What's Vol Got to Do With It*

Itamar Drechsler†
Amir Yaron‡

First Draft: July 2007
Current Draft: June 2008

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 Wharton, the CREATES workshop 'New Hope for the C-CAPM?', Imperial College, 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, and our discussants George Tauchen and Luca Benzoni. The authors gratefully acknowledge the financial support of the Rodney White Center at the Wharton School.

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

1 Introduction

That 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 and Zhou (2007). 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

[1]See 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.

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

[2]Bansal, 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 decline in economic activity.

[3]Tauchen (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.

(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. Some recent papers linking prices of derivatives with recursive preferences and/or long-run risks fundamentals include Bansal, Gallant, and Tauchen (2007), Bhamra, Kuhn, and Strebulaev (2007), Chen (2008), Benzoni, Collin-Dufresne, and Goldstein (2005), Eraker and Shaliastovich (2008), Liu, Pan, and Wang (2005), and Tauchen (2005).

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 and Zhou (2007) 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 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. 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 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 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 and Zhou (2007) 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. 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^2s 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 used by Bollerslev and Zhou (2007). They calculate the variance premium by subtracting from the squared VIX the previous month's realized variance. It is apparent from the table that the mean of the variance premium is somewhat smaller 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 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. 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 and Zhou (2007). For example, in conjunction with the price-earnings ratio, the in-sample R^2 increases to as much as 12.4%.[4] 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

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

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.

A natural question that arises is whether such R^2s 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.[5] In our sample, s_0 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^2s 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^2s 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^2s 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.

[5] This 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.

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

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

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,

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

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,

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

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 the $\ln R_{c,t+1}$ 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. The general framework 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 hit by both Gaussian and Poisson-driven jump shocks:

$$Y_{t+1} = \mu + FY_t + G_t z_{t+1} + J_{t+1} \tag{4}$$

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 $\exp\big(\Psi_{t,k}(u_k)\big)$, 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 it 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 $\exp\big(\sum_k \lambda_{t,k}(\psi_k(u_k) - 1)\big)$, or more compactly, $\exp\big(\lambda_t'(\psi(u) - 1)\big)$.

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:

$$Y_{t+1} = \begin{pmatrix} \Delta c_{t+1} \\ x_{t+1} \\ \sigma^2_{t+1} \\ \Delta d_{t+1} \end{pmatrix} \qquad F = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & \rho_x & 0 & 0 \\ 0 & 0 & \rho_\sigma & 0 \\ 0 & \phi & 0 & 0 \end{pmatrix}$$

The vector of Gaussian shocks is $z_{t+1} = (z_{c,t+1}, z_{x,t+1}, z_{\sigma,t+1}, z_{d,t+1}) \sim \mathcal{N}(0, \mathcal{I})$ and $J_{t+1} = (0, J_{x,t+1}, J_{\sigma,t+1}, 0)$ is the jump vector. The matrix G_t solves $G_t G_t' = h + H_\sigma \sigma_t^2$, so that the conditional variance-covariance matrix of the gaussian shocks is driven by the variable σ_t^2. Finally, 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 drives variation in the intensities of the jumps.[6] Since σ_t^2 is positive valued, positivity of the jump intensities is implied.

This generalized LRR specification is quite flexible and nests a number of related models. In particular, it nests the original Bansal and Yaron (2004) long-run risks model. 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. As mentioned above, σ_t^2 controls variation in the volatility of Gaussian shocks and jump intensities. 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 $h = \mathrm{diag}\big([0, 0, \varphi_\sigma, 0]\big)$ and $H_\sigma = \mathrm{diag}\big([\varphi_c, \varphi_x, 0, \varphi_d]\big)$. In the Bansal and Yaron (2004) specification, the volatility of σ_t^2 shocks is constant. Tauchen (2005) makes the volatility of σ_t^2 shocks stochastic via a square-root specification. To get this type of specification, set $H_\sigma = \mathrm{diag}\big([\varphi_c, \varphi_x, \varphi_\sigma, \varphi_d]\big)$ (and $h = 0$). Finally, as the specification above shows, we will consider jumps in both σ_t^2 and x_t, but not in the immediate innovations to Δc_{t+1} and Δd_{t+1}. As will be discussed below, these 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.

[6] Here $l_{1,\sigma}$ is the column multiplying σ_t^2 in the expression $l_1 Y_t$, which means it is just the third 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 unconditonal mean of v_t:

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

This approach is also taken by Bansal and Yaron (2004), Eraker and Shaliastovich (2008), and Bansal, Kiku, and Yaron (2007). 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 $A = (A_c, A_x, A_\sigma, A_d)'$ 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 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 satisfy the following system of equations:

$$0 = \theta \ln \delta + \theta\kappa_0 + \theta(\kappa_1 - 1)A_0 + \mathbf{f}\left(\theta(1 - \frac{1}{\psi})e_c + \theta\kappa_1 A\right) \tag{6}$$

$$0 = \mathbf{g}\left(\theta(1 - \frac{1}{\psi})e_c + \theta\kappa_1 A\right) - A\theta \tag{7}$$

where $e_c = (1, 0, 0, 0)'$ and where the functions $\mathbf{f}(u)$ and $\mathbf{g}(u)$ are defined in Appendix A.1. Equation (6) is a scalar and (7) is an $n \times 1$ system of equation which jointly determine A and A_0.

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, we solve for these linearization constants jointly by adding equations for them to the system that is solved numerically. Further details are given in Appendix A.2.

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 $t+1$:

$$\begin{aligned} m_{t+1} &= \theta \ln \delta - \frac{\theta}{\psi}\Delta c_{t+1} + (\theta - 1) r_{c,t+1} \\ &= \theta \ln \delta + (\theta - 1)\kappa_0 + (\theta - 1)(\kappa_1 - 1)A_0 - (\theta - 1)A' Y_t - \Lambda' Y_{t+1} \end{aligned} \tag{8}$$

where $\Lambda = (\gamma e_c + (1-\theta)\kappa_1 A)$. The innovation to the pricing kernel, conditional on the time t information set, has the simple form:

$$m_{t+1} - E_t(m_{t+1}) = -\Lambda'(Y_{t+1} - E_t(Y_{t+1})) = -\Lambda'\left(G_t z_{t+1} + J_{t+1} - E_t(J_{t+1})\right) \tag{9}$$

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, $\Lambda = (\gamma,\ \kappa_1 A_x(1-\theta),\ \kappa_1 A_\sigma(1-\theta),\ 0)'$.

