The Risk Premia Embedded in Index Options*

Torben G. Andersen† Nicola Fusari‡ Viktor Todorov§

March 6, 2014

Abstract
We study the dynamic relation between aggregate stock market risks and risk premia via an exploration of the time series of equity-index option surfaces. The analysis is based on estimating a general parametric asset pricing model for the risk-neutral equity market dynamics using a panel of options on the S&P 500 index, while remaining fully nonparametric about the actual evolution of market risks. We find that the risk-neutral jump intensity, which controls the pricing of left tail risk, cannot be spanned by the market volatility (and its components), so an additional factor is required to account for its dynamics. This tail factor has no incremental predictive power for future equity return volatility or jumps beyond what is captured by the current and past level of volatility. In contrast, the novel factor is critical in predicting the future market excess returns over horizons up to one year, and it explains a large fraction of the future variance risk premium. We contrast our findings with those implied by structural asset pricing models that seek to rationalize the predictive power of option data. Relative to those studies, our findings suggest a wider wedge between the dynamics of equity market risks and the corresponding risk premia with the latter typically displaying a far more persistent reaction following market crises.

Keywords: Option Pricing, Risk Premia, Jumps, Stochastic Volatility, Return Predictability, Risk Aversion, Extreme Events.

JEL classification: C51, C52, G12.

* Andersen gratefully acknowledges support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. Todorov’s work was partially supported by NSF Grant SES-0957330. We are grateful to two anonymous referees and the editor, Ken Singleton, for many helpful suggestions that have led to significant improvements. We also thank Snehil Banerjee, Peter Carr, Anna Cieslak, Bjorn Eraker, Ravi Jaganathan, Bryan Kelly (our discussant at the NBER Meeting), Robert Merton, Sergio Rebelo, Myron Scholes, Ivan Shaliastovich (our discussant in Montreal), Liuren Wu as well as seminar participants at Kellogg School of Management, Northwestern University, Duke University, the "Montreal 2013 Econometrics Conference: Time Series and Financial Econometrics," the 40th Annual Meeting of the Danish Econometric Society, Sandbjerg, Denmark, the European University Institute, Florence, 2013 Workshop on "Measuring and Modeling Financial Risk and High Frequency Data", the 2013 "International Conference in Financial Econometrics" at Shandong University, Jinan, China, the “40 Years after the Black-Scholes-Merton Model” conference at the Stern School of Business, October 2013, the NBER Asset Pricing Meeting at Stanford University, November 2013, Georgia Tech, and Boston University for helpful comments.

†Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208; e-mail: t-andersen@northwestern.edu.
‡The Johns Hopkins University Carey Business School, Baltimore, MD 21202; e-mail: nicola.fusari@jhu.edu.
§Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208; e-mail: v-todorov@northwestern.edu.
1 Introduction

Equity markets are subject to pronounced time-variation in volatility as well as abrupt shifts, or jumps. Moreover, these risk features are related in intricate ways, inducing a complex equity return dynamics. Hence, the markets are incomplete and derivative securities, written on the equity index, are non-redundant assets. This partially rationalizes the rapid expansion in the trading of contracts offering distinct exposures to volatility and jump risks. From an economic perspective, it suggests that derivatives data contain important information regarding the risk and risk pricing of the underlying asset. Indeed, recent evidence, exploiting parametric models, e.g., Broadie et al. (2007), or nonparametric techniques, e.g., Bollerslev and Todorov (2011b), finds the pricing of jump risk, implied by option data, to account for a significant fraction of the equity risk premium.

Standard no-arbitrage and equilibrium-based asset pricing models imply a tight relationship between the dynamics of the options and the underlying asset. This arises from the assumptions concerning the pricing of risk in the no-arbitrage setting and the endogenous pricing kernels implied by the equilibrium models. A prominent example is the illustrative double-jump model of Duffie et al. (2000) in which the return volatility itself follows an affine jump diffusion. In this context, the entire option surface is governed by the evolution of market volatility, i.e., the dynamics of all options is driven by a single latent Markov (volatility) process.

Recent empirical evidence reveals, however, that the dynamics of the option surface is far more complex. For example, the term structure of the volatility index, VIX, shifts over time in a manner that is incompatible with the surface being driven by a single factor. Likewise, Bates (2000) documents that a two-factor stochastic volatility model for the risk-neutral market dynamics provides a significant improvement over a one-factor version. Moreover, Bollerslev and Todorov (2011b) find that even the short-term option dynamics cannot be captured adequately by a single factor as the risk-neutral tails display independent variation relative to market volatility, thus driving a wedge between the dynamics of the option surface and the underlying asset prices.

The objective of the current paper is to characterize the risk premia, implied by the large panel of S&P 500 index options, and its relation with the aggregate market risks in the economy. As discussed in Andersen et al. (2013), the option panel contains rich information both for the evolution of volatility and jump risks and their pricing. Consequently, we let the option data speak for themselves in determining the risk premium dynamics and discriminating among alternative hypotheses regarding the source of variation in risk as well as risk pricing.

The standard no-arbitrage approach starts by estimating a parametric model characterizing the evolution of the underlying asset price. Risk premia are then introduced through a pricing
kernel which implies that risk compensation is obtained through parameter shifts. This ensures, conveniently, that the associated risk-neutral dynamics remains within the same parametric class entertained for the statistical measure, see, e.g., Singleton (2006), Chapter 15. However, this approach tends to tie the equity market and option surface dynamics closely together. In particular, the equity risk premia are typically linear in (components of) volatility. In contrast, we find the options to display risk price variation that is largely unrelated to, and effectively unidentifiable from, the underlying asset prices alone.

This motivates our “reverse” approach of starting with a parametric model for the risk-neutral dynamics and estimating it exclusively from option data along with no-arbitrage restrictions based on model-free volatility measures constructed from the underlying asset data. In this manner, we avoid letting a (possibly misspecified) parametric structure for the $\mathbb{P}$-dynamics impact the identification of option risk premia. We then explore the risk premia dynamics by combining the extracted state vector with high-frequency data on the equity index.

We document that even a very general two-factor stochastic volatility model with jumps both in price and volatility as well as a time-varying jump intensity produces systematic biases in the fit to the option surface. The problems are particularly acute following periods like the Asian crisis in 1997 or the great recession originating in the Fall of 2008. After these events, market volatility reverts back towards its pre-crisis level fairly quickly, while out-of-the-money put options remain expensive, inducing a steepening of the (Black-Scholes) implied volatility curve. This type of variation in the option surface is difficult to accommodate for standard (no-arbitrage or equilibrium) models as they imply that the priced jump tail risk, which in turn determines the out-of-the-money put prices, is governed solely by (components of) market volatility. This feature implies, in particular, that these models tend to generate risk premia that are too low in the aftermath of crises.

We follow Andersen et al. (2013) by introducing a third factor driving the risk-neutral jump intensity. This novel tail factor is not part of the volatility dynamics although it may be correlated with the level of volatility. We model it as purely jump driven, with one component jointly governed by the volatility jumps while another is independent of the volatility process. This feature allows the jump intensity to escalate – through so-called self-excitation of the jumps – in periods of crises when price and volatility jumps are prevalent, thus “magnifying” the response of the jump intensity to major (negative) market shocks. The extended model remains within the popular class of affine jump-diffusion models of Duffie et al. (2000) and exemplifies the flexibility of such models for generating intricate, yet analytically tractable, dynamic interactions between volatility and jump risks. This type of extension has not been explored in prior empirical option pricing studies.
Relative to the empirical analysis in Andersen et al. (2013), we estimate the model by mini-
mizing, not the squared, but the relative squared option pricing error across the full sample. This
reduces the weight assigned to highly turbulent periods where the bid-ask spreads and pricing errors
increase sharply. We find that the tail factor improves the characterization of the option surface
dynamics very significantly. In particular, the new model no longer undervalues short-maturity
out-of-the-money puts in the aftermaths of crises. Hence, our extended risk-neutral model provides
a more suitable basis for studying the dynamic properties of market risk premia.

In turn, the presence of this independently evolving tail factor implies that part of the risk
premium dynamics cannot be captured by the state variables driving the underlying asset price
dynamics. This implies that this jump risk factor may have predictive power for future risk premia
over and above what is implied by the volatility factors. This is indeed what we find. The novel
tail factor is significant in forecasting future excess market returns for horizons up to one year while
the volatility factors are insignificant. Similarly, the new tail factor is important for predicting the
future variance risk premium in conjunction with one of the volatility components. Taken together,
our findings rationalize why the variance risk premium provides superior forecasts for future returns
relative to volatility itself, as documented in Bollerslev et al. (2009). The key is the existence of
the new factor driving the left jump tails of the risk-neutral distribution.

Importantly, while the new jump factor has predictive power for risk premia, it contains no in-
cremental information regarding the future evolution of volatility and jump risks for the underlying
asset relative to the traditional volatility factors. Hence, our findings indicate that option markets
embody critical information about the market risk premia and its dynamics which is essentially
unidentifiable from stock market data alone. Moreover, the option surface dynamics contains in-
formation that can improve the modeling and forecasting of future volatility and jump risks, but
such applications necessitate an initial untangling of the components in the risk premia that evolve
independently from the volatility process. Overall, our empirical results suggest that there is a
wedge between the stochastic evolution of risks in the economy and their pricing, with the latter
typically having a far more persistent response to (negative) tail events than the former.

Our finding of a substantial wedge between the dynamics of the option and stock markets
presents a challenge for traditional structural asset pricing models. Specifically, the standard
exponentially-affine equilibrium models with a representative agent equipped with Epstein-Zin
preferences imply that the ratio of the risk-neutral and statistical jump tails is constant. On
the contrary, the new factor, extracted from the option data, drives the risk-neutral jump tail but
has no discernable impact on the statistical jump tail. We conjecture that this wider gap between
fundamentals and asset prices may be accounted for through an extension of the preferences via some form for time-varying risk aversion and/or ambiguity aversion towards extreme downside risk.

The rest of the paper is organized as follows. Section 2 describes our data. Section 3 presents our extended three-factor model which, combined with flexibility in the modeling of the jump distributions, encompass most existing models in the empirical option pricing literature. Section 4 introduces the estimation methodology and discusses the parameter estimates and the formal diagnostic tests for the fit to the option surface. Section 5 explores alternative diagnostics for model fit and different robustness checks. This analysis brings out some of the mechanisms behind the improved fit of our three factor model. Section 6 is dedicated to an out-of-sample analysis, documenting the robustness of the estimation results and inference. In Section 7 we exploit the estimation results to study the risk premium dynamics and its implication for return and variance predictability. Our findings are contrasted to corresponding predictability results implied by popular structural equilibrium models. Section 8 concludes. In a Supplementary Appendix we report estimation results on subsamples as well as for various alternative specifications for the risk-neutral dynamics of the underlying index. The Supplementary Appendix also contains additional diagnostic tests related to the parametric fit for the option surface.

