equity premium can be written:

$$\ln E_t(R_{m,t+1}) - r_{f,t} = B'_r G_t G'_t \Lambda + \lambda'_t(\psi(B_r) - 1) - \lambda'_t(\psi(B_r - \Lambda) - \psi(-\Lambda)) \quad (\text{A.4.3})$$

The first term on the left side of (A.4.3) is the contribution to the equity premium from the Gaussian shocks whereas the other two terms combine to give the contribution of the compound Poisson (jump) shocks.

A.5 Risk-free Rate

To derive the risk-free rate at time t , set $r_{j,t+1} = r_{f,t}$ in the Euler equation (3). Then substitute in for m_{t+1} and collect the constant terms, terms in Y_t , and Y_{t+1} . To evaluate the expectation, use the result in (A.1.2). Then, taking logs of both sides of the equation and solving for $r_{f,t}$ gives:

$$r_{f,t} = r_{f,0} - (\mathbf{g}(-\Lambda) - (\theta - 1)A)' Y_t \quad (\text{A.5.1})$$

where $r_{f,0} = -\theta \ln \delta + (1 - \theta) [\kappa_0 + (\kappa_1 - 1)A_0] - \mathbf{f}(-\Lambda)$.

B A Variance of Variance Model

The model discussed in this Appendix has pedagogical value for understanding the drift difference, variance premium, and return predictability. The model is a simplified version of the reference model in the main text, though there is a new state variable. The simplifications relative to the reference model are that (i) the Poisson shocks are shut off (i.e. $\lambda_t \equiv 0$) (ii) the Gaussian shocks are uncorrelated and (iii) the variable $\bar{\sigma}_t^2$ (the long-run mean of σ_t^2) is set constant, i.e. its shocks are shut off and it is no longer a state variable. We add a new state variable, q_t , that drives the volatility of innovations to σ_{t+1}^2 . In other words, q_t is the conditional variance of shocks to σ_{t+1}^2 . The processes for these two state variables are then written as:

$$\sigma_{t+1}^2 = \bar{\sigma}^2 + \rho_\sigma(\sigma_t^2 - \bar{\sigma}^2) + q_t^{1/2} z_{\sigma,t+1} \quad (\text{B.1})$$

$$q_{t+1} = \bar{q} + \rho_q(q_t - \bar{q}) + \varphi_q z_{q,t+1} \quad (\text{B.2})$$

Note that this specification maps easily into the general framework in (4) and is very similar to a model analyzed in Tauchen (2005). Solving the model for prices of risk, the market return and return variance follows the general procedure outlined in the main text. Under this model, we get that $v_t = A_0 + A_x x_t + A_\sigma \sigma_t^2 + A_q q_t$ and, importantly, that $\Lambda = (\gamma, (1 - \theta)\kappa_1 A_x, (1 - \theta)\kappa_1 A_\sigma, (1 - \theta)\kappa_1 A_q, 0)'$, i.e. shocks to q_t are also priced. The price-dividend ratio is $v_{m,t} = A_{0,m} + A_{x,m} x_t + A_{\sigma,m} \sigma_t^2 + A_{q,m} q_t$. The market return variance is given by (10). Writing out all the terms in expanded form gives:

$$\text{var}_t(r_{m,t+1}) = \sigma_{r,t}^2 = (\beta_{r,x}^2 \varphi_x^2 + \varphi_d^2) \sigma_t^2 + \beta_{r,\sigma}^2 q_t + \beta_{r,q}^2 \varphi_q^2 \quad (\text{B.3})$$

where $\beta_r = \kappa_{1,m}A_m + e_d$ exactly as in Section 3.3.2.

Since this model is a pure Gaussian model, the level difference is 0. The variance premium is then equal to the drift difference, which is nonzero because σ_t^2 and q_t have different drifts under P and Q :

$$E_t^Q[\sigma_{t+1}^2] - E_t[\sigma_{t+1}^2] = -\lambda_\sigma q_t \quad (\text{B.4})$$

$$E_t^Q[q_{t+1}] - E_t[q_{t+1}] = -\lambda_q \varphi_q^2 \quad (\text{B.5})$$

It then easily follows that:

$$vp_{t,t+1} = -(\beta_{r,x}^2 \varphi_x^2 + \varphi_d^2) \lambda_\sigma q_t - \beta_{r,\sigma}^2 \lambda_q \varphi_q^2 \quad (\text{B.6})$$

From (B.6) we see that time-variation in this model's variance premium is driven by q_t , the conditional variance of shocks to σ_t^2 . Since σ_t^2 controls the conditional variance of the other shocks, q_t is like the 'variance of variance'. A high q_t indicates high uncertainty about future conditional variance, and this uncertainty is reflected in the variance premium.