The expression for Λ shows that the signs of the risk prices depend on the signs of the A coefficients and $(1-\theta)$. The signs of the A's themselves depend only on the relation between the preference parameters γ and ψ. Thus, it is the relation between the preference parameters that determines the prices of all risks. 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 were 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 $\gamma > \frac{1}{\psi}$, this parametrization of preferences is identified by Epstein and Zin (1989) as implying a preference for early resolution of uncertainty.

For comparison, consider two cases in which risk aversion is greater than 1 but now $\psi < 1$. In the first case let $\gamma < \frac{1}{\psi}$ (preference for late resolution of uncertainty). In this case, $A_x < 0$ 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.

In the second case, let $\gamma > \frac{1}{\psi}$ (preference for early resolution of uncertainty). In this case, $A_x < 0$ and $A_\sigma > 0$, but now $(1-\theta) < 0$ and hence $\Lambda_x > 0$ and $\Lambda_\sigma < 0$. So for this parameter configuration, the prices of risk have the same sign as for $\gamma > 1$, $\psi > 1$ (our preferences of interest), but the A coefficients have the opposite sign. 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, 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 $v_{m,t+1}$:

$$r_{m,t+1} = \kappa_0 + \kappa_1 v_{m,t+1} - v_{m,t} + \Delta d_{t+1} \tag{10}$$

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

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

where $A_m = (A_{c,m}, A_{x,m}, A_{\sigma,m}, A_{d,m})'$ 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 to a system of equations, analogous to (6) and (7), that 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:

$$r_{m,t+1} = r_0 + (B'_r F - A'_m) Y_t + B'_r G_t z_{t+1} + B'_r J_{t+1} \tag{11}$$

where r_0 is a constant, $B_r = (\kappa_{1,m} A_m + e_d)$, and e_d is $(0, 0, 0, 1)'$ (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:

$$\mathrm{var}_t(r_{m,t+1}) = B'_r G_t G'_t B_r + \sum_i B_r^2(i) \mathrm{var}_t(J_{t+1,i})$$

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 $\Psi_{t,i}(u) = \lambda_{t,i} \psi_i(u)$, so the conditional variance can be rewritten concisely as:

$$\begin{aligned}\mathrm{var}_t(r_{m,t+1}) &= B_r' G_t G_t' B_r + B_r^{2\prime} \Psi_t^{(2)}(0) \\ &= B_r' G_t G_t' B_r + B_r^{2\prime} \mathrm{diag}\left(\psi^{(2)}(0)\right) \lambda_t \end{aligned} \tag{12}$$

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

We can also derive the conditional expected return on the market by taking conditional expectations of (11), obtaining:

$$E_t(r_{m,t+1}) = r_0 + (B_r' F - A_m')Y_t + B_r' E_t\left(J_{t+1}\right)$$

Using the cgf, we have $E_t\left(J_{t+1}\right) = \Psi_t^{(1)}(0)$, which in the compound Poisson case equals $\mathrm{diag}\left(\psi^{(1)}(0)\right) \lambda_t$. Substituting into the expression for the conditional expectation and breaking up λ_t into $l_0 + l_1 Y_t$ leads to the following:

$$E_t(r_{m,t+1}) = \tilde{r}_0 + \left(B_r' \tilde{F} - A_m'\right) Y_t \tag{13}$$

where $\tilde{F} = \left(F + \mathrm{diag}\left[\psi^{(1)}(0)\right] l_1\right)$ and $\tilde{r}_0 = r_0 + \mathrm{diag}\left(\psi^{(1)}(0)\right)$ is a constant.

Equation (13) shows that the conditional expectation of the market return loads on the state vector according to $(B_r'\tilde{F} - A_m')$. Since B_r is a function of A_m, these loadings are effectively determined by the endogenous A_m coefficients that come out of the model solution. Thus, a state variable increases in influence as a driver of time variation in expected returns as it's A_m coefficient increases. The sign of the A_m coefficient determines the direction that expected returns are driven by the state variable.

Consider a state variable with a relatively high loading in $(B_r'\tilde{F} - A_m')$. In other words, the variable is influential in driving expected returns. If expected returns load positively on this state variable, then increases in the state variable will be associated with increases in expected returns. If the variable is subject to large shocks, then expected returns will reflect these shocks in their variation over time. In this paper we argue that a state variable with these properties drives the intensities of jumps in σ_t^2 and x_t. Below we show that this driver of jump intensity is reflected strongly in the variance premium so that the variance premium is a stronger predictor of expected returns. In our reference parametrization σ_t^2 completely determines λ_t, the jump intensity vector. Thus, its endogenously determined influence on returns will reflect its importance as the driver of jump intensities as well as its determination of the volatility of Gaussian shocks.[7]

Equation (13) can be derived more immediately if the model dynamics in (4) are first demeaned. Let $\tilde{J}_{t+1} = J_{t+1} - E_t(J_{t+1})$ denote the (conditionally) demeaned compound Poisson processes. Then the model dynamics can be rewritten in this 'innovations' form by using again the cgf, $E_t(J_{t+1}) = \operatorname{diag}\left(\psi^{(1)}(0)\right)\lambda_t$, and the identity $\lambda_t = l_0 + l_1 Y_t$ to obtain:

$$Y_{t+1} = \tilde{\mu} + \tilde{F} Y_t + G_t z_{t+1} + \tilde{J}_{t+1} \tag{14}$$

where $\tilde{F}$ was just defined above, and $\tilde{\mu} = \mu + \operatorname{diag}\left(\psi^{(1)}(0)\right) l_0$. We use this representation of the model when we consider how the dynamics are altered by changing to the risk neutral probability measure.[8]