2 Data and Preliminary Analysis

We use European style S&P 500 equity-index (SPX) options traded at the CBOE. We exploit the closing bid and ask prices reported by OptionMetrics, applying standard filters and discarding all in-the-money options, options with time-to-maturity of less than 7 days, as well as options with zero bid prices. For all remaining options, we compute the mid bid-ask Black-Scholes implied volatility. The data spans January 1, 1996–April 23, 2013. It is further divided into an in-sample period covering January 1, 1996–July 21, 2010, and an out-of-sample period consisting of July 22, 2010–April 23, 2013. Following earlier empirical work, e.g., Bates (2000) and Christoffersen et al. (2009), we sample every Wednesday. The in-sample period includes 760 trading days, and the estimation is based on an average of 234 bid-ask quotes per day The out-of-sample period contains 142 trading days and features substantially more quotes, and we end up exploiting an average of 708 option contracts per day from this sample. The nonparametric estimate of volatility used for penalizing the objective function below is constructed from one-minute data on the S&P 500 futures covering the time span of the options. The same data are used to construct measures of volatility.

1 This is a “true” out-of-sample period as the entire analysis was completed for the in-sample period before we obtained the additional data covering the more recent years.

2 Due to extreme violations of no-arbitrage-conditions, we replaced October 8, 2008, with October 6, 2008.
and jump risks for the predictive regressions in Section 7. Finally, we employ the returns on the SPY ETF traded on the NYSE, which tracks the S&P 500 index portfolio, and the 3-month T-bill rate to proxy for the risk-free rate, when implementing these regressions.

2.1 The Option Panel

We denote European-style out-of-the-money option prices for the asset $X$ at time $t$ by $O_{t,k,\tau}$. Assuming frictionless trading in the options market, the option prices at time $t$ are given as,

$$
O_{t,k,\tau} = \begin{cases} 
E_t^Q \left[ e^{-\int_t^{t+\tau} r_s \, ds} (X_{t+\tau} - K)^+ \right], & \text{if } K > F_{t,t+\tau}, \\
E_t^Q \left[ e^{-\int_t^{t+\tau} r_s \, ds} (K - X_{t+\tau})^+ \right], & \text{if } K \leq F_{t,t+\tau},
\end{cases}
$$

(2.1)

where $\tau$ is the tenor of the option, $K$ is the strike price, $F_{t,t+\tau}$ is the futures price for the underlying asset at time $t$ referring to date $t + \tau$, for $\tau > 0$, $k = \ln(K/F_{t,t+\tau})$ is the log-moneyness, and $r_t$ is the instantaneous risk-free interest rate. Finally, we denote the annualized Black-Scholes implied volatility corresponding to the option price $O_{t,k,\tau}$ by $\kappa_{t,k,\tau}$. This merely represents an alternative notational convention, as the Black-Scholes implied volatility is a strictly monotone transformation of the ratio $\frac{e^{r_{t,t+\tau}O_{t,k,\tau}}}{F_{t,t+\tau}}$, where $r_{t,t+\tau}$ denotes the risk-free rate over the period $[t, t + \tau]$.

The empirical work explicitly accounts for measurement error in the option prices. We denote the average of the bid and ask quotes (expressed in Black Scholes implied volatility units) by $\bar{\kappa}_{t,k,\tau}$, and view this as a noisy observation of underlying value. To the extent the measurement errors are not strongly correlated across a large fraction of the surface, we improve the efficiency of the inference by incorporating the full option cross-section in our estimation and testing procedures, effectively averaging out idiosyncratic observation errors. The size of the spread varies over time and is positively correlated with the volatility level. In addition, there are systematic differences in the relative spread across moneyness. For example, the spread is about 8% of the mid-spread level for deep OTM puts, on average, implying that a typical implied volatility (IV) reading of 40% is associated with bid and ask quotes of 38.4% and 41.6%. Similarly, for an IV of 18% for ATM options, the quotes are generally around 17.6%-18.4%, while a typical set of quotes for far OTM calls are 18.8%-21.2% for a mid-point value of 20%.

The options underlying the implied volatility (IV) surface are highly heterogeneous in terms of moneyness and tenor across time. To facilitate comparison, we create a uniform set of regions based on the option characteristics. Specifically, we define the volatility-adjusted moneyness, $m$, at time $t$ for tenor $\tau$, by standardizing the log-moneyness with the at-the-money implied volatility,

$$
m = \frac{k}{\kappa_{t,0,\tau} \cdot \sqrt{\tau}}.
$$
Table 3 shows how the observations in our sample are distributed across the option surface. The four regions of moneyness represent deep OTM put options, OTM put options, ATM options and OTM call options, while the two categories for time-to-maturity provide a rough split into short versus long dated options. Not surprisingly, there is particularly good coverage for ATM options, which represent over 44% of the in-sample observations. The quotes for the OTM call options are somewhat limited and amount to almost 16% of the total options quotes, roughly matching the proportion of deep OTM put options. Finally, over 24% fall in the OTM put option region.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau \leq 60 )</td>
<td>10.36</td>
<td>18.46</td>
</tr>
<tr>
<td>( \tau &gt; 60 )</td>
<td>5.23</td>
<td>8.62</td>
</tr>
<tr>
<td>( m \leq -3 )</td>
<td>12.34</td>
<td>12.31</td>
</tr>
<tr>
<td>( -3 &lt; m \leq -1 )</td>
<td>11.89</td>
<td>11.03</td>
</tr>
<tr>
<td>( -1 &lt; m \leq -1 )</td>
<td>19.58</td>
<td>14.58</td>
</tr>
<tr>
<td>( m &gt; 1 )</td>
<td>8.71</td>
<td>10.34</td>
</tr>
<tr>
<td>( \tau \leq 60 )</td>
<td>10.36</td>
<td>18.46</td>
</tr>
<tr>
<td>( \tau &gt; 60 )</td>
<td>5.23</td>
<td>8.62</td>
</tr>
</tbody>
</table>

Table 1: Relative Number of contracts. We report the percentage of option contracts that, on average, fall within the different combinations of moneyness and tenor for the indicated sample.

In the out-of-sample period, the daily number of active quotes is much higher, especially for deep OTM options. Since we would like to compare model performance across the two samples, we have truncated the set of OTM options included in the recent period to lie within the boundaries of -7.1 for the puts and 2.4 for the calls. These cut-offs correspond to the average minimum and maximum moneyness of the option quotes for our in-sample period. This standardization limits the heterogeneity across the samples. The relative proportion of OTM calls and puts is stable, but we observe a non-trivial shift from ATM to deep OTM put options, with the former now representing about 32.5% and the latter 27% of the overall observations. To the extent variation in the pricing of deep OTM put option prices is harder to accommodate than for ATM prices, this composition effect will – all else equal – imply a worse out-of-sample fit than would otherwise be observed.

The option IV surface displays a highly persistent and nonlinear dynamic which is difficult to convey effectively through a few summary measures. We provide a couple of alternative depictions that highlight different aspects of the dynamic. The first approach emphasizes the evolution of separate surface characteristics, as represented by the (annualized) IV level, term structure, skew, and skew term structure. These quantities are plotted in Figure 1. Our second approach consists of a standard principal component analysis of the IV surface, and we explore these in detail below.
In Figure 1, the level captures the average IV for ATM short-dated options, the term structure reflects the difference between the IV of long and short maturity ATM options, the skew measures the IV gap between short-dated OTM put and OTM call options, and, finally, the skew term structure is the difference between the skew computed from long- and short-dated options. The IV level displays occasional erratic spikes to the upside, but also displays strong persistence, as expected for a series reflecting the general level of volatility. Inspecting the remaining three panels, the degree of commonality is striking. Every major spike in the IV level is visible in the other characteristics, albeit in the opposite direction for the IV and skew term structures. That is, rapidly rising volatility is accompanied by a downward sloping term structure, a steepening of the (short) IV skew, and a sharp increase in the short versus long skew.

Table 2 supplement Figure 1 with summary statistics for the IV surface characteristics. The correlation matrix confirms the strong covariability between the IV level and the remaining features. Moving to the individual characteristics, we see that the IV term structure is moderately positive, apart from episodic large negative outliers. The skew is consistently strongly positive, exceeding 10% over the vast majority of the sample and averaging 18%. The skew also displays strong persistence and is particularly elevated when markets are turbulent. Finally, the skew term structure is negative for almost the entire sample, so the “smirk” flattens substantially with option tenor. This latter feature accounts for a great deal of variation in the surface shape, as the skew is extremely negatively correlated with the skew term structure. Thus, when volatility soars and the skew steepens, the effect is typically much less pronounced at the longer maturities. Hence, this type of excitement of the short left part of the IV surface must be associated with shocks to volatility and jump intensities of low to moderate persistence, as the effect is strongly at longer horizons. This is also consistent with the covariation of the IV and skew term structures as well as the comparatively mild negative association between the skew and IV term structure.

Turning to the principal component (PC) analysis, Figure 2 depicts the in-sample realizations of the four first PCs. It is evident that the first PC is closely related to the IV level, while the second PC displays commonality with the IV term structure. However, the last two PCs appear largely unrelated to the characteristics depicted in Figure 1. The first PC captures about 95.7% of the total variation, while the following PCs account for, respectively, 2.3%, 0.7% and 0.3%. Clearly, there is a dominant level type effect, but this “factor” also accounts for a great deal of variation in the skew, term structure and skew term structure, leaving, relatively speaking, only minor residual variation to explain for the remaining PCs.
Figure 1: **Implied volatility (IV) surface characteristics.** Top left Panel: IV Level defined as the average IV for short ATM options. Top Right Panel: IV Term Structure defined as the difference in IV between long and short ATM options. Bottom left Panel: IV skew defined as the difference between IV of short OTM put and OTM call options. Bottom right Panel: skew term structure, defined as the difference between long and short skew, where the long skew is defined as the short skew, but using long-dated options. We define short options as those with less than 21 days to maturity and long options as those with more than 180 days to maturity. We define ATM options with volatility-adjusted moneyness $m$ within $[-0.2, 0.2]$; OTM put options with $m < -3$ (short maturities) and $m < -1.5$ (long maturities); OTM call options with $m > 1$.

<table>
<thead>
<tr>
<th>Summary statistics</th>
<th>Correlation Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td><strong>Median</strong></td>
</tr>
<tr>
<td>Level</td>
<td>0.18</td>
</tr>
<tr>
<td>TS</td>
<td>0.02</td>
</tr>
<tr>
<td>Skew</td>
<td>0.20</td>
</tr>
<tr>
<td>Skew TS</td>
<td>-0.05</td>
</tr>
</tbody>
</table>

Table 2: **Summary statistics and correlation matrix for surface characteristics.** The statistics for the implied volatility Level, Term Structure (TS), Skew, and Skew Term Structure are computed over the in-sample period, January 1996–July 2010.
Figure 2: **Principal Components of the Implied Volatility Surface.** The panels depict the first four principal component extracted from the S&P 500 implied volatility surface from January 1996 to July 2010. On each day we interpolate the IV surface to generate standardized options with the same volatility-adjusted moneyness, $m$, and tenor across the sample. Specifically, we obtain option prices for $m$ in the set $\{-4, -3, -2, -1, 0, 1\}$ and tenor equal to three values, $\{0.1, 0.3, 1\}$ (in years). This produces a total of 18 synthesized option contracts per day. For each panel, we also report the percentage of the overall variation explained by the given principal component.
To further explore how the PCs interact with the IV surface characteristics, Table 3 reports on the in-sample regression of characteristics on PCs. The table confirms the strong association of the first PC with the IV level, featuring a $t$-statistic beyond 100. For this PC, we also obtain the negative association with the IV and skew term structures and positive relation with the skew, matching the correlation patterns for the IV level in Table 2. In addition, we find the second PC to be associated with the IV term structure, corroborating the visual impression from Figure 2. The third PC is weakly associated with the skew as well as the skew term structure and level. However, the skew is much more directly associated with the first PC so, effectively, the third PC captures only residual skew variation that is largely orthogonal to the level. Finally, the fourth PC has no identifiable relation to any of the characteristics. In particular, the variation in the skew term structure is effectively accounted for through the first two or three PCs.