Finally, the conditional equity premium is:

$$\beta_{r,x} \lambda_x \varphi_x^2 \sigma_t^2 + \beta_{r,\sigma} \lambda_\sigma q_t + \beta_{r,q} \lambda_q^2 \varphi_q \quad (\text{B.7})$$

This shows that the loading on q_t is priced. When $\gamma > 1$, $\psi > 1$, then $\lambda_\sigma < 0$ (the agent is averse to increases in volatility/uncertainty) and $\beta_{r,\sigma} < 0$ (increases in volatility decrease the market return). For these preferences, these last two expressions then show that there is a positive covariation between $vp_{t,t+1}$ and the conditional equity premium, i.e. $vp_{t,t+1}$ will predict stock returns. Simple algebra shows that the projection coefficient of (B.7) on (B.6) is $\frac{-\beta_{r,\sigma}}{\beta_{r,x}^2 \varphi_x^2 + \varphi_d^2}$, which is positive for $\gamma > 1$, $\psi > 1$.

C Drift Difference with Jumps

Here we conclude the discussion about the drift difference from section 4.3 by discussing the drift difference when the Poisson jumps are included. Since section 4.3 already derives the drift difference for the Gaussian-related part of $\text{var}_t(r_{m,t+1})$, we now consider only the Poisson-related part. From (10) and (21) this is $B_r^{2'} \text{diag}(\psi^{(2)}(*)) \lambda_t$ where $* = 0$ un-

der P and $* = -\Lambda$ under Q . Thus, under P the one period drift in this quantity is: $B_r^{2'} \text{diag}(\psi^{(2)}(0)) [E_t^P(\lambda_{t+1}) - \lambda_t]$, while under Q it is: $B_r^{2'} \text{diag}(\psi^{(2)}(-\Lambda)) [E_t^Q(\lambda_{t+1}) - \lambda_t]$. The drift difference is then just the Q -related term minus the P -related term. While we can use the derived dynamics for Y_t under Q and P to write the λ_t expressions more explicitly, we stop at this point and simply note that, as in the pure Gaussian case, the choice of preferences determines the sign of this jump-related component of the drift difference. The main issue is the relation between $E_t^Q(\lambda_{t+1})$ and $E_t^P(\lambda_{t+1})$ and it parallels the discussion above of the Gaussian case, e.g. $E_t^Q(\lambda_{t+1}) > E_t^P(\lambda_{t+1})$ when $\gamma > 1, \psi > 1$. Finally, we note that the Poisson part of the drift difference is a linear function of the drift in λ_t . When the drift in λ_t is linear in λ_t , as in our model, then this represents the second component of $vp_{t,t+1}$ that is driven by the latent jump intensity.

C.1 Adding Up the Parts

To get the total $vp_{t,t+1}$ just add the expressions for the level difference and drift difference. Algebraically, the expression is a bit messy. However, our discussion has shown that the mapping from preferences to the sign of each of the components is consistent, so the components generally augment each other. We have discussed how the components reveal latent elements of the state vector that are important drivers of conditional risk premia. Although it is not conceptually difficult to derive algebraic expressions for the projection coefficient of excess returns on $vp_{t,t+1}$, they do not add much insight beyond our previous discussions, which point out that they will have the right sign under the $\gamma > 1, \psi > 1$ preferences.

References

- Bansal, Ravi, Robert F. Dittmar, and Christian Lundblad, 2005, Consumption, dividends, and the cross-section of equity returns, *Journal of Finance* 60, 1639–1672.
- Bansal, Ravi, A. Ronald Gallant, and George Tauchen, 2007, Rational pessimism, Rational Exuberance, and Asset Pricing Models, *Review of Economic Studies* forthcoming.
- Bansal, Ravi, Varoujan Khatchatrian, and Amir Yaron, 2005, Interpretable asset markets?, *European Economic Review* 49, 531–560.
- Bansal, Ravi, Dana Kiku, and Amir Yaron, 2007a, A Note on the Economics and Statistics of Predictability: A Long Run Risks Perspective, Working paper, The Wharton School, University of Pennsylvania.
- Bansal, Ravi, Dana Kiku, and Amir Yaron, 2007b, Risks For the Long Run: Estimation and Inference, Working paper, The Wharton School, University of Pennsylvania.
- Bansal, Ravi, and Amir Yaron, 2004, Risks for the long run: A potential resolution of asset pricing puzzles, *Journal of Finance* 59, 1481–1509.
- Benzoni, Luca, Pierre Collin-Dufresne, and Robert S. Goldstein, 2005, Can Standard Preferences Explain the Prices of Out-of-the-Money S&P 500 Put Options, Working paper, University of Minnesota.
- Bhamra, Harjoat, Lars-Alexander Kuehn, and Ilya Strebulaev, 2007, The Levered Equity Risk Premium and Credit Spreads: A Unified Framework, Working paper, Stanford University.
- Bloom, Nick, 2007, The Impact of Uncertainty Shocks, Working paper, Stanford University.
- Bollerslev, Tim, Michael Gibson, and Hao Zhou, 2009, Dynamic Estimation of Volatility Risk Premia and Investor Risk Aversion from Option-Implied and Realized Volatilities, .
- Bollerslev, Tim, George Tauchen, and Hao Zhou, 2009, Expected stock returns and variance risk premia, .
- Boudoukh, Jacob, Matthew Richardson, and Robert Whitelaw, 2009, The Myth of Long-Horizon Predictability, *Review of Financial Studies* forthcoming.