3.3.3 Risk Premia

Appendix A.4 uses the Euler equation (3) to derive the risk-free rate, $r_{f,t}$. It equals $r_{f,0} - (\mathbf{g}(-\Lambda) - (\theta-1)A)' Y_t$, where $r_{f,0}$ is a constant given in the appendix. The conditional risk premium is obtained by subtracting $r_{f,t}$ from (13) and equals:

$$E_t(r_{m,t+1} - r_{f,t}) = \tilde{r}_0 - r_{f,0} + \left(B_r' \tilde{F} - A_m' + \mathbf{g}(-\Lambda)' - (\theta-1)A'\right) Y_t \tag{15}$$

For the model parameterizations we consider in the calibration, $r_{f,t}$ has very low variability compared to $r_{m,t}$, as is the case in the data. Thus, variation in $r_{m,t+1} - r_{f,t}$ is essentially identical to that of $r_{m,t+1}$ and the loadings on the state vector for $r_{m,t+1}$ are very close to those of $r_{m,t+1} - r_{f,t}$. In particular, the level of uncertainty and jump intensity, driven by σ_t^2 has a similar influential effect on expected returns and expected risk premia. On the other

[7] An interesting extension of our reference configuration would be to separate 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 should not materially change the underlying mechanisms at work in the model, it will substantially increase the complexity of the model and it's calibration. We believe the reference configuration strikes a good balance between parsimony and achievement of the main objectives of the model.

[8] The innovations form of the dynamics is also more intuitive to use for model calibration and for determining unconditional moments of the state vector.

hand, in our reference configuration the long-run risk variable x_t only effects the risk-free rate and so cancels out of the market risk-premium.

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 the 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:

$$vp_{t,t+1} = E_t^Q[\sum_{i=1}^{n-1} \text{var}^Q_{t+\frac{i-1}{n}}(r_{m,t+\frac{i-1}{n},t+\frac{i}{n}})] - E_t^P[\sum_{i=1}^{n-1} \text{var}^P_{t+\frac{i-1}{n}}(r_{m,t+\frac{i-1}{n},t+\frac{i}{n}})] \tag{16}$$

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 term the quantity $\text{var}_t^Q(r_{m,t+1}) - \text{var}_t^P(r_{m,t+1})$ the "level difference". 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 term it the "drift difference".

Equation (16) 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:

$$vp_{t,t+1} \equiv E_t^Q[\text{var}_t^Q(r_{m,t+1})] - E_t^P[\text{var}_t^P(r_{m,t+1})] \tag{17}$$

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 distribution 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 (9) 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 $f^Q(z_{t+1})$ 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 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

$$z_{t+1} \overset{Q}{\sim} \mathcal{N}(-G'_t\Lambda, I) \tag{18}$$

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:

$$\Psi_{t,k}^{Q}(u_k) = \lambda_{t,k}\psi_k(-\Lambda_k)\left(\frac{\psi_k(u_k - \Lambda_k)}{\psi_k(-\Lambda_k)} - 1\right) \tag{19}$$

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 (19). 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:

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

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 (19) 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 size is amplified or diminished, while the magnitude of the risk price controls the degree of the change.

Given (19), we can now easily compute the moments of J_{t+1} under Q by taking derivatives of the Q measure cgf:

$$E_t^Q(J_{t+1,k}) = \Psi_{t,k}^{Q\;(1)}(0) = \lambda_{t,k}\psi_k^{(1)}(-\Lambda_k) \tag{20}$$

$$\operatorname{var}_t^Q(J_{t+1,k}) = \Psi_{t,k}^{Q\;(2)}(0) = \lambda_{t,k}\psi_k^{(2)}(-\Lambda_k) \tag{21}$$

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

$$Y_{t+1} = \mu + FY_t - G_tG_t'\Lambda + G_t\tilde{z}_{t+1} + J_{t+1}^Q \tag{22}$$

where J_{t+1}^Q denotes the vector of independent compound Poisson processes with cgf given under Q by (19).

4.2 The Variance Premium and the Risk of Jumps

We first focus on the "level difference", $\text{var}_t^Q(r_{m,t+1}) - \text{var}_t^P(r_{m,t+1})$. It follows from (11), (18), and (21) that:

$$\begin{aligned} \text{var}_t^Q(r_{m,t+1}) &= B_r'G_tG_t'B_r + B_r^{2\prime}\Psi_t^{Q(2)}(0) \\ &= B_r'G_tG_t'B_r + B_r^{2\prime}(\text{diag}\left(\psi^{(2)}(-\Lambda)\right)\lambda_t \end{aligned} \tag{23}$$

Subtracting $\text{var}_t^P(r_{m,t+1})$ (equation (12)) from $\text{var}_t^Q(r_{m,t+1})$ then gives the level difference:

$$\text{var}_t^Q(r_{m,t+1}) - \text{var}_t(r_{m,t+1}) = B_r^{2\prime}\text{diag}\left(\psi^{(2)}(-\Lambda) - \psi^{(2)}(0)\right)\lambda_t \tag{24}$$

Some observations are now possible. First, note that the part of conditional variance coming from the Gaussian shocks, $B_r'G_tG_t'B_r$ 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.[9]

Secondly, expression (24) 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

[9]This 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.

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.[10]

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 (24) 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 $E_t\left(\xi_\sigma^2\left[\exp(-\Lambda_\sigma\xi_\sigma) - 1\right]\right)$. 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, his risk-neutral measure puts greater weight on states where there was 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* 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:

$$\psi_x^{(2)}(-\Lambda_x) - \psi_x^{(2)}(0) = \exp\left(\frac{1}{2}\Lambda_x^2\sigma_x^2\right)\Lambda_x^2\sigma_x^4 + \exp\left(\frac{1}{2}\Lambda_x^2\sigma_x^2\right)\sigma_x^2 - \sigma_x^2 \tag{25}$$

[10] The 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.