<table>
<thead>
<tr>
<th></th>
<th>Level</th>
<th>TS</th>
<th>Skew</th>
<th>Skew TS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC1</td>
<td>0.19 (147.92)</td>
<td>-0.05 (-34.01)</td>
<td>0.19 (66.82)</td>
<td>-0.09 (-27.52)</td>
</tr>
<tr>
<td>PC2</td>
<td>-0.28 (-32.92)</td>
<td>0.46 (51.98)</td>
<td>0.01 (0.32)</td>
<td>0.45 (20.57)</td>
</tr>
<tr>
<td>PC3</td>
<td>-0.18 (-11.11)</td>
<td>-0.07 (-4.02)</td>
<td>0.64 (18.34)</td>
<td>-0.43 (-10.52)</td>
</tr>
<tr>
<td>PC4</td>
<td>0.14 (6.13)</td>
<td>-0.14 (-5.70)</td>
<td>-0.45 (-8.90)</td>
<td>-0.02 (-0.34)</td>
</tr>
<tr>
<td>R2</td>
<td>0.97</td>
<td>0.84</td>
<td>0.87</td>
<td>0.63</td>
</tr>
<tr>
<td>AC(1)</td>
<td>0.35</td>
<td>0.47</td>
<td>0.42</td>
<td>0.53</td>
</tr>
<tr>
<td>AC(2:10)</td>
<td>0.18</td>
<td>0.18</td>
<td>0.08</td>
<td>0.15</td>
</tr>
<tr>
<td>AC(11:20)</td>
<td>0.10</td>
<td>0.10</td>
<td>0.07</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Table 3: Relating Option Characteristics and Principal Components. We report the coefficients with $t$-statistics (in parentheses) and the $R^2$ from the linear regression of different options characteristics (level, term structure, slope, and slope term structure) on the first four principal components extracted from the S&P 500 implied volatility surface from January 1996 to July 2010. We also report the first sample autocorrelation coefficient and the average sample autocorrelation coefficient over two to ten and eleven to twenty lags of the regression residuals.

These observations have a number of implications. First, there is no simple mapping between PCs and characteristics. The latter are intrinsically interconnected and covary strongly. Hence, it is impossible to associate a specific factor with features of a given characteristic. Such factors will generally exert a significant joint impact on multiple characteristics. Second, the difficulty of separating the forces driving the individual characteristics is, of course, a general feature of systems with pronounced nonlinearities. We see further evidence of potential nonlinearities in the
large degree of serial correlation in the residuals of the regression in Table 3. The first order autocorrelation coefficient is very large across all the characteristics, and they die out extremely slowly, suggesting highly persistent deviations from the linear approximation associated with the regression analysis. In short, the IV surface dynamics is likely highly nonlinear in the underlying factors. Third, the persistent residuals can be interpreted as evidence for missing factors and, indeed, we found standard tests for the number of factors, developed in, e.g., Bai and Ng (2002), to indicate 7 to 8 factors. Given the nonlinear association between option IV and factors, this likely reflects the failure of the linear approximation rather than an indication of the true number of underlying (linear) factors. Fourth, given the nonlinear relation between the the IV surface and underlying factors, it is problematic to associate the first few (linear) PCs with actual factors.

In summary, accounting for a given model’s ability to track the IV surface dynamics provides an intuitive and useful way to highlight the implications of model misspecification. Indeed, we later illustrate the quality of our own model fit by comparing the model-implied evolution of the characteristics to the IV surface characteristics generated by the data. On the other hand, there is no direct association between the inability to fit the dynamics of a specific set of characteristics and the lack of a given factor, because the factors’ impact on the option IVs are highly nonlinear and create correlated dynamic interactions across the characteristics. Later in the paper, we exploit new diagnostic tools for the model fit to broader regions of the IV surface, which provide a more formal basis for guiding model specification and selection.

3 Parametric Modeling of the Option Panel

We adopt a general-to-specific approach to the parametric modeling of the risk-neutral equity-index returns, with the objective of capturing the salient features of the IV surface dynamics succinctly. The model subsequently serves as the basis for our analysis of the equity and variance risk premiums. The main initial decision concerns the number of latent state variables to include. The vast majority of empirical option pricing studies employs a single stochastic volatility factor, but the literature on estimating the return dynamics under the physical measure as well as a few option pricing studies, e.g., Bates (2000), Christoffersen et al. (2009), Christoffersen et al. (2012), and Andersen et al.

For a simple illustration, we calibrated a one-factor Heston model and derived the sensitivity of the IV surface to the volatility factor. For a given location in the surface, i.e., a given mildly OTM option, the sensitivity (derivative of the option IV with respect to volatility) will typically vary dramatically with the level of volatility. Thus, strongly correlated approximation errors will arise from linearizing the dependence of the surface characteristics to factors, as the true sensitivities fluctuate strongly over time. This illustration is provided in our supplementary web appendix.

In this respect, the dynamics of the IV surface deviates sharply from that of the the term structure of interest rates. For the latter, the first three PCs extracted from the variation in the yields across the available maturities provide excellent linear spanning for the full yield curve.
(2013), point to a minimum of two factors. This is also consistent with our descriptive analysis of the IV surface in Section 2 which found the characteristics to load strongly on the first two principal components and quite significantly on the third PC as well. Hence, we follow Andersen et al. (2013) in proposing a general three-factor model which, apart from the specification of the jump distribution, embeds all existing continuous-time models in the literature as special cases. We document below that exponentially distributed price jumps provide a superior fit relative to the more commonly adopted Gaussian specification, justifying the alternative representation in our benchmark model. After eliminating those features, that are jointly economically and statistically insignificant, we arrive at a tractable and relatively parsimonious three-factor model that provides a satisfactory fit to the option surface, both in- and out-of-sample.

Our three-factor model for the risk-neutral equity index dynamics is given by the following extension of the model proposed in Andersen et al. (2013),

\[
\begin{align*}
\frac{dX_t}{X_{t-}} &= (r_t - \delta_t) dt + \sqrt{V_{1,t}} dW_{1,t}^Q + \sqrt{V_{2,t}} dW_{2,t}^Q + \eta \sqrt{U_t} dW_{3,t}^Q + \int_{\mathbb{R}^2} (e^x - 1) \tilde{\mu}^Q(dt, dx, dy), \\
V_{1,t} &= \kappa_1 (V_1 - V_{1,t}) dt + \sigma_1 \sqrt{V_{1,t}} dB_{1,t}^Q + \mu_1 \int_{\mathbb{R}^2} x^2 1_{\{x<0\}} \mu(dt, dx, dy), \\
V_{2,t} &= \kappa_2 (V_2 - V_{2,t}) dt + \sigma_2 \sqrt{V_{2,t}} dB_{2,t}^Q, \\
U_t &= -\kappa_u U_t dt + \mu_u \int_{\mathbb{R}^2} [(1 - \rho_u) x^2 1_{\{x<0\}} + \rho_u y^2] \mu(dt, dx, dy),
\end{align*}
\]

where \((W_{1,t}^Q, W_{2,t}^Q, W_{3,t}^Q, B_{1,t}^Q, B_{2,t}^Q)\) is a five-dimensional Brownian motion with \(\text{corr} \left( W_{1,t}^Q, B_{1,t}^Q \right) = \rho_1\) and \(\text{corr} \left( W_{2,t}^Q, B_{2,t}^Q \right) = \rho_2\), while the remaining Brownian motions are mutually independent. In addition, \(\mu\) is an integer-valued counting variable (random measure on \(\mathbb{R}_+ \times \mathbb{R}^2\)), representing the jumps in the price, \(X\), as well as the state vector, \((V_{1,t}, V_{2,t}, U_t)\). The corresponding (instantaneous) jump intensities, under the risk-neutral measure, is \(dt \otimes \nu_t^Q(dx, dy)\). This process is also known as the (jump) compensator, and the difference between the jump realizations and the compensator, 

\(\tilde{\mu}^Q(dt, dx, dy) = \mu(dt, dx, dy) - dt \otimes \nu_t^Q(dx, dy)\), constitutes the associated martingale (jump) measure.

The jump specification involves two separate components, \(x\) and \(y\). The former captures co-jumps that occur simultaneously in the price, the first volatility factor, \(V_{1,t}\), and, potentially, in the \(U_t\) factor (if \(\rho_u < 1\)), while the \(y\) jumps represent independent shocks to the \(U_t\) factor. The latter can also generate a jump in return volatility (if \(\eta > 0\), but the main effect is through the jump intensities. The compensator characterizes the conditional jump distribution. It is given by,

\[
\frac{\nu_t^Q(dx, dy)}{dx dy} = (e^{-r(t)} \cdot 1_{\{x<0\}} \lambda_- e^{-\lambda_- |x|} + e^{r(t)} \cdot 1_{\{x>0\}} \lambda_+ e^{-\lambda_+ x}) 1_{\{y=0\}} + c^{-}(t) \cdot 1_{\{x=0, y<0\}} \lambda_- e^{-\lambda_- |y|}.
\]
The first term on the right hand side, referring to the $x \neq 0, y = 0$ case, reflects co-jumps in price and volatility, while the second term, $x = 0, y < 0$, captures independent shocks to the $U_t$ factor. Hence, the individual (strictly positive) jumps in $U_t$ are either independent from $V_1$ or proportional to the (simultaneous) jump in $V_1$. The price jumps are exponentially distributed, with separate tail decay parameters, $\lambda_-$ and $\lambda_+$, respectively, for negative and positive jumps. Moreover, for parsimony, the independent shocks to the $U_t$ factor is distributed identically to the negative price jumps. Finally, the time-varying jump intensities are governed by the $c^-(t)$ and $c^+(t)$ coefficients. These coefficients evolve as affine functions of the state vector,

$$
c^-(t) = c^+_0 + c^-_1 V_{1,t-} + c^-_2 V_{2,t-} + U_{t-}, \quad c^+(t) = c^+_0 + c^+_1 V_{1,t-} + c^+_2 V_{2,t-} + c^+_u U_{t-}.
$$

This representation involves a large set of parameters that can be hard to identify separately. At the estimation stage, we eliminate those that are insignificant and have no discernable impact on model fit. However, generality along this dimension is important, as the jump specification turns out to be critical for a suitable representation of the IV surface dynamics.