- Britten-Jones, M., and A. Neuberger, 2000, Option Prices, Implied Price Processes, and Stochastic Volatility, *Journal of Finance* 55(2), 839–866.
- Broadie, Mark, Mikhail Chernov, and Michael Johannes, 2007, Model Specification and Risk Premiums: Evidence from Futures Options, *The Journal of Finance* 62, 1453–1490.
- Campbell, John, and Robert Shiller, 1988, Stock Prices, Earnings, and Expected Dividends, *Journal of Finance* 43, 661–676.
- Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay, 1997, *The Econometrics of Financial Markets*. (Princeton University Press Princeton, New Jersey).
- Campbell, John Y., and Samuel B. Thompson, 2007, Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?, *Review of Financial Studies* forthcoming.
- Carr, Peter, and Liuren Wu, 2007, Variance Risk Premia, *Review of Financial Studies* forthcoming.
- Chen, Hui, 2008, Macroeconomic Conditions and the Puzzles of Credit Spreads and Capital Structure, Working paper, MIT.
- Chernov, Mikhail, and Eric Ghysels, 2000, A Study Towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Options Valuation, *The Journal of Financial Economics* 56, 407–458.
- Cochrane, John H., 1999, Portfolio advice for a multifactor world, *Economic Perspectives* XXXIII(3), (Federal Reserve Bank of Chicago).
- Cont, Rama, and Peter Tankov, 2004, *Financial Modeling with Jump Processes*. (Chapman & Hall).
- Demeterfi, K., E. Derman, M. Kamal, and J. Zou, 1999, A Guide to Volatility and Variance Swaps, *Journal of Derivatives* 6, 9–32.
- Duffie, D., J. Pan, and K. J. Singleton, 2000, Transform Analysis and Asset Pricing for Affine Jump-Diffusions, *Econometrica* 68, 1343–1376.

- Epstein, Larry G., and Stanley E. Zin, 1989, Substitution, risk aversion, and the intertemporal behavior of consumption and asset returns: A theoretical framework, *Econometrica* 57, 937–969.
- Eraker, Bjorn, 2004, Do Equity Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices, *The Journal of Finance* 56, 1367–1403.
- Eraker, Bjorn, 2007, Affine General Equilibrium Models, *Management Science* forthcoming.
- Eraker, Bjorn, and Ivan Shaliastovich, 2008, An Equilibrium Guide to Designing Affine Pricing Models, *Mathematical Finance* forthcoming.
- Hansen, Lars, and Ravi Jagannathan, 1991, Implications of Security Market Data for Models of Dynamic Economies, *Journal of Political Economy* 99, 225–262.
- Hansen, Lars, and Thomas Sargent, 2008, Fragile Beliefs and the Price of Model Uncertainty, Working paper, .
- Heston, Steven L., 1993, A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, *Review of Financial Studies* 6, 2, 327–343.
- Hodrick, Robert, 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *Review of Financial Studies* 5, 357–386.
- Jiang, George, and Yisong Tian, 2005, Model-Free Implied Volatility and Its Information Content, *Review of Financial Studies* 18, 1305–1342.
- Lettau, Martin, Sydney Ludvigson, and Jessica Wachter, 2007, The Declining Equity Premium: What Role Does Macroeconomic Risk Play?, *Review of Financial Studies* Forthcoming.
- Liu, Jun, Jun Pan, and Tan Wang, 2005, An Equilibrium Model of Rare-Event Premia and Its Implication for Option Smirks, *The Review of Financial Studies* 18, 131–164.
- Pan, Jun, 2002, The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study, *The Journal of Financial Economics* 63, 3–50.
- Stambaugh, Robert F., 1999, Predictive Regressions, *Journal of Financial Economics* 54, 375–421.

Tauchen, George, 2005, Stochastic Volatility in General Equilibrium, Working paper, Duke University.