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. Recall from (15) that the loading of the risk premium on the state vector is given by $\left(B_r'\tilde{F} - A_m' + \mathbf{g}(-\Lambda)' - (\theta - 1)A'\right)$. For our reference configuration, the market risk premium loads only on σ_t^2, so for notational simplicity we denote this loading by $\beta_{r,\sigma}$. When $\gamma > 1$, $\psi > 1$, $\beta_{r,\sigma}$ is positive as a result of $A_{m,\sigma} < 0$.

According to the level difference equation (24), and since in our configuration $\lambda_t = l_{1,\sigma}\sigma_t^2$, we can rewrite $\mathrm{var}_t^Q(r_{m,t+1}) - \mathrm{var}_t(r_{m,t+1})$ as $\beta_{\mathrm{lev},\sigma}\sigma_t^2$. As discussed above, $\gamma > 1$, $\psi > 1$ implies that the level difference is positive, so $\beta_{\mathrm{lev},\sigma} > 0$.

Now, consider the predictive regression for excess market returns:

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

Substituting in the expressions gives

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

Therefore, $\operatorname{var}_t^Q(r_{m,t+1}) - \operatorname{var}_t(r_{m,t+1})$ predicts excess returns on the market. The predictive coefficient is positive, as in the data. The intuition is as follows. The state variable σ_t^2, which controls the intensity of jumps, 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 high conditional risk premia. 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 variance rises with the jump intensity leading to the positive covariation between the variance premium and risk premia.

4.3 Drift Difference

We now examine the contribution to $vp_{t,t+1}$ from the "drift difference": the difference between the quantities (a) $E_t^Q[\operatorname{var}_{t+1}^Q(r_{m,t+2})] - \operatorname{var}_t^Q(r_{m,t+1})$ and (b) $E_t^P[\operatorname{var}_{t+1}^P(r_{m,t+2})] - \operatorname{var}_t^P(r_{m,t+1})$. This is the difference in the "drift" of the conditional variance between the two measures. 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 $\operatorname{var}_t^Q(r_{m,t+1}) = \operatorname{var}_t^P(r_{m,t+1})$. Hence, the drift difference is simply $E_{t+1}^Q[\operatorname{var}_{t+1}(r_{m,t+2})] - E_{t+1}^P[\operatorname{var}_{t+1}(r_{m,t+2})]$. From (12) we have that, in the pure Gaussian case, $\operatorname{var}_t(r_{m,t+1}) = B_r' G_t G_t' B_r$. In our reference configuration, $G_t G_t' = h + H_\sigma \sigma_t^2$, so that $\operatorname{var}_t(r_{m,t+1}) = B_r' h B_r + B_r' H_\sigma B_r \sigma_t^2$ and the drift difference is just $B_r' H_\sigma B_r \left[E_t^Q(\sigma_{t+1}^2) - E_t^P(\sigma_{t+1}^2)\right]$, 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 (22) we see that for the reference configuration the pure Gaussian case gives:

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

Let σ_t^2 correspond to row i of Y_t (in our reference model $i = 3$). Assume for simplicity, that shocks to σ_t^2 are uncorrelated with the other shocks. Then in the i-th row of $h + H_\sigma \sigma_t^2$ only the diagonal element is non-zero and the drift difference is simply:

$$-B_r' H_\sigma B_r \left[h(i,i) + H_\sigma(i,i)\sigma_t^2\right] \Lambda_\sigma$$

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.

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.

We wish to note that 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 With Jumps

To conclude, we discuss the drift difference when the Poisson jumps are included. Since we have already derived the drift difference for the Gaussian-related part of $\mathrm{var}_t(r_{m,t+1})$, we now consider only the Poisson-related part. From (12) and (23) this is $B_r^{2\prime}\mathrm{diag}\left(\psi^{(2)}(*)\right)\lambda_t$ where $* = 0$ under P and $* = -\Lambda$ under Q. Thus, under P the one period drift in this quantity is: $B_r^{2\prime}\mathrm{diag}\left(\psi^{(2)}(0)\right)\left[E_t^P(\lambda_{t+1}) - \lambda_t\right]$, while under Q it is: $B_r^{2\prime}\mathrm{diag}\left(\psi^{(2)}(-\Lambda)\right)\left[E_t^Q(\lambda_{t+1}) - \lambda_t\right]$. 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 $E_t^Q(\lambda_t)$ and $E_t^P(\lambda_t)$ and it parallels the discussion above of the Gaussian case, e.g. $E_t^Q(\lambda_t) > E_t^P(\lambda_t)$ when $\gamma > 1$, $\psi > 1$. Finally, we note that this Poisson part of the drift difference is a linear function of the λ_t. Thus, it is the second component of $vp_{t,t+1}$ that is driven by the latent jump intensity.

4.3.3 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. 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 several model calibrations.

5 Calibration Results

5.1 Parametrization

We first discuss in more detail the parametrization of the model. 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. The symmetric distribution is a zero-mean normal distribution: $\xi_x \sim \mathcal{N}(0, \sigma_x^2)$. The asymmetric distribution is a demeaned gamma distribution: $\Gamma(\nu_x, \frac{\mu_x}{\nu_x}) - \mu_x$. Demeaning the jump size prevents σ_t^2 from entering into the equation for the expected change in x_t. Otherwise, it would become a factor in the x_t equation, since it drives the jump intensity and therefore 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. $\xi_x \sim -\Gamma(\nu_x, \frac{\mu_x}{\nu_x}) + \mu_x$.

It is easiest to specify the model parameters using the 'innovations' form of the dynamics specified in equation (14). This is more intuitive than using (4) and also makes it clearer whether a given set of parameters implies stationary dynamics.[11] Therefore, we specify $\tilde{F}$ in (14) rather than F. For our specifications, the difference between them is in the equation for σ^2_{t+1}. This is the result of the jumps in σ^2_t, which have a non-zero mean. Since σ^2_t itself drives the intensity of the jumps, $\tilde{F}$ implies that the true autoregressive parameter for σ^2_t is larger than the parameter ρ_σ in (4). We label the true autoregressive parameter $\tilde{\rho}_\sigma$ and write:

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

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, $E[\sigma^2_t] = 1$. 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^2_t (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: $h(i,i) + H_\sigma(i,i)\sigma^2_t = \varphi^2_i(1-w_i)E(\sigma^2_t) + \varphi^2_i w_i \sigma^2_t$ 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^2_t. The parameter w_i is the weighting that controls the conditional shock variance's sensitivity to changes in σ^2_t.