Our three-factor model possesses a number of distinctive features. The factors $(V_1, V_2, U)$ drive both the diffusive volatility and the jump intensities. $V_1$ and $V_2$ are always present in the diffusive volatility, as in traditional multi-factor volatility models, while $U$ contributes to diffusive volatility only if $\eta > 0$. In fact, the constrained model arising for $\eta = 0$ (or very small) is of separate interest. It implies that the factor $U_t$ affects only the jump intensities, with no (little) impact on diffusive volatility. Furthermore, in an extension to existing option pricing models, we allow positive and negative jump intensities to have different loadings on the latent factors. In particular, some factors affect only positive or only negative jump intensities. Such flexibility in modeling the jump intensities is important given the nonparametric evidence in Bollerslev and Todorov (2011b).

The jump modeling also involves several distinctive features. First, the price jumps are exponentially distributed. This is unlike most earlier option pricing studies which rely on Gaussian price jumps, following Merton (1976). Nonparametric evidence in Bollerslev and Todorov (2013) suggests the exponential price jumps capture the behavior of short-maturity OTM puts much better, and we confirm this in our empirical analysis. Next, the jumps in the factor $V_1$ is linked deterministically to the price jumps, with squared price jumps impacting the volatility dynamics in a manner reminiscent of GARCH models. For parsimony and ease of identification, we allow only the negative price jumps to impact the volatility dynamics. Finally, $U$ is driven in part by the (same) squared negative price jumps and in part by independent jumps, with the parameter $\rho_u$ controlling

\[ \text{Support for this type of price-volatility jump dependence, albeit under } P, \text{ may be found in Todorov (2011).} \]
the contribution of each component in the dynamics of \( U \). As such, the model accommodates both perfect dependence \( (\rho_u = 0) \) and full independence \( (\rho_u = 1) \) between the jump risks of \( V_1 \) and \( U \). Moreover, these state variables, governing important features of the option surface dynamics, are related through the time-variation in the jump intensity. Our specification allows for “cross self-excitation” in which jumps in \( V_1 \) enhance the probability of future jumps in \( U \), and vice versa.\(^6\)

Differences in the jump distributions aside, our model nests most existing models as special cases. In particular, in the one-factor setting, with \( V_2 \) and \( U \) absent, we recover the double jump volatility model of Duffie et al. (2000), estimated using option data by Broadie et al. (2007). In the two-factor setting, with \( U \) absent and excluding volatility jumps (in \( V_1 \)), we obtain the Bates (2000) jump-diffusion, and further ruling out price jumps leads to the two-factor diffusive model of Christoffersen et al. (2009). Finally, our jump dynamics shares some qualitative features with the discrete-time model in Christoffersen et al. (2012). Consequently, we generalize existing two-factor continuous-time models by incorporating their main features jointly, while further allowing for volatility jumps, an ingredient stressed by Broadie et al. (2007). Finally, we introduce the new jump-driven factor, \( U \), which accommodates additional nonlinearities in the option surface dynamics, while still retaining the tractability afforded by the exponentially-affine setting.

The main departure from prior work stems from the inclusion of the new \( U \) factor. Given the rather unconventional representation, we briefly discuss how this factor enhances the features of the risk-neutral dynamics. It is best seen by focusing on a restricted version of the model where we eliminate \( U \) from the diffusive volatility, i.e., set \( \eta = 0 \), and further let \( \rho_u = 0 \), so that there is no independent jump component driving \( U \). In this scenario, there is no distinct source of risk impacting \( U \). It is driven solely by the squared negative price jumps. Nonetheless, since \( U \) affects the jump intensities separately from \( V_1 \), it may still influence the option surface dynamics in critical ways. That is, to convey the current state of the system, \( U \) must be included among the components of the state vector.\(^7\) In other words, even if the source of risk in \( U \) is spanned by the jumps in \( X \) and \( V_1 \), \( U \) is still necessary for characterizing the distribution of future log returns or predicting the future evolution of the factors, even after controlling for the current values of \( V_1 \) and \( V_2 \).

---

\(^6\)We stress that the model (3.1) still belongs to the affine family covered by Duffie et al. (2003) and, as shown in Andersen et al. (2013), the following parameter constraints ensure covariance stationarity of the latent factors,

\[
\kappa_1 > \frac{2\kappa_1 \mu_1}{\lambda^2}, \quad \kappa_3 > \frac{2\kappa_3 \kappa_1 \mu_u}{\kappa_1 \lambda^2 - 2\kappa_1 \mu_1}, \quad \text{and} \quad \kappa_2 < 0, \quad \text{and} \quad \sigma_i^2 \leq 2\kappa_i \bar{\eta}_i, \quad i = 1, 2.
\]

\(^7\)This is a reflection of the fact that there, in general, is no direct association between the dimensionality of the sources of risk and the dimension of the state vector required to characterize the conditional dynamics of the system. This phenomenon arises in continuous-time ARMA model, see, e.g., Brockwell (2001). It is also a well-known feature of the so-called quadratic term structure models, see, e.g. Ahn et al. (2002) and Leippold and Wu (2002).

14
course, the role of $U$ only expands, if it is subject to independent shocks as well. In the empirical section below, we detail how the $U$ factor impacts the IV surface characteristics over time.

The model in equation (3.1) pertains to the risk-neutral dynamics of $X$. However, due to the equivalence of $Q$ and $P$, the assumed dynamics have implications for the dynamics of $X$ under $P$ as well. In general, these implications are limited to those features of model (3.1) which hold almost surely. They consist of the following. First, the spot diffusive variance is invariant to the change of measure. In the model, this is given by $V_t = V_{1,t} + V_{2,t} + \eta^2 U_t$. Second, regarding the jumps, the only property that applies almost surely is the identity of the realized jumps in the returns and state variables. In particular,

$$
\Delta V_{1,t} = \mu_1 (\Delta \log(X))^2 1_{\{\Delta \log(X) < 0\}},
$$

and, if $\rho_u = 0$, we also have,

$$
\Delta U_t = (\Delta \log(X))^2 1_{\{\Delta \log(X) < 0\}},
$$

under both measures.

In most empirical option pricing applications, additional assumptions are invoked when changing measure from $P$ to $Q$. For example, it is commonly assumed that the model class is identical, and affine, under both measures. This is convenient as affine models offer a great deal of tractability. However, this approach severely restricts the dynamics of the risk premiums. In addition, such “structure preserving transformations” (SPTs) impose auxiliary restrictions, extending to the model parameters. In particular, in our affine setting, the SPT assumption implies that $\sigma_1$, $\sigma_2$, $\sigma_3$, $\rho_1$, $\rho_2$, $\rho_u$, $\mu_1$ and $\kappa_3$ are identical, while the remaining parameters may differ under $P$ and $Q$.

Given the above discussion, it is clear that data on the underlying equity-index values can be helpful in the estimation of the risk-neutral model. Below, we impose the pathwise (almost sure) restriction regarding the spot variance in estimation. Given the difficulty in recovering spot volatility jumps from high-frequency data – due both to estimation uncertainty and lack of overnight observations – we do not imposing restrictions regarding the pathwise (realized) price and volatility jump relationship implied by the risk-neutral model (3.1). As detailed above, if one constrains the pricing of risk through a SPT from $P$ to $Q$, additional information from the $P$ dynamics can be “imported” from the underlying return data during estimation of the risk-neutral dynamics. However, as we document below, the option panel is very informative about the risk neutral dynamics, so we avoid auxiliary restrictions that may induce model misspecification.
4 Estimation

4.1 Estimation Approach

The development of formal tools for parametric inference in the context of an option panel is challenging. There are pronounced time series dependencies in the latent volatility components and, potentially, the jump intensities. At the same time, sizeable bid-ask spreads influence the observed transaction prices and quotes for the options. These measurement errors for the options are strongly heterogeneous and correlated with the overall return variation. Finally, there are no-arbitrage constraints that, one, at any point in time link the individual option prices across strikes, and, two, equate the spot diffusive volatility for the underlying asset (imperfectly observable from high-frequency returns) with the volatility implied by the contemporaneous state vector (also extracted with some degree of statistical error) every time we observe the option cross-section. The interactions of these effects generate complex nonlinear interactions and dependencies in the system which should be accounted for in any practical assessment of estimation precision of quality of fit to the option surface.

In this paper, we adapt the parametric estimation and inference approach put forth in Andersen et al. (2013) which is designed to deal with this exact type of environment. It exploits in-fill asymptotics in the option cross-section, i.e., it develops inference tools under the assumption that, for a limited set of maturity dates, option prices are observed across a broad range of strike prices with only small gaps between the exercise prices. The procedure accommodates that the option panel is highly unbalanced, with a different set of maturities and strikes available for each cross-section, the measurement errors are heterogeneous and correlated with the overall return variation, the explicit incorporation of the no-arbitrage constraint linking spot volatility under the physical and risk-neutral measure, and avoids relying on a structure preserving transformation from $\mathbb{P}$ to $\mathbb{Q}$ that involves ad hoc restrictions on the risk pricing.

The approach provides consistent period-by-period estimates for the state vector along with valid asymptotic inference for the state vector as well as the model parameters and the model fit to specific regions of the IV surface for any observation date. This is feasible only due to the in-fill asymptotic scheme for the strikes in the option cross-section.\footnote{Alternative inference techniques can potentially provide unbiased estimates for the state vector determining the current values of the multiple volatility components and the jump intensities, but they cannot achieve consistency within this setting, thus rendering formal inference regarding the state vector and IV surface fit on a period-by-period basis infeasible.} At the practical level, this reflects the given actual structure of the panel. The option quotes are clustered closely across the strike range while the cross-sections are observed weekly. Thus, the well-recognized advantages of
inference via “high-frequency” observations apply naturally to the cross-section, not the time series, in the given setting. Moreover, the wealth of information embedded in the option cross-section is, of course, well-recognized. It enables nonparametric extraction of the conditional risk-neutral density estimates for the underlying asset returns across the sequence of maturities represented by the available tenors at a given date. The evolution of these conditional densities over time speak to the number of factors, the intertemporal variation in volatility and jump intensities and persistence of the different characteristics. Given the identified factors and model parameters, the individual cross-section identifies the current state vector. In fact, theoretically, the entire system can be identified and estimated consistently from a single option cross-section, but the identification from a single date is weak and extremely imprecise given the presence of measurement errors.\footnote{On the other hand, quite reasonable identification can usually be obtained from a single year of option data – along with the high-frequency returns on the underlying asset used to enforce the (statistical) equality of spot volatility across the P to Q measures.}

Finally, we emphasize the advantage of obtaining formal diagnostics for model fit. The approach can be used to derive summary measures for whether we, statistically, suitably accommodate specific regions of the IV surface on a period-by-period basis. Below, we illustrate this via Z-scores that capture the overall fit for OTM put and call options along with ATM options at either the short or long maturities. By comparing these period-by-period fit across the strike range, we assess the quality of fit to the skew, while comparison across the short and long tenors reveal the accommodation of the IV term structure. These indications of overall under- or overvaluation of specific options explicitly account for the estimation uncertainty induced my measurement errors in the options, the estimation uncertainty for the model parameters and period-specific state vector and the imprecision in inferring the spot volatility from the high-frequency return data.\footnote{The massive Monte Carlo study in Andersen et al. (2013) confirms that these test procedures perform well in this type of setting.}