Table I
Summary Statistics

	Excess Returns		Variances			
	S&P 500	VWRet	VIX ²	Fut ²	Ind ²	Daily ²
Mean	0.528%	0.526%	33.30	22.17	14.74	20.69
Median	0.957%	1.023%	25.14	14.19	8.99	13.51
Std.-Dev.	4.01%	4.13%	24.13	22.44	15.30	21.95
Skewness	-0.635	-0.836	2.00	2.62	2.78	2.68
Kurtosis	4.217	4.547	8.89	11.10	13.26	11.91
AR(1)	-0.04	0.02	0.79	0.65	0.73	0.62

Table I presents descriptive statistics for excess returns and realized variances. The sample is monthly and covers 1990m1 to 2007m3. VWRet is the value-weighted return on the combined NYSE-AMEX-NASDAQ. VIX² is the square of the VIX index divided by 12, to convert it into a monthly quantity. The value for a particular month is the last observation of that month. Fut² is constructed by summing the squares of the log returns on the S&P 500 futures over 5-minute intervals during a month. Ind² does the same for the log returns on the S&P 500 Index. Daily² sums squared daily returns on the S&P 500 index over a month. All three realized variance measures are multiplied by 10⁴ to convert them into squared percentages and make them comparable to VIX².

Table II
Conditional Volatility

Dept. Variable	Regressors		intercept	β_1	β_2	R^2
	X1	X2				
Daily $^2_{t+1}$ (t-stat)	Daily 2_t	MA(1)	3.70 (2.76)	0.82 (13.19)	-0.35 (-3.37)	0.40
Ind $^2_{t+1}$ (t-stat)	Ind 2_t	VIX 2_t	0.10 (0.11)	0.40 (3.74)	0.26 (4.18)	0.59
Fut $^2_{t+1}$ (t-stat)	Ind 2_t	VIX 2_t	-0.89 (-0.61)	0.29 (2.06)	0.56 (6.19)	0.59

Table II presents estimates from regressions of realized variance measures on lagged predictors. The sample is monthly and covers 1990m1 to 2007m3. Reported t-statistics are Newey-West (HAC) corrected.

Table III
Properties of the Variance Premium

	VP(BTZ)	VP(Ind-forecast)	VP(Daily-MA(1))	VP(Fut-forecast)
Mean	18.56	18.61	12.67	11.27
Median	14.21	15.06	7.97	8.92
Std.-Dev.	15.34	13.55	14.38	7.61
Minimum	-26.05	4.54	-4.02	3.27
Skewness	2.13	2.33	2.45	2.39
Kurtosis	11.86	11.60	12.62	12.03
AR(1)	0.50	0.69	0.54	0.65

Table III presents summary statistics for various measures of the conditional variance premium. The sample is monthly and covers 1990m1 to 2007m3. Each measure is equal to VIX^2 minus a particular quantity. VP(BTZ) subtracts Ind_t^2 , the contemporaneous month's realization of Ind^2 . This measure is used in Bollerslev, Tauchen, and Zhou (2009). VP(Ind-forecast) subtracts the forecast of Ind_{t+1}^2 that comes from the second regression in Table II. VP(Daily-MA(1)) subtracts the forecast of $Daily^2$ that comes from the first regression in Table II. VP(Fut-forecast) subtracts the forecast of Fut_{t+1}^2 that comes from the third regression in Table II.

Table IV
Return Predictability by the Variance Premium

Dependent	Regressors		OLS			Robust Reg.		
	X1	X2	β_1	β_2	$R^2(\%)$	β_1	β_2	$R^2(\%)$
r_{t+1}	VP_t (t-stat)		0.76 (2.18)		1.46	1.12 (2.77)		3.20
r_{t+1}	VP_{t-1} (t-stat)		1.26 (3.90)		4.07	1.21 (2.97)		3.75
r_{t+3}	VP_t (t-stat)		0.86 (3.19)		5.92	0.87 (4.12)		6.09
r_{t+1}	VP_t (t-stat)	$\log(P/E)_t$	1.39 (3.00)	-48.67 (-3.04)	8.30	1.81 (4.33)	-50.52 (-4.36)	10.77
r_{t+1}	VP_{t-1} (t-stat)	$\log(P/E)_t$	2.09 (4.82)	-58.12 (-3.50)	13.43	1.98 (4.68)	-57.30 (-4.85)	12.61

Table IV presents return predictability regressions. The sample is monthly and covers 1990m1 to 2007m3. Reported t-statistics are Newey-West (HAC) corrected. P/E is the price-earnings ratio for the S&P 500. The dependent variable is the total return (annualized and in percent) on the S&P 500 Index over the following one and three months, as indicated. The three month returns series is overlapping. OLS denotes estimates from an ordinary least-squares regression. Robust Reg. denotes estimates from robust regressions utilizing a bisquare weighting function.