[11] However, for simulation (4) is easier. We map the explicitly specified parameters in (14) into the ones in (4) so they can be used in the model simulations.

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 and let φ be the vector of φ_i and w be the vector of w_i. Then we set:

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

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.[12] 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.

5.2 Results

In calibrating the model we use the following guidelines. We assume a monthly decision interval. We would like to find a specification for the long run, volatility, and jump shocks such that (i) once time-averaged to annual data 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's variance premium generates statistics as well as return projection results that are consistent with the data.

In Table V we provide the parameter specification for the model economy described above with jump shocks that are normally distributed. Table VI provides the data and the 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 500 simulations each with the same finite sample length as its data counterpart. For the consumption and dividend dynamics we utilize the longest sample available; hence, the simulations are based on 924 monthly

[12]To 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 $w_c = 0.5$, $w_d = 0.25$, for which this quantity equals $0.97 \times \Omega_{cd}$.

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. 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 500 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 of the data statistics.

The top panel in Table VI shows that the model captures quite well several 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. It should be noted that the model parameters in Table V are generally close to those in Bansal and Yaron (2004) and these results indicate that the jump components do not effect the annual cashflow dynamics in a significant manner. Table VI also presents the model based asset pricing implications. The middle panel, labeled *returns*, pertains to annual data on the market, risk free rate and price-divided ratio. As discussed earlier, the corresponding model statistics are 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 panel in Table VI provides several statistics pertaining to the variance premium, all of which are given at the monthly frequency. The median of the model generated mean variance premium is somewhat smaller than its data counterpart. However, the model's 90% confidence interval easily includes the data point estimate of the mean variance premium. The rest of the model statistics are amazingly in line with the data estimates. In particular, the model's median for the volatility, skewness and kurtosis of the variance premium are essentially the same as their data counterpart. Further, the volatility and first two autocorrelations of the conditional volatility of the market return are quite close to their data estimates. While there is no single parameter that uniquely governs these moments, we show below that the jump properties clearly affect these moments in a sizeable manner.

At the outset of this paper, we highlighted the ability of the variance premium to predict future returns. The model is able to replicate this feature of 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 the short horizons. The model median R^2 for the one-month ahead projection is about 2% and the 90% finite sample distribution of R^2 clearly includes the 1.5% R^2 from the data. Furthermore the 5.9% R^2 for the 3-month ahead projection is quite close to the model's median R^2 estimate. 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. It should be noted that, although we do not formally estimate the model, the number of reported statistics exceeds the number of parameters in the model so that capturing the long list of moments in Table VI is by no means an obvious outcome. Finally, it is important to recognize that the preference parameters used here (e.g., risk aversion of 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 (2007)). This provides some cross-validation of these type of preferences.

Table VII provides the parameter configuration for a model in which the jump sizes are drawn from a gamma distribution. As discussed above this configuration allows us to consider more non-symmetric jumps. Table VIII provides the corresponding output from this model. Again, one can easily observe that the model produces cashflow statistics that are consistent with their data counterparts. The market return and equity premium are now slightly larger and match their data counterpart. In essence, it is quite difficult to distinguish this configuration from the one given in Table VI purely along these cashflow and return dimensions. The main fit improvement of this model relative to the one with normal shocks is in matching the skewness and kurtosis of the variance premium. While the median estimate were slightly too large relative to their data counterpart in the case of the normal distribution, the model with gamma shocks gets these dimensions more precisely. Furthermore, the skewed shock structure emanating from this specification leads to larger R^2s in variance premium's ability to predict future returns.

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 VI, the corresponding level difference component has a median size and standard deviation that are approximately 75 and 82 percent of the total variance premium's size and standard deviation, respectively. For Table VIII, the corresponding percentages are 78 and 85. 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.

In Table IX 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. 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 3.6%-4.6% (from 6.5% in the case of Table VI). 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 two panels and essentially zero in the last. Moreover, when the jump in σ_t^2 is shut in the first and third panels the volatility of the conditional variance of the market return is much 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^2s that are far below their data counterparts and the predictive regression coefficients are very unstable. This shows that the variance premium moments convey much information on the time variation in conditional cashflow moments.

Throughout this paper we have been motivated by the connection between the variance premium, the risk of influential shocks to the economy, and return predictability. Predictability in our model depends to a large extent on the risk of a large shock to either uncertainty/volatility or x_{t+1}, the small, persistent component in cash flow growth. These large shocks are quite infrequent compared to the small Gaussian shocks that occur on a normal basis. As evidenced by the calibration results, they have only a small effect at the annual level on cash flows and return volatility but are important for variation in conditional asset pricing moments, particularly at horizons of a few months. These asset pricing moments clearly show the effect of (ex-ante) risk – the possibility of an influential (large or important) shock – though actual (ex-post) realizations may materialize much less frequently. In this sense, these risks incorporate a 'rare-events' element to them, though it is important to note that the jump intensities in our calibrations imply a jump realization, on average, every year or two. In the calibration with normal jump sizes, the large shock realizations are symmetrically good and bad. The curvature of the utility function means that the benefit of good shocks is outweighed by the loss due to the bad ones. In the asymmetric case of demeaned gamma jump sizes, the skewness of the jump sizes implies that the shocks are likely to be small and positive, but infrequently are larger, negative shocks. For both variants of the model, there may be extended periods where large, negative shocks are not realized, though the risk of them is real and varies through time. This is reflected in the finite-sample R-square statistics, which correspond to a sample of the same length as the corresponding data. Note that in Tables VI and VIII, the right tails of the R-square 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-squares implied by the model pricing kernel are close to these median values.

6 Conclusion

This paper shows that the variance premium is useful for measuring agents' perceptions of uncertainty and the risk of influential shocks to the 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. 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 the predictive regression for market returns.