Letting $n$ denote the (equidistant) frequency by which we sample the high-frequency returns of the underlying asset, our estimator takes the form,

$$
\left( \{ \hat{S}_n^t \}_{t=1,...,T}, \hat{\theta}^n \right) = \arg\min_{\{ Z_t \}, \theta \in \Theta} \sum_{t=1}^{T} \frac{1}{N_t} \sum_{j=1}^{N_t} \frac{\left( U_{t,k,\tau_j} - \kappa(k_j, \tau_j, Z_t, \theta) \right)^2}{\hat{V}_t^{n}} + \zeta_n \left( \frac{\hat{V}_t^{(n,m_n)} - V_t(Z_t, \theta) }{\hat{V}_t^{n}} \right)^2,
$$

for a deterministic sequence of nonnegative numbers $\{ \zeta_n \}$ (decreasing asymptotically to zero, as $n$ diverges) and $V_t(Z_t, \theta) = V_{1,t} + V_{2,t} + \eta^2 U_t$ is the model-implied value of the spot (diffusive) variance. $\hat{V}_t^{n} \equiv \hat{V}_t^{(n,n)}$ and $\hat{V}_t^{(n,m_n)}$ are nonparametric estimators of the diffusive integrated variance constructed from the intraday record of the log-futures price of the underlying asset. Specifically,
\( \hat{V}_t^{(n,m_n)} \) exploits \( m_n \) returns sampled (at frequency \( n \)) just prior to the observation time for the option cross-section (the end of the trading day). For \( m_n = n \), \( \hat{V}_t^{(n,m_n)} \) is the truncated realized return variation, see, e.g., Mancini (2009), which consistently estimates the integrated diffusive variance over a fixed time period prior to \( t \). For \( m_n/n \to 0 \), \( \hat{V}_t^{(n,m_n)} \) is a consistent estimator of the spot variance at \( t \) and corresponds to the truncated realized variation computed over an (asymptotically shrinking) fraction of the day just prior to the option quotes. In our implementation, we sample every minute over a 6.75 hours trading day, excluding the initial five minutes, resulting in \( n = 400 \). We employ \( m_n = 300 \) for \( \hat{V}_t^{(n,m_n)} \). Moreover, we explicitly account for the pronounced intraday volatility pattern. Details regarding our construction of these estimators are provided in the appendix.

The estimator (4.2) minimizes the weighted mean squared error in fitting the panel of observed option implied volatilities, with a penalization term that reflects how much the model-implied spot variance deviates from a model-free spot variance estimate. The presence of \( \hat{V}_t^{(n,m_n)} \) in the objective function serves as a regularization device that helps identify the parameter vector by penalizing values that imply “unreasonable” volatility levels. Moreover, the standardization by \( \hat{V}_t^n \) allows us to weigh option observations on high and low volatility days differently to improve efficiency by accounting for the fact that the measurement errors in the option prices and the imprecision in estimating the state vector generally rises with market volatility. This represents a departure from the estimation procedure in the empirical analysis in Andersen et al. (2013).\(^{11}\)

The estimator (4.2) is derived through joint optimization over the parameters and state vector realizations but, in practice, it is most conveniently performed iteratively, exploiting two simple steps. For a given parameter vector, we first determine the optimal value of the state vector on each day of the sample. Second, for the concentrated objective function, we perform MCMC based optimization of the parameter vector with a chain of length 10,000. For this new \( \theta \) estimate, we then move back to step one and iterate the procedure.

### 4.2 Estimation Results

The parameter estimates for model (3.1) are reported in Table 4. The basic features of the volatility factors, \( V_1 \) and \( V_2 \), are similar to those obtained via model (??). Nonetheless, the size of \( V_2 \) has declined significantly relative to \( V_1 \). Moreover, \( V_2 \) is less persistent than before with a half life of about 18 months versus one month for \( V_1 \). In addition, the volatility factors exhibit a stronger

\(^{11}\)The rationale for our weighting scheme is apparent from Figure ?? in the supplementary appendix. It demonstrates that the option pricing errors are roughly linear in market volatility, motivating our scaling of the errors by the level of volatility.
Table 4: Parameter Estimates for the extended three factor model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std.</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Std.</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Std.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_1$</td>
<td>-0.913</td>
<td>0.028</td>
<td>$\sigma_2$</td>
<td>0.110</td>
<td>0.006</td>
<td>$c_2^-$</td>
<td>0.913</td>
<td>3.730</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.007</td>
<td>0.000</td>
<td>$\mu_1$</td>
<td>1.756</td>
<td>0.647</td>
<td>$c_2^+$</td>
<td>14.269</td>
<td>4.986</td>
</tr>
<tr>
<td>$\kappa_1$</td>
<td>8.325</td>
<td>0.167</td>
<td>$\kappa_3$</td>
<td>0.522</td>
<td>0.080</td>
<td>$c_3^-$</td>
<td>19.836</td>
<td>5.460</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.323</td>
<td>0.014</td>
<td>$\rho_3$</td>
<td>0.117</td>
<td>0.614</td>
<td>$\lambda_-$</td>
<td>21.157</td>
<td>0.240</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>-0.945</td>
<td>0.036</td>
<td>$c_0^+$</td>
<td>0.723</td>
<td>0.079</td>
<td>$\lambda_+$</td>
<td>48.365</td>
<td>2.053</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>0.016</td>
<td>0.001</td>
<td>$c_{1-}$</td>
<td>34.592</td>
<td>1.931</td>
<td>$\mu_1$</td>
<td>11.602</td>
<td>0.262</td>
</tr>
<tr>
<td>$\kappa_2$</td>
<td>0.480</td>
<td>0.033</td>
<td>$c_{1+}$</td>
<td>88.178</td>
<td>14.711</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Parameter estimates of the three-factor model (3.1). The model is estimated using S&P 500 equity-index option data sampled every Wednesday over the period January 1996-July 2010.

negative correlation with the diffusive price innovations. While most of the jump parameters are not readily comparable to those in the preceding models, we note that the tail decay parameters only change slightly and the overall features are quite similar to what we established with model (3.1). For example, the negative and positive jump intensities are now about 3.4 and 2.6 per year, while the jump sizes are about -4.7% and 2.1%, respectively, representing only minor offsetting changes in the jump intensities and sizes. Likewise, we again find that the jump intensity is driven almost exclusively by the time-varying component. We also observe that the $c_2^-$ and $c_2^+$ coefficients are pairwise statistically distinct, indicating a different degree of time-variation in the left and right jump tail, consistent with the nonparametric evidence in Bollerslev and Todorov (2011b). Further, the state variable $U$ is highly persistent and, as discussed in detail later, is the primary determinant behind the time variation of the left jump tail. Finally, the parameter $\rho_3$ cannot be estimated with precision, implying that the degree of independent jump variation in $U$ is hard to identify from the option panel alone.

The extended three-factor model (3.1) improves the fit to the option surface substantially relative to model (3.1). The RMSE for the option-implied volatilities drops by close to 15% and now equals 1.75%. Even more strikingly, the Z-scores in Figure ?? portray a set of statistics that largely reside within the 95% confidence bands. The tabulation of option pricing violations at the 1% level across the six strike-maturity categories in the Supplementary Appendix corroborates the point.

\footnote{As before, the estimates of $c_0^-$ were insignificant and fixed at zero.}
The week-by-week conditional fit for model (??) versus (3.1) yields violation frequencies ranging from 20% to 39% versus 4% to 15%. That is, every option category is more suitably priced in the three-factor model than is the case for the best fit obtained across all six categories in the two-factor setting. Thus, although evidence of persistent mispricing remains, for example for the OTM put and call options during crisis periods, the violations are smaller and more short-lived, and specifically, the ATM options are priced accurately. We conclude that model (3.1) captures the critical dynamic features reasonably well through the stipulated law of motion for the three state variables. Consequently, the residual mispricing should exert less of an impact on our subsequent analysis relative to the more standard specifications explored above.\textsuperscript{13}

5 Diagnostic of the Option Panel Fit and Model Performance

We have established that model (3.1) provides a much improved statistical fit to the option surface relative to more standard representations of the risk-neutral asset price dynamics. However, it is less evident whether the superior fit translates into a better characterization of the key dynamic factors that determine the future evolution of, and associated risks related to, the implied volatility surface as well as the corresponding risk premiums. The remainder of this section is dedicated to illustrating the critical differences in the implied volatility surface dynamics across the alternative models. The question concerning the risk premiums is addressed in the following section.

5.1 State Vector Dynamics, Volatility Surface Dynamics, and Jump Intensities

<table>
<thead>
<tr>
<th>Summary statistics</th>
<th>Correlation Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Std</td>
</tr>
<tr>
<td>V_{1,t} 0.0195</td>
<td>0.0345</td>
</tr>
<tr>
<td>V_{2,t} 0.0102</td>
<td>0.0081</td>
</tr>
<tr>
<td>U_t     0.1362</td>
<td>0.1378</td>
</tr>
</tbody>
</table>

Table 5: Summary statistics of the Recovered State Vector. AC stands for autocorrelation. Sample averages are for the period January 1996 - July 2010.

In the estimation step for the three-factor model, we extract estimates for the realizations of

\textsuperscript{13} In principle, some of the remaining valuation problems can be rectified via a time-varying jump tail index, i.e., allowing the parameter $\lambda_\cdot$ to vary, see, e.g., the nonparametric evidence in Bollerslev and Todorov (2013). However, time variation in the tail index takes us outside the affine setting, implying that we lose the analytical tractability in pricing the option surface which is critical for our current implementation procedures.
the state vector \((V_1, V_2, U)\) for each week in the sample. Since the state vector, along with the point estimates for the various model parameters, governs the (diffusive) volatility as well as the jump probabilities in the model, they directly imply a path for volatility and for both positive and negative jump intensities over time. Figure ?? displays the implied time series for volatility and the negative jump intensity over our sample period.

![Figure 3: Spot Volatility and the U Factor](image)

**Figure 3: Spot Volatility and the U Factor.** The top panel displays the spot diffusive volatility estimated from the high-frequency data (the dark line) and from the option panel (the light-colored line). The bottom panel displays the estimate of \(\sqrt{c^3} U_t\).