Table V
Calibration – Model Parameters

Preferences	δ	γ	ψ			
	0.999	9.5	2.0			
Δc_{t+1}	$E[\Delta c]$	φ_c	w_c			
	0.0016	0.0066	0.5			
x_{t+1}	ρ_x	φ_x	w_x	$l_{1,\sigma}(x)$	μ_x	ν_x
	0.976	$0.032 \times \varphi_c$	1	0.8/12	$3.645 \times \varphi_x$	1
Δd_{t+1}	$E[\Delta d]$	ϕ	φ_d	w_d	Ω_{cd}	
	0.0016	2.5	$5.7 \times \varphi_c$	0.125	0.20	
σ_{t+1}^2	$\tilde{\rho}_\sigma$	φ_σ	$l_{1,\sigma}(\sigma)$	μ_σ	ν_σ	
	0.87	0.35	0.8/12	2.55	1	
$\bar{\sigma}_{t+1}^2$	$\rho_{\bar{\sigma}}$	$\varphi_{\bar{\sigma}}$				
	0.985	0.10				

Table V presents the parameters for the benchmark model with $\xi_x \sim -\Gamma(\nu_x, \frac{\mu_x}{\nu_x}) + \mu_x$.

Table VI
Cash Flows

Statistic	Data		Model		
			5%	50%	95%
<i>Cashflow Dynamics</i>					
$E[\Delta c]$	1.88	(0.32)	0.99	1.89	2.81
$\sigma(\Delta c)$	2.21	(0.52)	1.978	2.29	3.02
$AC1(\Delta c)$	0.43	(0.12)	0.22	0.43	0.63
$E[\Delta d]$	1.54	(1.53)	-1.36	1.84	4.85
$\sigma(\Delta d)$	13.69	(1.91)	9.43	11.02	13.02
$AC1(\Delta d)$	0.14	(0.14)	0.09	0.28	0.44
$corr(\Delta c, \Delta d)$	0.59	(0.11)	0.12	0.32	0.52
$kurt(\Delta c)$ (Q)	4.49	(0.51)	3.24	4.52	8.14

Table VI presents moments of consumption and dividend dynamics from the data and the model of Table V. The data are real, sampled at an annual frequency and cover the period 1930 to 2006, except for the last line, which is quarterly and covers 1947 to 2006. Standard errors are Newey-West with 4 lags. For the model we report percentiles of these statistics based on 1000 model simulations with each statistic calculated using a sample size equal to its data counterpart. In both the model and data, consumption and dividend growth rates are time-averaged.

Table VII
Equity Return and Risk-free Rate

Statistic	Data		Model		
			5%	50%	95%
<i>Returns</i>					
$E[r_m]$	6.23	(1.96)	3.91	6.93	10.00
$E[r_f]$	0.82	(0.35)	0.20	0.92	1.42
$\sigma(r_m)$	19.37	(1.94)	14.84	17.43	21.27
$\sigma(r_f)$	1.89	(0.17)	0.80	1.48	2.66
$E[p - d]$	3.15	(0.07)	2.90	2.99	3.07
$\sigma(p - d)$	0.31	(0.02)	0.13	0.17	0.23
$skew(r_m - r_f)$ (M)	-0.43	(0.54)	-1.03	-0.35	0.15
$kurt(r_m - r_f)$ (M)	9.93	(1.26)	4.23	7.25	15.41
$AC1(r_m - r_f)$ (M)	0.09	(0.06)	-0.10	-0.02	0.07
$kurt(r_m - r_f)$ (A)	3.80	(0.55)	2.85	3.92	8.35

Table VII presents asset pricing moments from the data and the model of Table V. The data are real, sampled at an annual frequency and cover the period 1930 to 2006, except for the lines labeled with “(M)”, which are sampled at a monthly frequency. Standard errors are Newey-West with 4 lags. For the model we report percentiles of these statistics based on 1000 model simulations with each statistic calculated using a sample size equal to its data counterpart.