A possible direction for generating interesting transient dynamics like the ones documented here is by generalizing preferences to include features of ambiguity aversion and a desire for robustness. As Hansen and Sargent (2006) demonstrate, a desire for robustness can lead to interesting time-varying misspecification risk premia components. The derivative related features of the data could be a fruitful ground for assessing the role these additional dimensions may provide in enhancing the model's ability to confront the data.

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 (2007) 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 papers 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

We use the Euler equation 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:

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

Now substitute in the conjecture $v_t = A_0 + A' Y_t$ to get the equation in 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:

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

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

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

$$\begin{aligned} E \left[\exp(u' Y_{t+1} \,|\, Y_t) \right] &= \exp(\mathbf{f}(u) + \mathbf{g}(u)' Y_t) && \text{(A.1.2)} \\ \mathbf{f}(u) &= \mu' u + \frac{1}{2} u' h u + l_0' \left(\psi(u) - 1 \right) && \text{(A.1.3)} \\ \mathbf{g}(u) &= F' u + \frac{1}{2} [u' H_i u]_{i \in \{1 \ldots n\}} + l_1' \left(\psi(u) - 1 \right) && \text{(A.1.4)} \end{aligned}$$

and $[u' H_i u]_{i \in \{1 \ldots n\}}$ 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 expression into three terms:

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

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

$$
E_t\left(\exp(u'G_t Z_{t+1})\right) = \exp(\frac{1}{2}u'G_t G_t' u) = \exp\left(\frac{1}{2}u'hu + \frac{1}{2}\sum_i u'H_i u'Y_t(i)\right)
$$

$$
E_t\left(\exp(u'J_{t+1})\right) = \exp\left(\lambda_t'(\psi(u)-1)\right) = \exp\left(l_0'(\psi(u)-1) + (l_1 Y_t)'(\psi(u)-1)\right)
$$

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:

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

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:

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

$$
0 = \mathbf{g}\left(\theta(1-\frac{1}{\psi})e_c + \theta\kappa_1 A\right) - A\theta \quad \text{(A.1.7)}
$$

A.2 Numerical Solution

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

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

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:

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

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

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

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 (8) (2) substitute in (10) 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 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:

$$r_{f,t} = r_{f,0} - (\mathbf{g}(-\Lambda) - (\theta - 1)A)' Y_t \quad \text{(A.4.1)}$$

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 helps to clarify a few points made in the main text about the drift difference and predictability by the variance premium. The model is a simplified version of the reference model in the main text, but with the addition of a state variable. The simplification relative to the reference model is that the Poisson shocks are shut off (i.e. $\lambda_t \equiv 0$) and the Gaussian shocks are uncorrelated. 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 can written as:

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

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

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 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 (12). Writing out all the terms in expanded form gives:

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

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 is equal to the drift difference, which is nonzero as σ_t^2 and q_t have different drifts under P and Q.

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

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

It then easily follows that:

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

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:

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

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 $\gamma > 1$, $\psi > 1$.

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, 2007, 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 Kuhn, and Ilya Strebulaev, 2007, The Levered Equity Risk Premium and Credit Spreads: A Unified Framework, Working paper, Stanford.

Bloom, Nick, 2007, The Impact of Uncertainty Shocks, Working paper, Stanford University.

Bollerslev, Tim, and Hao Zhou, 2007, Expected stock returns and variance risk premia, Working paper, Finance and Economics Discussion Series 2007-11, Board of Governors of the Federal Reserve System (U.S.).

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, 2006, 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.

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. VIX^2 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. Fut^2 is constructed by summing the squares of the log returns on the S&P 500 futures over 5-minute intervals during a month. Ind^2 does the same for the log returns on the S&P 500 Index. $Daily^2$ sums squared daily returns on the S&P 500 index over a month. All three realized variance measures are multiplied by 10^4 to convert them into squared percentages and make them comparable to VIX^2.

Table II

Conditional Volatility

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

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(BZ)	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(BZ) subtracts Ind_t^2, the contemporaneous month's realization of Ind^2. This measure is used in Bollerslev and Zhou (2007). 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		0.76		1.46	1.12		3.20
	(t-stat)		(2.18)			(2.77)		
r_{t+1}	VP_{t-1}		1.26		4.07	1.21		3.75
	(t-stat)		(3.90)			(2.97)		
r_{t+3}	VP_t		0.86		5.92	0.87		6.09
	(t-stat)		(3.19)			(4.12)		
r_{t+1}	VP_t	$\log{(\mathrm{P/E})_t}$	1.39	-48.67	8.30	1.81	-50.52	10.77
	(t-stat)		(3.00)	(-3.04)		(4.33)	(-4.36)	
r_{t+1}	VP_{t-1}	$\log{(\mathrm{P/E})_t}$	2.09	-58.12	13.43	1.98	-57.30	12.61
	(t-stat)		(4.82)	(-3.50)		(4.68)	(-4.85)	

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 Parameter Configuration

Preferences	δ	γ	ψ		
	0.999	10	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	$l_{1,\sigma}(x)$	σ_x
	0.975	$0.042 \times \varphi_c$	0.43	0.75/12	$2.5 \times \varphi_x$
Δd_{t+1}	$E[\Delta d]$	ϕ	φ_d	w_d	Ω_{cd}
	0.0016	3	$6.7 \times \varphi_c$	0.25	0.20
σ^2_{t+1}	$\tilde{\rho}_\sigma$	φ_σ	$l_{1,\sigma}(\sigma)$	μ_σ	ν_σ
	0.8975	0.30	0.75/12	2.5	1.0

Table V presents the parameters for the version of the reference model with $\xi_x \sim \mathcal{N}(0, \sigma_x^2)$.