There are many fascinating aspects to the plots in Figure ??: First, we see spikes in market volatility and – even more strikingly – in the negative jump intensity around well-known crises. Second, the negative jump intensity portrays a very different picture than the (diffusive) volatility. For example, the peaks of the (negative) jump intensity in the 1997 and 1998 crises as well as the sovereign debt crisis in Europe match or exceed those observed from 2000 through 2003, yet the model-implied volatility is substantially higher in the latter period than during the 1997-1998 episodes or the European crisis. Hence, the alternating shapes of the option surface signal that the episodes represent very different types of exposure to volatility and (negative) jump risks. Third,
not surprisingly, the financial crisis stands out as the most dramatic period. It is also evident that jump intensities mean-revert less quickly following the financial crisis. This pattern is repeated across all the episodes identified above in which the jump intensity rises strongly relative to the volatility. Fourth, focusing on the part of the jump intensity that is accounted for by the new jump factor $U$, we see a similar pattern. For the Asian, Russian and European crises, the elevated jump intensity is almost exclusively driven by $U$, while this is not the case for the volatile episodes surrounding 9-11 and the bursting of the internet bubble in 2002. Likewise, the initial spike in the jump intensity during the financial crisis was not caused solely by the $U$ factor, but the latter was clearly the determinant behind the very slow reversion thereafter. Finally, we note that the overall pattern for both volatility and jump intensities qualitatively matches the corresponding figure in Andersen et al. (2013), based on the identical model, but estimated via a pure RMSE criterion for the option-implied volatility. The current model is estimated by the same criterion, but applied to the weighted option-implied volatility errors. The weighting allows the model to accommodate the shape of the option surface equally across different volatility states, while the pure (unweighted) RMSE criterion forces the model to fit the surface as closely as possible during periods with extreme option price fluctuations. This effect is evident here, as the model-implied variance estimates in Andersen et al. (2013) peaks above 0.35 during the financial crisis and also features other values around 0.30. In contrast, for the current estimation approach, the volatility estimates never exceed 0.28. The discrepancies between the two sets of underlying parameter estimates are discussed further in the following subsection.

### 5.2 Fitting the Option Characteristics

Figure ?? summarizes key time series implications of model (3.1). In order to understand whether the model represents a major deviation from the one- and two-factor models estimated previously, we compare the implied properties of the resulting systems along critical dimensions. Figure ?? plots pairwise model-implied slopes for the implied volatility term structure and the volatility skew for all dates in our sample and contrasts them with corresponding values obtained from actual option prices. The first row demonstrates, strikingly, how the one-factor jump diffusions fail to generate the diverse combinations of volatility term structure and smirk slopes observed in the data. For both models, the points are clustered around a steep negatively sloped line. This contrasts dramatically with the combinations obtained from the options which are much more widely dispersed. Specifically, the term structure of implied volatility can be sharply upward or downward sloping when the implied volatility skew is steep. The one-factor models simply cannot generate this feature. The second row shows that the two-factor model provides a major step
forward in this respect. However, the scatter plot is too asymmetric relative to the actual data. In particular, model (??) fails to generate scenarios with a steep term structure and relatively low implied volatility skew. This feature is accommodated by model (3.1), illustrating the added flexibility afforded by the new, partially independent, jump factor.

Figure 4: Fit to Implied Volatility surface characteristics by the three-factor model. The definitions of characteristics as for Figure 1. The dark line corresponds to the options data while the light-colored line to the three-factor model (3.1).

5.3 Robustness of Estimation Results and Inference

We next explore the robustness of our empirical findings in a number of directions. First, as indicated previously, our three-factor model (3.1) has been estimated already in Andersen et al. (2013), but our weighting of the criterion function in the current implementation should improve the efficiency and finite sample robustness of the inference. We already noted in Section ?? that the qualitative features of the extracted factors over the sample are similar across the alternative estimation procedures. The most striking difference in point estimates for the parameters concerns the second volatility factor, $V_2$. The estimates obtained via the unweighted criterion in Andersen et al.
Figure 5: **Fit to Implied Volatility surface characteristics by the two-factor Gaussian model.** The definitions of characteristics as for Figure 1. The dark line corresponds to the options data while the light line to the two-factor Gaussian model (?).
Figure 6: Effect of $U_t$ on the implied Volatility surface characteristics in the three-factor exponential jump model. The definitions of characteristics as for Figure 1. The dark line corresponds to the three-factor model (3.1), while the light-colored line represent the same model but without the contribution of $U_t$. 
(2013) imply that this factor is substantially larger, more persistent, has more volatile innovations and displays less negative correlation with the returns than is the case for the present estimates. This discrepancy is likely to impact the quality of fit. In fact, the formal diagnostics, reported in the Supplementary Appendix, corroborate this conjecture. The unweighted RMSE criterion, by construction, minimizes the RMSE, which equals 1.61% in Andersen et al. (2013) compared to 1.75% in the current implementation. However, the overall fit to the option surface is clearly superior for the current estimates. The pricing violations at the 1% significance level across our six separate option categories are uniformly higher for the estimates obtained via the unweighted criterion function, ranging from 18% to 37% in Andersen et al. (2013), than reported here, where they fall between 4% and 15%. This uniform improvement is also evident visually when comparing Figure ?? to the corresponding Z-score statistics in Andersen et al. (2013). This suggests that outliers are unduly influential in the inference based on the unweighted objective function relative to the weighted RMSE criterion used in the current work.

Second, one may wonder if the improved performance simply reflects the advantages afforded by a three-factor model relative to less richly parameterized one- and two-factor specifications. Thus, we compare the current model to estimation results obtained from a more traditional representation where the third factor is also a volatility factor in line with those in model (3.1). The exact model specification, parameter estimates and diagnostic tests are presented in the Supplementary Appendix. The overall fit is only slightly worse than for the present model – with a RMSE of 1.86% versus 1.75%. However, as above, the model fails to provide a consistent fit to the option surface across time. Although the model is estimated by the identical weighted criterion function, the pricing violations at the 1% level for the six different option categories range from 13% to 45% versus the 4%-15% span for model (3.1). Moreover, every single option category has a significantly lower number of violations across the sample for the current model. Specifically, for the short maturities, going from OTM puts through ATM options to OTM calls, the violations are 21% versus 8%, 13% versus 4%, and 35% versus 15%. Moreover, the relative fit is even worse for the more traditional model at long maturities. Hence, the introduction of a partially independent (negative) jump factor is critical for accommodating the shifting shape of the option surface. In other words, the option surface dynamics provides rich information that can be used to separate jump risks from volatility risks, but the identification of these distinct features hinges crucially on a suitable modeling framework that enables the jumps to exert a unique and distinct impact on the model-implied surface relative to the volatility factors.

Third, the Supplementary Appendix reports results for subsample estimation covering the pe-
periods 1996-2006 and 2007-2010. We find the fit to be excellent for the long 1996-2006 subsample, with an overall RMSE of 1.05%, and with violations at the 1% level for three of the six option categories to be around 1%, while the remaining rates range from 8% to 18.5%, with the hardest category to fit being the long maturity OTM put options. While this sample excludes the financial crisis and thus poses less of a challenge to the option pricing model, the sample still covers some dramatic episodes around 1997, 1998 and the internet bubble, along with some extremely low volatility levels during 2004-2006. Obviously, the shorter and extremely turbulent 2007-2010 period is much harder to accommodate, and the RMSE is now 2.36%, with violations of the fit to the six option categories ranging from 16% to 61% and the short maturity OTM call options being especially hard to price appropriately. Nonetheless, the similarities in the implied jump processes across the two subsamples and the full sample results for model (3.1) reflects a remarkable stability in the inference. The distinct decay parameters for the left and right tails are quite close and the mean jump sizes are between -4.5% and -5% for the negative tail and around 2% for the positive tail in both subsamples. Moreover, both imply a half life for volatility shocks of about 5-6 weeks for $V_1$ and about one year for $V_2$, which aligns well with the implied quantities for the full sample. Thus, the critical dynamic features of the estimated model are remarkably stable over time.

6 Out-of-Sample Performance

In order to test the robustness of the results obtained so far, we use the parameters estimated for each model over the period from January 1st 1996 to July 21st 2010 to price options over the subsequent period from July 22nd 2010 to April 23rd 2013. The average implied volatility (across all options) over the period is equal to 24.17% while the average at-the-money implied volatility over the considered period is equal to 18.40.

In Table 6 we report the RMSE in implied volatility for each model. $RMSE_{IV}$ is defined as:

$$RMSE_{IV} = \sqrt{\frac{\sum_{t=1}^{T} \sum_{j=1}^{N_t} (\tilde{\kappa}_{j,k_j,\tau_j} - \kappa(k_j, \tau_j, Z_t, \theta))^2}{\sum_{t=1}^{T} N_t}}.$$  

\footnotesize
\textsuperscript{14} The sharp increase in the average number of options in the out-of-sample analysis with respect to our in-sample estimation is for two reasons. On the one hand we observe an ever increasing number of options quoted for a given options’ maturity -thus filling the moneyness dimension. At the same time, the introduction of weekly options increases the number of times-to-maturity available on each day -thus filling the maturity dimension. Weekly options were introduced by the CBOE in 2005, but only from 2011 their volume has become a significant (almost 20%) fraction of all the SPX options contracts.

\textsuperscript{15} As a reference, in the first part of the sample, between January 1996 and December 1998, we have an average of 158 contracts per day, while in the last part of the sample, between January 2008 to July 2010, we have an average of 490 contracts per day.
We conclude that the novel three-factor model (3.1) provides a significant improvement over the traditional specification of the risk-neutral equity return dynamics driving the evolution of the option surface. Furthermore, the more critical features of the system seem to be quite robustly estimated across highly diverse market conditions. As such, the system should, comparatively, provide a sound basis for the exploration of the equity and volatility risk premiums which we turn to in the remainder of the paper.

7 Risk Premia Dynamics and Predictability

In this section we relate our findings based on the option panel to the underlying return data. In particular, we study the links between the state vector, or risk factors, extracted from the option panel with the volatility and jump risks inferred from the dynamics of the underlying stock market index. This sets the stage for a direct exploration of the associated equity and variance risk premiums and their relation with the option-implied factors. Given the superior performance, we rely on the three factor model (3.1) for tracking the dynamics of the option surface over time. We continue to avoid making strong assumptions regarding the dynamics of the market risks under the actual probability measure. Consequently, this part of our analysis is fully nonparametric, and we invoke only minimal stationarity conditions regarding the \( P \)-dynamics.

7.1 Connecting the Information in the Option Panel and the Underlying Asset

To define risk premia, we must first develop consistent notation concerning the pricing of each of the sources of risks in the model (3.1). We define

\[
W_{1,t}^P = W_{1,t}^Q - \int_0^t \lambda_s W_1^s ds, \quad W_{2,t}^P = W_{2,t}^Q - \int_0^t \lambda_s W_2^s ds, \quad B_{1,t}^P = B_{1,t}^Q - \int_0^t \lambda_s B_1^s ds, \quad B_{2,t}^P = B_{2,t}^Q - \int_0^t \lambda_s B_2^s ds,
\]

(7.1)

where \( W_{1,t}^P, W_{2,t}^P, B_{1,t}^P \) and \( B_{2,t}^P \) are \( P \) Brownian motions and \( \lambda_t^{W_1}, \lambda_t^{W_2}, \lambda_t^{B_1} \) and \( \lambda_t^{B_2} \) denote the associated prices of risk. The compensator of the jump measure, \( \mu \), under the \( P \) measure is given by \( dt \otimes \nu_t^P(dx,dy) \), and the mapping \( \nu_t^P(dx,dy) \to \nu_t^Q(dx,dy) \), defined for every jump size \( x \) and \( y \) and every point in time \( t \), reflects the compensation for jump risk.