Table VIII
Variance Premium

Statistic	Data		Model		
			5%	50%	95%
<i>Variance Premium</i>					
$\sigma(\text{var}_t(r_m))$	17.18	(2.21)	5.35	18.38	55.40
$AC1(\text{var}_t(r_m))$	0.81	(0.04)	0.65	0.81	0.91
$AC2(\text{var}_t(r_m))$	0.64	(0.08)	0.42	0.65	0.82
$E[VP]$	11.27	(0.93)	3.77	9.06	21.65
$\sigma(VP)$	7.61	(1.08)	3.74	12.28	36.80
$skew(VP)$	2.39	(0.59)	1.70	3.14	5.33
$kurt(VP)$	12.03	(3.30)	6.22	14.74	37.45
$kurt(\Delta VIX)$	18.83	(5.28)	10.30	26.00	76.67
<i>Return Predictability (VP)</i>					
$\beta(1m)$	0.76	(0.35)	0.00	0.83	2.12
$R^2(1m)$	1.46	(1.52)	0.05	2.70	11.17
$\beta(3m)$	0.86	(0.27)	0.06	0.69	1.79
$R^2(3m)$	5.92	(4.67)	0.20	5.89	24.89
<i>Return Predictability (VP, p-d)</i>					
$\beta_1(1m)$	0.86	(0.40)	-0.35	0.76	2.13
$\beta_2(1m)$	-37.73	(21.30)	-110.65	-20.56	35.28
$R^2(1m)$	4.15	(3.31)	0.38	3.53	11.86
$\beta_1(3m)$	0.95	(0.28)	-0.30	0.60	1.75
$\beta_2(3m)$	-34.40	(17.76)	-100.33	-20.45	30.01
$R^2(3m)$	12.82	(7.51)	0.85	8.20	26.35

Table VIII presents moments pertaining to the variance premium from the data and for the model of Table V. The top panel presents moments for $\text{var}_t(r_m)$ and VP_t (constructed in section 2). The middle and bottom panel present results from predictive regressions of excess market returns on VP (middle panel) and VP and $p-d$ (bottom panel) for horizons of 1 and 3 months. For these regressions the excess returns are expressed as annualized percentages. In all panels the data are sampled at a monthly frequency and cover the period 1990.1-2007.3. For the model we report percentiles of these statistics based on 1000 model simulations with each statistic calculated using a sample size equal to its data counterpart. Standard errors are Newey-West with 4 lags.

Table IX
Long-Horizon Predictability

Statistic	Data		Model		
			5%	50%	95%
<i>Consumption Predictability</i>					
$R^2(1y)$	18.73	(10.75)	0.249	7.72	29.72
$R^2(3y)$	8.19	(7.47)	0.19	8.94	33.73
$R^2(5y)$	4.58	(6.04)	0.11	7.08	31.59
<i>Return Predictability: OLS</i>					
$R^2(1y)$	5.51	(5.18)	0.12	4.57	15.20
$R^2(3y)$	19.30	(10.66)	0.11	7.27	25.18
$R^2(5y)$	36.85	(13.70)	0.11	7.94	28.75
<i>Return Predictability: VAR</i>					
$R^2(1y)$	5.00	(4.30)	0.45	5.99	17.58
$R^2(3y)$	11.27	(8.92)	0.69	9.24	28.56
$R^2(5y)$	14.15	(10.59)	0.67	8.72	29.91

Table IX presents R-squared from predictive regressions of consumption growth (top panel) and excess market returns on the log price-dividend ratio for the data and the model in Table V. Horizons of 1,3 and 5 years are reported. The middle panel reports R-squared from OLS regressions. The bottom panel reports R-squared implied by a monthly VAR of excess returns and the log price-dividend ratio. The sampling frequency used in the regressions reflects data availability: the consumption regression uses non-overlapping annual data, the OLS return regressions use series sampled at the monthly frequency (overlapping), and the VAR is estimated on monthly data. The period covered is 1930 to 2006. For the model we report percentiles for the statistics based on 1000 model simulations with each statistic calculated using a sample size equal to its data counterpart. Standard errors are Newey-West with 4 lags.

Table X
Predictability of Volatility

Statistic	Data		Model		
			5%	50%	95%
<i>Consumption Volatility</i>					
$\beta(1y)$	-0.93	(0.39)	-1.86	-0.65	0.79
$R^2(1y)$	6.75	(4.93)	0.01	1.22	7.99
$\beta(5y)$	-0.56	(0.31)	-1.12	-0.37	0.37
$R^2(5y)$	10.16	(10.03)	0.03	3.31	19.82
<i>Return Volatility</i>					
$\beta(1y)$	-0.09	(0.049)	-0.22	-0.09	-0.00
$R^2(1y)$	7.77	(6.67)	0.13	4.55	18.79
$\beta(5y)$	-0.02	(0.038)	-0.16	-0.05	0.03
$R^2(5y)$	0.95	(3.32)	0.03	3.67	23.24

Table X presents results from predictive regressions of consumption volatility and excess market return volatility on the log price-dividend ratio for the data and the model in Table V. Horizons of 1 and 5 years are reported. Consumption volatility is calculated as the log of the sum of absolute residuals from an AR(1) model of consumption growth. The regression is sampled at an annual frequency and uses the end-of-year price-dividend ratios. Excess market return volatility is calculated as the standard deviation of monthly excess market returns over 1, 3, and 5 year horizons, multiplied by $\sqrt{12}$ for annualization. The estimates are rolling and the regression is sampled at the monthly frequency. The period covered is 1930 to 2006. For the model we report percentiles for the statistics based on 1000 model simulations with each statistic calculated using a sample size equal to its data counterpart. Standard errors are Newey-West with $(\text{horizon} + 1)$ lags where horizon is in units of the sampling frequency.