Table VI
Model Calibration Results

Statistic	Data		Model 5%	Model 50%	Model 95%
Cashflow Dynamics					
$E[\Delta c]$	1.88	(0.32)	0.90	1.86	2.88
$\sigma(\Delta c)$	2.21	(0.52)	1.94	2.34	2.95
$AC1(\Delta c)$	0.43	(0.12)	0.26	0.46	0.64
$E[\Delta d]$	1.54	(1.53)	-1.58	1.74	5.65
$\sigma(\Delta d)$	13.69	(1.91)	11.04	13.23	15.72
$AC1(\Delta d)$	0.14	(0.14)	0.13	0.31	0.50
$corr(\Delta c, \Delta d)$	0.59	(0.11)	0.11	0.38	0.56
Returns					
$E[r_m]$	6.23	(1.96)	3.29	6.49	10.22
$E[r_f]$	0.82	(0.35)	0.52	1.08	1.53
$\sigma(r_m)$	19.37	(1.94)	16.30	19.42	23.90
$\sigma(r_f)$	1.89	(0.17)	0.80	1.22	2.38
$E[p-d]$	3.15	(0.07)	2.98	3.05	3.13
$\sigma(p-d)$	0.31	(0.02)	0.13	0.17	0.22
$skew(r_m - r_f)$ (M)	-0.43	(0.54)	-0.99	-0.21	0.30
$kurt(r_m - r_f)$ (M)	9.93	(1.26)	4.08	7.12	14.70
$AC1(r_m - r_f)$ (M)	0.09	(0.06)	-0.09	-0.01	0.06
Variance Premium					
$\sigma(\mathrm{var_t}(r_m))$	17.18	(2.21)	6.62	23.46	73.23
$AC1(\mathrm{var_t}(r_m))$	0.81	(0.04)	0.66	0.82	0.92
$AC2(\mathrm{var_t}(r_m))$	0.64	(0.08)	0.45	0.67	0.85
$E[VP]$	11.27	(0.93)	4.02	7.57	17.63
$\sigma(VP)$	7.61	(1.08)	3.00	10.65	33.23
$skew(VP)$	2.39	(0.59)	1.84	3.36	5.36
$kurt(VP)$	12.03	(3.30)	6.52	15.74	38.00
$\beta(1)$	0.76	(0.35)	-0.39	0.83	2.63
$R^2(1)$	1.46	(1.52)	0.02	1.94	9.73
$\beta(3)$	0.86	(0.27)	-0.27	0.76	2.09
$R^2(3)$	5.92	(4.67)	0.04	4.21	23.80
$\beta(6)$	0.49	(0.24)	-0.38	0.55	1.68
$R^2(6)$	3.97	(4.74)	0.07	5.66	33.64

Table VI presents (a) consumption and dividend dynamics (b) asset pricing moments (c) moments pertaining to the variance premium. For each statistic the table reports its data and model corresponding values. The data for consumption, dividends, the market return, risk free rate, and price-dividend ratio correspond to the period from 1930 to 2006. The data pertaining to the variance premium is based on monthly data from 1990.1-2007.3. For the model we report finite sample statistics based on 500 simulations each with the corresponding sample size the same as its data counterpart. For the annual data the statistics are based on time-averaged data. The parameters for calibrating the model are given in Table V. Standard errors are calculated using the Newey-West variance-covariance estimator with 4 lags.

Table VII

Calibration – Model Parameter Configuration

Preferences	δ	γ	ψ			
	0.999	10.49	2.0			
Δc_{t+1}	$E[\Delta c]$	φ_c	w_c			
	0.0017	0.0066	0.5			
x_{t+1}	ρ_x	φ_x	w_x	$l_{1,\sigma}(x)$	μ_x	ν_x
	0.975	$0.042 \times \varphi_c$	0.28	0.5/12	$3 \times \varphi_x$	1
Δd_{t+1}	$E[\Delta d]$	ϕ	φ_d	w_d	Ω_{cd}	
	0.0017	3	$6.7 \times \varphi_c$	0.25	0.20	
σ^2_{t+1}	$\tilde{\rho}_\sigma$	φ_σ	$l_{1,\sigma}(\sigma)$	μ_σ	ν_σ	
	0.90	0.40	1.25/12	1.9	1.3	

Table VII presents the parameters for the version of the reference model with $\xi_x \sim -\Gamma(\nu_x, \frac{\mu_x}{\nu_x}) + \mu_x$.

Table VIII

Model Calibration Results

Statistic	**Data**		**Model** 5%	50%	95%
Cashflow Dynamics					
$E[\Delta c]$	1.88	(0.32)	1.00	1.98	2.93
$\sigma(\Delta c)$	2.21	(0.52)	1.94	2.41	3.03
$AC1(\Delta c)$	0.43	(0.12)	0.27	0.47	0.67
$E[\Delta d]$	1.54	(1.53)	-1.57	1.86	5.40
$\sigma(\Delta d)$	13.69	(1.91)	11.09	13.25	15.84
$AC1(\Delta d)$	0.14	(0.14)	0.12	0.30	0.49
$corr(\Delta c, \Delta d)$	0.59	(0.11)	0.13	0.36	0.59
Returns					
$E[r_m]$	6.23	(1.96)	3.55	6.78	10.14
$E[r_f]$	0.82	(0.35)	0.46	0.99	1.49
$\sigma(r_m)$	19.37	(1.94)	16.64	19.32	23.50
$\sigma(r_f)$	1.89	(0.17)	0.84	1.35	2.21
$E[p-d]$	3.15	(0.07)	2.94	3.01	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)	-0.74	-0.13	0.36
$kurt(r_m - r_f)$ (M)	9.93	(1.26)	3.81	6.00	10.89
$AC1(r_m - r_f)$ (M)	0.09	(0.06)	-0.09	-0.02	0.07
Variance Premium					
$\sigma(\mathrm{var_t}(r_m))$	17.18	(2.21)	8.85	22.41	61.88
$AC1(\mathrm{var_t}(r_m))$	0.81	(0.04)	0.65	0.83	0.93
$AC2(\mathrm{var_t}(r_m))$	0.64	(0.08)	0.42	0.68	0.86
$E[VP]$	11.27	(0.93)	3.77	7.05	16.69
$\sigma(VP)$	7.61	(1.08)	4.06	10.28	28.39
$skew(VP)$	2.39	(0.59)	1.90	2.87	4.60
$kurt(VP)$	12.03	(3.30)	6.36	12.29	29.22
$\beta(1)$	0.76	(0.35)	-0.11	0.94	2.59
$R^2(1)$	1.46	(1.52)	0.01	2.17	8.82
$\beta(3)$	0.86	(0.27)	-0.16	0.83	2.14
$R^2(3)$	5.92	(4.67)	0.08	5.07	19.76
$\beta(6)$	0.49	(0.24)	-0.21	0.58	1.60
$R^2(6)$	3.97	(4.74)	0.11	5.74	30.54