The dynamics of the stock price process under the physical probability measure \( P \) is then,

\[
\frac{dX_t}{X_t} = \alpha_t dt + \sqrt{V_{1,t}} dW_{1,t}^P + \sqrt{V_{2,t}} dW_{2,t}^P + \int_{\mathbb{R}^2} (e^x - 1) \tilde{\nu}_t^P(dt, dx, dy),
\]

(7.2)

where

\[
\alpha_t - (r_t - \delta_t) = \lambda_t^{W_1} \sqrt{V_{1,t}} + \lambda_t^{W_2} \sqrt{V_{2,t}} + \int_{\mathbb{R}^2} (e^x - 1) \nu_t^P(dx, dy) - \int_{\mathbb{R}^2} (e^x - 1) \nu_t^Q(dx, dy).
\]

(7.3)
is the instantaneous equity risk premium, reflecting compensation for diffusive and (price) jump risks.

Application of Itô formula yields the following representation for \( \log(X_t) \) under \( \mathbb{P} \),

\[
d \log(X_t) = \left[ \alpha_t - q_t^\mathbb{P} \right] dt + \sqrt{V_{1,t}} dW_{1,t}^\mathbb{P} + \sqrt{V_{2,t}} dW_{2,t}^\mathbb{P} + \int_{\mathbb{R}^2} x \mu^{\mathbb{P}}(dt, dx, dy),
\]

where \( q_t^\mathbb{P} = \frac{1}{2} V_t + \int_{\mathbb{R}^2} (e^x - 1 - x) \nu_t^\mathbb{P} (dx, dy) \), and similarly under \( \mathbb{Q} \) with \( \alpha_t \) replaced by \( r_t - \delta_t \) and all superscripts \( \mathbb{P} \) replaced with \( \mathbb{Q} \) in the expression above.

We may then define the (conditional) cum-dividend equity risk premium over the horizon \( \tau \),

\[
\text{ERP}_t^\tau \equiv \frac{1}{\tau} \left[ \mathbb{E}_t^\mathbb{P} \left( \log \left( \frac{X_{t+\tau}}{X_t} \right) + \int_t^{t+\tau} (\delta_s + q_s^\mathbb{P}) ds \right) - \mathbb{E}_t^\mathbb{Q} \left( \log \left( \frac{X_{t+\tau}}{X_t} \right) + \int_t^{t+\tau} (\delta_s + q_s^\mathbb{Q}) ds \right) \right]
- \frac{1}{\tau} \left[ \mathbb{E}_t^\mathbb{P} \left( \int_t^{t+\tau} r_s ds \right) - \mathbb{E}_t^\mathbb{Q} \left( \int_t^{t+\tau} r_s ds \right) \right] \quad (7.5)
\]

Notice that since a long position in the market index involves a commitment of capital, a part of the wedge in the \( \mathbb{P} \) and \( \mathbb{Q} \) expectations of the cum-dividend equity returns reflects compensation for the time-variation in the risk-free interest rate.\(^{16}\)

We next define corresponding risk measures for the return variation. The most natural and popular one is the quadratic variation over \([t, t+\tau]\) which we denote \( QV_{t,t+\tau} \). This captures the return variation over the given horizon and is given by,

\[
QV_{t,t+\tau} = QV_{t,t+\tau}^c + QV_{t,t+\tau}^j,
\]

\[
QV_{t,t+\tau}^c = \int_t^{t+\tau} (V_{1,s} + V_{2,s}) ds, \quad QV_{t,t+\tau}^j = \int_t^{t+\tau} \int_{\mathbb{R}^2} x^2 \mu(ds, dx, dy), \quad (7.6)
\]

where we decompose the return variation into terms generated by the continuous and jump component of \( X \), \( QV_{t,t+\tau}^c \) and \( QV_{t,t+\tau}^j \). Further, note that the (realized) quadratic variation is independent of the probability measure. The variance risk premium is defined as,

\[
\text{VRP}_t^\tau \equiv \frac{1}{\tau} \left[ \mathbb{E}_t^\mathbb{P} \left( QV_{t,t+\tau} \right) - \mathbb{E}_t^\mathbb{Q} \left( QV_{t,t+\tau} \right) \right], \quad (7.7)
\]

which provides compensation for the variance risk associated with both the continuous and the jump component of \( X \).

We are also interested in assessing directly the risks and risk premiums associated with jumps. In particular, we want to gauge the compensation for large price jumps and to allow for a separate

\(^{16}\)This term is, of course, absent if we instead define the equity risk premium using futures on the market index.
risk premium for the negative versus positive jumps. We obtain direct measures of the jump risks by simply counting the number of “big” jumps over the relevant horizon,

\[ LT_{t,t+\tau}^K = \int_t^{t+\tau} \int_{\mathbb{R}^2} 1_{\{x \leq -K\}} \mu(ds, dx, dy), \quad RT_{t,t+\tau}^K = \int_t^{t+\tau} \int_{\mathbb{R}^2} 1_{\{x \geq K\}} \mu(ds, dx, dy), \]  

(7.8)

where \( K \) is a prespecified threshold. We set \( K = 0.5\% \) in the subsequent analysis.\(^{17}\) From the properties of the compensator for a jump measure, we have,

\[ \begin{align*}
LT_{t,t+\tau}^K &= \int_t^{t+\tau} \int_{\mathbb{R}^2} 1_{\{x \leq -K\}} \nu^p_s(dx, dy) ds + \epsilon^L_{t,t+\tau}, & \mathbb{E}^p(\epsilon^L_{t,t+\tau}) = 0, \\
RT_{t,t+\tau}^K &= \int_t^{t+\tau} \int_{\mathbb{R}^2} 1_{\{x \geq K\}} \nu^p_s(dx, dy) ds + \epsilon^R_{t,t+\tau}, & \mathbb{E}^p(\epsilon^R_{t,t+\tau}) = 0.
\end{align*} \]

(7.9)

Hence, up to martingale difference sequences, \( LT_{t,t+\tau}^K \) and \( RT_{t,t+\tau}^K \) measure the \( \mathbb{P} \) jump intensity of “large” jumps.

The risk measures we have introduced, including \( QV_{t,t+\tau}^c, QV_{t,t+\tau}^j, LT_{t,t+\tau}^K \) and \( RT_{t,t+\tau}^K \), are not directly observable, but we can estimate them using high-frequency futures data. Naturally, since intraday data are available only during active trading, our high-frequency measures pertain exclusively to the trading days within \([t, t + \tau]\). We denote the latter with superscript \( i \) (for intraday based measures), e.g., \( QV_{t,t+\tau}^i = \sum_{s=t+1}^{t+\tau} QV_{s+s,i+1} \). \( QV_{t,t+\tau}^{c,i}, QV_{t,t+\tau}^{j,i}, LT_{t,t+\tau}^{K,i} \) and \( RT_{t,t+\tau}^{K,i} \) are defined analogously.

We then introduce the following empirical measures,

\[ \begin{align*}
\hat{LT}_{t,t+\tau}^{K,i} &= \sum_{s=t+1}^{t+\tau} \sum_{i=1}^{n} 1_{\{\Delta_n^{c,i} f < -K_i\}}, \\
\hat{RT}_{t,t+\tau}^{K,i} &= \sum_{s=t+1}^{t+\tau} \sum_{i=1}^{n} 1_{\{\Delta_n^{c,i} f > K_i\}}, \\
\hat{QV}_{t,t+\tau}^{i} &= \sum_{s=t+1}^{t+\tau} \sum_{i=1}^{n} |\Delta^{c,i}_n f|^2 1_{\{\Delta_n^{c,i} f \leq \alpha_i\}}, \quad \hat{QV}_{t,t+\tau}^{c,i} = \sum_{s=t+1}^{t+\tau} \sum_{i=1}^{n} |\Delta^{c,i}_n f|^2 1_{\{\Delta_n^{c,i} f \leq \alpha_i\}}.
\end{align*} \]

(7.10)

(7.11)

where the time-varying jump thresholds \( K_i \) and \( \alpha_i \) are chosen according to the procedures outlined in the supplementary appendix.\(^{18}\) Under general conditions, see, e.g., Jacod (2008),

\[ \begin{align*}
\hat{LT}_{t,t+\tau}^{K,i} \xrightarrow{p} LT_{t,t+\tau}^{K,i}, & \quad \hat{RT}_{t,t+\tau}^{K,i} \xrightarrow{p} RT_{t,t+\tau}^{K,i}, \quad \hat{QV}_{t,t+\tau}^{i} \xrightarrow{p} QV_{t,t+\tau}^{i}, \quad \hat{QV}_{t,t+\tau}^{c,i} \xrightarrow{p} QV_{t,t+\tau}^{c,i}.
\end{align*} \]

(7.12)

\(^{17}\)This (fixed) threshold \( K \) is large enough that we can separate returns exceeding this (absolute) level from diffusive volatility using 1-minute observations. Experiments with alternative cutoffs produced similar results.

\(^{18}\)The variation in the threshold \( \alpha_i \) ensures that we can separate jumps from large diffusive moves asymptotically. \( K_i \) is defined as the maximum of \( K \) and \( \alpha_i \), so that \( K_i = K \) except when volatility is unusually high, in which case we adopt \( K_i = \alpha_i \) to guard against jump misclassification.
For the overnight periods within \([t, t + \tau]\), we cannot separate the diffusive volatility from jumps, and we simply estimate the total (realized) overnight variance via,

\[
\hat{QV}_{t,t+\tau}^o = \sum_{s=t+1}^{t+\tau} (f_{s-1+\tau} - f_{s-1})^2. \tag{7.13}
\]

Our estimate for the total variation, \(QV_{t,t+\tau}\), is then given by \(\hat{QV}_{t,t+\tau} = \hat{QV}_{t,t+\tau}^i + \hat{QV}_{t,t+\tau}^o\).