Table XI
Cash Flows

Statistic	Data		Model		
			5%	50%	95%
<i>Cashflow Dynamics</i>					
$E[\Delta c]$	1.88	(0.32)	1.08	1.94	2.85
$\sigma(\Delta c)$	2.21	(0.52)	1.80	2.28	2.99
$AC1(\Delta c)$	0.43	(0.12)	0.22	0.44	0.61
$E[\Delta d]$	1.54	(1.53)	-1.06	2.00	5.04
$\sigma(\Delta d)$	13.69	(1.91)	9.42	11.06	13.09
$AC1(\Delta d)$	0.14	(0.14)	0.11	0.29	0.46
$corr(\Delta c, \Delta d)$	0.59	(0.11)	0.11	0.32	0.53
$kurt(\Delta c)$ (Q)	4.49	(0.51)	3.24	4.45	8.19

Table XI replicates the results of Table VI for the version of the reference model with $\xi_x \sim \mathcal{N}(0, \sigma_x^2)$. The calibration parameters are in Table V except $\sigma_x = \mu_x$ (μ_x and ν_x are eliminated) and $\gamma = 10$.

Table XII
Equity Return and Risk-free Rate

Statistic	Data		Model		
			5%	50%	95%
<i>Returns</i>					
$E[r_m]$	6.23	(1.96)	4.30	7.21	10.17
$E[r_f]$	0.82	(0.35)	0.33	0.96	1.44
$\sigma(r_m)$	19.37	(1.94)	14.92	17.42	21.50
$\sigma(r_f)$	1.89	(0.17)	0.82	1.42	2.55
$E[p - d]$	3.15	(0.07)	2.88	2.97	3.05
$\sigma(p - d)$	0.31	(0.02)	0.13	0.17	0.23
$skew(r_m - r_f)$ (M)	-0.43	(0.54)	-1.03	-0.23	0.26
$kurt(r_m - r_f)$ (M)	9.93	(1.26)	4.26	7.18	15.89
$AC1(r_m - r_f)$ (M)	0.09	(0.06)	-0.09	-0.02	0.06
$kurt(r_m - r_f)$ (A)	3.80	(0.55)	2.85	3.92	8.20

Table XII replicates the results of Table VII for the version of the reference model with $\xi_x \sim \mathcal{N}(0, \sigma_x^2)$. The calibration parameters are as in Table V except $\sigma_x = \mu_x$ (μ_x and ν_x are eliminated) and $\gamma = 10$.

Table XIII
Variance Premium

Statistic	Data		Model		
			5%	50%	95%
<i>Variance Premium</i>					
$\sigma(\text{var}_t(r_m))$	17.18	(2.21)	5.05	18.15	50.59
$AC1(\text{var}_t(r_m))$	0.81	(0.04)	0.66	0.80	0.91
$AC2(\text{var}_t(r_m))$	0.64	(0.08)	0.42	0.64	0.82
$E[VP]$	11.27	(0.93)	2.98	7.57	17.61
$\sigma(VP)$	7.61	(1.08)	2.96	10.35	28.68
$skew(VP)$	2.39	(0.59)	1.72	3.28	5.34
$kurt(VP)$	12.03	(3.30)	6.35	15.75	37.75
$kurt(\Delta VIX)$	18.83	(5.28)	10.29	26.51	77.89
<i>Return Predictability (vp)</i>					
$\beta(1m)$	0.76	(0.35)	-0.04	0.992	2.63
$R^2(1m)$	1.46	(1.52)	0.06	2.75	10.73
$\beta(3m)$	0.86	(0.27)	-0.03	0.82	2.13
$R^2(3m)$	5.92	(4.67)	0.15	6.27	23.77
<i>Return Predictability (vp, p-d)</i>					
$\beta_1(1m)$	0.86	(0.40)	-0.35	0.87	2.68
$\beta_2(1m)$	-37.73	(21.30)	-110.37	-24.45	37.25
$R^2(1m)$	4.15	(3.31)	0.32	3.72	11.56
$\beta_1(3m)$	0.95	(0.28)	-0.35	0.72	2.09
$\beta_2(3m)$	-34.40	(17.76)	-100.67	-24.54	30.62
$R^2(3m)$	12.82	(7.51)	0.94	8.47	25.38

Table XIII replicates the results of Table VIII for the version of the reference model with $\xi_x \sim \mathcal{N}(0, \sigma_x^2)$. The calibration parameters are in Table V except $\sigma_x = \mu_x$ (μ_x and ν_x are eliminated) and $\gamma = 10$.