Table VIII presents (a) consumption and dividend dynamics (b) asset pricing moments (c) moments pertaining to the variance premium. For each statistic the table reports its data and model corresponding values. The data for consumption, dividends, the market return, risk free rate, and price-dividend ratio correspond to the period from 1930 to 2006. The data pertaining to the variance premium is based on monthly data from 1990.1-2007.3. For the model we report finite sample statistics based on 500 simulations each with the corresponding sample size the same as its data counterpart. For the annual data the statistics are based on time-averaged data. The parameters for calibrating the model are given in Table VII. Standard errors are calculated using the Newey-West variance-covariance estimator with 4 lags.

Table IX
Model Calibration Results

Statistic	Data		Model 1-A 5%	Model 1-A 50%	Model 1-A 95%	Model 1-B 5%	Model 1-B 50%	Model 1-B 95%	Model 1-C 5%	Model 1-C 50%	Model 1-C 95%
Cashflow Dynamics											
$E[\Delta c]$	1.88	(0.32)	1.02	1.93	2.74	1.16	1.82	2.72	1.17	1.87	2.79
$\sigma(\Delta c)$	2.21	(0.52)	2.02	2.41	2.87	1.82	2.46	2.82	1.94	2.27	2.68
$AC1(\Delta c)$	0.43	(0.12)	0.31	0.47	0.63	0.21	0.43	0.59	0.24	0.43	0.60
$E[\Delta d]$	1.54	(1.53)	-1.44	1.86	5.19	-1.47	1.65	4.81	-1.47	1.70	5.04
$\sigma(\Delta d)$	13.69	(1.91)	11.52	13.39	15.36	11.11	12.86	15.07	11.06	13.00	15.37
$AC1(\Delta d)$	0.14	(0.14)	0.11	0.30	0.46	0.09	0.27	0.44	0.13	0.29	0.45
$corr(\Delta c, \Delta d)$	0.59	(0.11)	0.14	0.38	0.54	0.11	0.35	0.52	0.12	0.33	0.54
Returns											
$E[r_m]$	6.23	(1.96)	1.56	4.61	8.03	0.63	3.86	7.06	0.41	3.59	6.92
$E[r_f]$	0.82	(0.35)	0.92	1.33	1.65	0.95	1.37	1.75	1.13	1.44	1.80
$\sigma(r_m)$	19.37	(1.94)	16.59	18.73	21.05	15.52	17.78	20.65	15.70	17.84	19.70
$\sigma(r_f)$	1.89	(0.17)	0.64	0.81	1.05	0.61	0.85	1.41	0.52	0.69	0.90
$E[p-d]$	3.15	(0.07)	3.51	3.59	3.65	3.80	3.86	3.93	3.93	3.99	4.05
$\sigma(p-d)$	0.31	(0.02)	0.14	0.17	0.21	0.12	0.15	0.19	0.13	0.15	0.19
$skew(r_m - r_f)$ (M)	-0.43	(0.54)	-0.19	0.01	0.21	-0.23	0.05	0.41	-0.15	-0.10	0.15
$kurt(r_m - r_f)$ (M)	9.93	(1.26)	3.06	3.47	4.27	3.11	3.93	7.42	2.81	3.11	3.44
$AC1(r_m - r_f)$ (M)	0.09	(0.06)	-0.06	0.00	0.05	-0.07	0.00	0.06	-0.06	0.00	0.06
Variance Premium											
$\sigma(\mathrm{var_t}(r_m))$	17.18	(2.21)	4.64	6.22	9.44	3.40	10.51	32.21	3.45	4.76	7.15
$AC1(\mathrm{var_t}(r_m))$	0.81	(0.04)	0.79	0.87	0.93	0.65	0.81	0.924	0.81	0.87	0.93
$AC2(\mathrm{var_t}(r_m))$	0.64	(0.08)	0.63	0.75	0.86	0.43	0.66	0.84	0.64	0.75	0.87
$E[VP]$	11.27	(0.93)	0.22	0.30	0.40	0.23	0.41	1.10	0.02	0.02	0.03
$\sigma(VP)$	7.61	(1.08)	0.12	0.17	0.25	0.19	0.59	1.81	0.01	0.01	0.02
$skew(VP)$	2.39	(0.59)	0.32	0.88	1.71	1.79	3.26	5.12	0.36	0.82	1.60
$kurt(VP)$	12.03	(3.30)	2.35	3.33	6.72	6.79	15.22	35.98	2.34	3.26	6.39
$\beta(1)$	0.76	(0.35)	-33.43	7.43	60.41	-14.94	4.31	28.52	-478.07	39.43	604.86
$R^2(1)$	1.46	(1.52)	0.00	0.22	2.42	0.01	0.61	5.89	0.01	0.20	1.96
$\beta(3)$	0.86	(0.27)	-32.28	6.19	62.27	-13.83	3.27	21.25	-485.47	39.43	604.86
$R^2(3)$	5.92	(4.67)	0.00	0.50	7.13	0.01	1.39	10.70	0.01	0.53	5.24
$\beta(6)$	0.49	(0.24)	-33.47	4.44	47.04	-10.89	2.59	16.30	-447.01	29.35	522.27
$R^2(6)$	3.97	(4.74)	0.016	1.11	12.62	0.01	1.74	15.83	0.01	1.04	9.11

Table IX presents a three part comparative statics exercise for the model given in Table V. Each panel alters the model in TableV 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: $l_{1,\sigma} = l_{1,x} = 0$.