Table XIV
Comparative Statics Results

Statistic	Data		Model 1-A			Model 1-B			Model 1-C		
			5%	50%	95%	5%	50%	95%	5%	50%	95%
<i>Cashflow Dynamics</i>											
$E[\Delta c]$	1.88	(0.32)	1.06	1.95	2.83	1.18	1.90	2.64	1.28	1.92	2.64
$\sigma(\Delta c)$	2.21	(0.52)	1.87	2.31	2.84	1.67	2.11	2.74	1.74	2.12	2.59
$AC1(\Delta c)$	0.43	(0.12)	0.23	0.43	0.62	0.14	0.36	0.54	0.16	0.36	0.54
$E[\Delta d]$	1.54	(1.53)	-1.25	2.03	4.96	-1.00	1.95	4.72	-0.87	1.99	4.68
$\sigma(\Delta d)$	13.69	(1.91)	9.62	11.14	12.82	9.24	10.90	12.54	9.32	10.87	12.57
$AC1(\Delta d)$	0.14	(0.14)	0.10	0.29	0.45	0.08	0.26	0.43	0.08	0.27	0.42
$corr(\Delta c, \Delta d)$	0.59	(0.11)	0.10	0.33	0.52	0.05	0.27	0.45	0.06	0.27	0.46
<i>Returns</i>											
$E[r_m]$	6.23	(1.96)	1.40	4.63	7.53	0.24	3.15	5.95	0.02	2.85	5.60
$E[r_f]$	0.82	(0.35)	0.75	1.23	1.65	1.14	1.52	1.82	1.25	1.57	1.84
$\sigma(r_m)$	19.37	(1.94)	13.99	15.84	17.78	12.85	14.67	16.97	12.81	14.53	16.46
$\sigma(r_f)$	1.89	(0.17)	0.75	1.23	1.65	0.47	0.79	1.36	0.42	0.58	0.78
$E[p - d]$	3.15	(0.07)	3.58	3.65	3.93	4.38	4.44	4.49	4.65	4.70	4.75
$\sigma(p - d)$	0.31	(0.02)	0.11	0.14	0.19	0.09	0.12	0.15	0.09	0.11	0.15
$skew(r_m - r_f)$ (M)	-0.43	(0.54)	-0.74	-0.161	0.10	-0.23	0.07	0.40	-0.12	0.02	0.17
$kurt(r_m - r_f)$ (M)	9.93	(1.26)	3.11	3.96	8.19	3.07	3.81	5.94	2.85	3.13	3.48
$AC1(r_m - r_f)$ (M)	0.09	(0.06)	-0.06	0.00	0.05	-0.07	0.00	0.06	-0.06	0.00	0.05
<i>Variance Premium</i>											
$\sigma(\text{var}_t(r_m))$	17.18	(2.21)	3.17	5.15	7.90	2.07	7.39	20.24	2.17	3.53	5.50
$AC1(\text{var}_t(r_m))$	0.81	(0.04)	0.79	0.87	0.93	0.65	0.80	0.91	0.78	0.87	0.93
$AC2(\text{var}_t(r_m))$	0.64	(0.08)	0.62	0.76	0.86	0.42	0.64	0.81	0.62	0.76	0.87
$E[VP]$	11.27	(0.93)	1.44	2.84	4.77	0.13	0.34	0.81	0.01	0.02	0.03
$\sigma(VP)$	7.61	(1.08)	1.17	1.85	2.84	0.14	0.49	1.34	0.01	0.01	0.02
$skew(VP)$	2.39	(0.59)	0.36	0.93	1.81	1.73	3.25	5.52	0.44	0.96	1.79
$kurt(VP)$	12.03	(3.30)	2.35	3.50	6.70	6.29	15.61	39.34	2.38	3.57	6.62
$\beta(1)$	0.76	(0.35)	-1.83	1.64	5.60	-12.12	7.01	30.87	-395.40	95.03	672.82
$R^2(1)$	1.46	(1.52)	0.01	0.44	3.48	0.00	0.68	6.42	0.00	0.33	2.42
$\beta(3)$	0.86	(0.27)	-1.89	1.59	5.63	-12.25	5.91	27.02	-397.71	88.91	632.68
$R^2(3)$	5.92	(4.67)	0.01	1.24	9.78	0.02	1.60	13.71	0.01	0.79	6.69

Table XIV presents a three part comparative statics exercise for the model in Table V. Each panel alters the model in Table V by shutting off a Poisson jump process. Model 1-A sets $l_{1,\sigma} = 0$ to shut off the Poisson component of σ_t^2 . Model 1-B sets $l_{1,x} = 0$ to shut off the Poisson component of x_t . Model 1-C shuts off both Poisson components: $l_{1,\sigma} = l_{1,x} = 0$.