Similarly, we estimate the equity and variance risk premiums using the relations,

\[
\log \left( \frac{X_{t+\tau}}{X_t} \right) - \frac{1}{\tau} \int_t^{t+\tau} (r_s - \delta_s) ds = ERP_t + \frac{1}{\tau} E_t^p \left( \int_t^{t+\tau} q_s^p ds \right) + \epsilon_t^E, \quad E_t^P (\epsilon_t^E) = 0,
\]

\[
\hat{VRP}_t = \frac{1}{\tau} \left[ \hat{QV}_{t,t+\tau} - E_t^P (QV_{t,t+\tau}) \right] = VRP_t + \hat{VRP}_t, \quad E_t (\hat{VRP}_t) = 0, \tag{7.14}
\]

where \(E_t^P (QV_{t,t+\tau})\) can be measured in model-free fashion via the VIX index computed by the CBOE from a portfolio of S&P 500 index options. Equation (7.14) shows that a martingale difference sequence separates \(\hat{VRP}_t\) from \(VRP_t\), and, likewise, a martingale difference sequence separates the log excess cum-dividend returns on the underlying asset from the unobservable ERP rate, i.e.,

\[
\frac{1}{\tau} \int_t^{t+\tau} q_s^p ds.
\]

In principle, we can remove the term stemming from the convexity adjustment, i.e., \(\frac{1}{\tau} E_t^P \left( \int_t^{t+\tau} q_s^p ds \right)\), via a consistent estimator for \(\int_t^{t+\tau} q_s^p ds\) (again up to a martingale difference term) obtained from high-frequency data, e.g.,

\[
\int_t^{t+\tau} q_s^p ds = \sum_{i \in (t,t+\tau)} \left[ \frac{1}{2} |\Delta^n f|^2 1_{\{|\Delta^n f| \leq \alpha_t\}} + (e^{\Delta^n f} - 1 - \Delta^n f) 1_{\{|\Delta^n f| > \alpha_t\}} \right].
\]

In practice, this adjustment is minute and the results are virtually unchanged if we implement it.\(^{20}\)

### 7.2 The Predictability of Equity and Variance Risk and Risk Premia

Equipped with empirically feasible estimators for the relevant risk measures and risk premia, we now explore the relationship between the option-implied factors, \(V_1, V_2\) and \(U\), that drive the dynamics of the option surface and the various risk measures and risk premia associated with the underlying asset. We rely on alternative versions of the following predictive regression,

\[
y_t = \alpha_0 + \alpha_1 V_{1,t} + \alpha_2 V_{2,t} + \alpha_3 U_t + \epsilon_t, \tag{7.15}
\]

where the left hand side represents, in turn, jump and diffusive variance risk measures as well as risk premia, i.e., \(y_t = \hat{E}T_{t,t+\tau}^{K,\ell}, \hat{RT}_{t,t+\tau}^{K,\ell}, \hat{QV}_{t,t+\tau}^{c,i}, \hat{QV}_{t,t+\tau}^{i,}, \log \left( \frac{X_{t+\tau}}{X_t} \right) - \frac{1}{\tau} \int_t^{t+\tau} (r_s - \delta_s) ds\) and

\(^{19}\)For the overnight periods we just take one half of the squared return.

\(^{20}\)We do not report results adjusting for this term to conserve space. They are available upon request.
Given the relationships explicated in equations (7.9), (7.10) and (7.14), it is evident that the regressions based on the alternative \( y_t \) variables, asymptotically, yield estimates identical to those based on the corresponding infeasible measures of interest, i.e.,

\[
\sum_{s=t+1}^{t+\tau} \int_{s+\tau}^{s+1} 1_{\{x \leq -K\}} \nu_s^p(dx, dy),
\sum_{s=t+1}^{t+\tau} \int_{s+\tau}^{s+1} 1_{\{x \geq K\}} \nu_s^p(dx, dy),
QV_{t,t+\tau}^c, QV_{t,t+\tau}, ERP_t^\tau \text{ and } VRP_t^\tau .
\]

Thus, the predictive regression in equation (7.15) speaks directly to the linkages between the option surface dynamics and the risks and risk premia associated with the equity-index market.

In general, if the premia for the diffusive and jump risks are spanned by the factors \( V_1, V_2 \) and \( U_t \), then the expectation of \( y_t \) conditional on time \( t \) information will be functionally related to \( V_{1,t}, V_{2,t} \) and \( U_t \). Moreover, in the standard case, almost universally adopted in option pricing applications, the measure change preserves the affine structure, so the conditional mean of \( y_t \) is linear in \( V_{1,t}, V_{2,t} \) and \( U_t \). Hence, the regression in (7.15) produces optimal (mean-square error) predictors for the volatility and jump risks at time \( t \). Furthermore, conceptually, our extraction of \( V_{1,t}, V_{2,t} \) and \( U_t \) provides a richer information set for forecasting the volatility and jump realizations than the history of underlying asset returns. The latter, at best, generates estimates of the path for \( \{V_{1,s} + V_{2,s}\}_{s \leq t} \) as well as associated jump variation measures.\(^{21}\)

We summarize the results from the predictive regressions on Figures 7 and 8. Since we are particularly interested in the incremental role of the novel factor \( U \), we initially project \( U \) linearly onto the two volatility factors and denote the residual by \( \hat{U} \). Consequently, \( \hat{U} \) reflects the features of the system not associated with the traditional volatility factors. Given our relatively short sample, we compute the predictive regressions for horizons up to one year only. Figure 7 shows that the state variables, extracted from the option panel, have significant explanatory power for the future evolution of risks. In particular, the plot pertaining to the count of positive and negative jumps documents that the jump intensities display highly predictable time-variation under the statistical measure, \( \mathbb{P} \), i.e., \( \nu_t^p(dx, dy) \) is truly a function of \( t \).\(^{22}\) Moreover, the state variables differ greatly in terms of their ability to forecast the future volatility and jump intensity. Specifically, once we control for the volatility factors, \( \hat{U} \) provides no incremental explanatory power. This is evident both from the insignificant \( t \)-statistics corresponding to \( \hat{U} \) as well as the trivial drop in \( R^2 \) when we exclude \( \hat{U} \) from the regressions.\(^{23}\) Hence, Figure 7 is consistent with a model for which the jump

\(^{21}\) However, it is, of course, essential that we initially “correct” the option data for the relevant risk premia through the estimation of the parametric model (3.1).

\(^{22}\) For further nonparametric evidence of time variation in the \( \mathbb{P} \)-jump intensity, obtained solely from high-frequency futures data, see Bollerslev and Todorov (2011a).

\(^{23}\) This is consistent with Andersen and Bondarenko (2011) who find that moderate-to-deep OTM put options have no predictive power for the future volatility of the underlying equity-index returns once they control for the information provided by ATM and OTM call options.
intensity, under $\mathbb{P}$, depends only on the two volatility factors $V_1$ and $V_2$, but not $\tilde{U}$.\footnote{In general, since $V_1$ contains jumps of time-varying intensity that load on all the factors, $V_1$, $V_2$ and $U$, the conditional forecast of future volatility $V_{1,t} + V_{2,t}$ still depends (critically) on the current value of the third factor $U$.}

We now turn to the predictability of the equity and variance risk premia. Figure 8 indicates that our three factors, extracted from the option panel, now take on very different roles. In particular, a significant part of the predictability of both the equity and variance risk premia is due to the $U$ factor. For the variance risk premia, $V_1$ also contributes strongly at shorter horizons, while, importantly, for the equity risk premium $\tilde{U}$ is the single dominant explanatory factor for all horizons. The importance of $U$ for predicting both the equity and variance risk premia is consistent with Bollerslev and Todorov (2011b) who find the equity and variance risk premia to
Figure 8: Predictive regressions for equity and variance risk premia. For each regression, the top panels depict the t-statistics for the individual parameter estimates while the bottom panels indicate the regression $R^2$. The predictive variables are $V_1$ (dashed-dotted line), $V_2$ (dashed line) and $\tilde{U}$ (solid line), where $\tilde{U}$ is the residual from the linear projection of $U$ on $V_1$ and $V_2$. The regression standard errors are constructed to also account for the estimation error in the projection generating $\tilde{U}$. The dashed lines in the $R^2$ plots correspond to constrained regressions, including only $V_1$ and $V_2$.

embed a common component stemming from compensation of left jump tail risk. The common dependence of the equity and variance risk premia on $U$, coupled with the significant persistence of the latter, help rationalize the predictive power of the variance risk premium for future excess returns, documented in Bollerslev et al. (2009) and Drechsler and Yaron (2011). The limited role for the two volatility factors mirrors the conclusions of many prior studies on the risk-return tradeoff, going back to, e.g., French et al. (1987) and Glosten et al. (1993). Finally, we note that our results are not driven by the events surrounding the financial crisis. Qualitatively identical results apply for subsamples that end prior to 2007.\textsuperscript{25}

Another way to gauge the importance of $U$, and proper model specification in general, for predicting the equity and variance risk premia is to contrast the above evidence to findings generated by standard two-factor jump-diffusive models. Since the results are entirely consistent with our prior conclusions, we briefly summarize the evidence and refer to the supplementary appendix for details. If we omit $U$ from the model, we are still left with an elaborate model which, besides the two volatility state variables driving the dynamics of the option surface, also incorporates both diffusive and jump leverage effects, co-jumps in returns and volatility, and separate decay rates for the left and right (exponential) jump tails.\textsuperscript{26} We find that the forecast performance regarding the future evolution of volatility and jump risks is similar to the corresponding results in Figure 7. Given that $\tilde{U}$ plays only a minimal role in forecasting these quantities within model (3.1), this is quite

\textsuperscript{25}The documentation of these results are available upon request.
\textsuperscript{26}The corresponding two-factor model with Gaussian jumps performs less well along all dimensions.

34
intuitive. Moreover, as before, when considering the equity risk premium, the volatility factors are largely insignificant (and as likely to be negative as positive), and the $R^2$ of the predictive regression for the future excess returns is dramatically reduced. In short, the evidence for predictability of the equity risk premium vanishes when the option surface dynamics is modeled in the common jump-diffusive framework and driven exclusively by volatility factors. Hence, the inclusion of the $U$ factor, allowing for the left risk-neutral tail to have a separate source of variation, is pivotal for capturing the predictability in the equity risk premium. Finally, for the variance risk premium, both volatility factors are significant in the two-factor model, but the $R^2$ of the variance risk premium regression is notably lower than for our three-factor model (3.1).

These findings have implications for the ability of existing structural economic models to rationalize the predictability of the equity and variance risk premia which we discuss next.

![Predictive regressions for volatility and jump risks in the two-factor model (??).](image)

Figure 9: Predictive regressions for volatility and jump risks in the two-factor model (??). The predictive variables are $V_1$ (dashed-dotted line) and $V_2$ (dashed line). The remainder of the notation follows the conventions in Figure 7.

### 7.3 Structural Implications of the Predictive Power of the Option Surface

Figures 7 and 8 demonstrate that the factor $U$, driving a substantial part of the OTM short maturity put option dynamics, has no impact on the actual volatility and jump dynamics of the underlying asset. In contrast, the factor has a critical effect on the pricing of volatility and jump risk. In other words, it resembles a risk premium and not a risk factor. Can we rationalize this finding from an
economic perspective? To guide intuition, we compare the findings in Figures 7 and 8 with those from recent structural models that link option prices to fundamental macroeconomic risks, such as aggregate consumption and dividends. This emerging literature has made important progress in tackling the challenging, yet critical, task of jointly explaining the equity return and risk premium dynamics in a coherent general equilibrium setting. For concreteness, we initially explore a couple of specific models featuring a representative agent with Epstein-Zin preferences exposed to risks in real consumption growth, namely Wachter (2013) and Drechsler and Yaron (2011).

In the one-factor model of Wachter (2013), the consumption growth is subject to infrequent, but large, negative jumps (rare disasters) with a time-varying arrival rate, resembling the mechanism in Gabaix (2012). This type of equilibrium model can account for many critical empirical features such as the correlation between volatility and jump risks, the time-varying jump arrival as well as the ability of the market variance risk premium to predict future equity excess returns.

The model of Drechsler and Yaron (2011) specifies consumption growth as conditionally Gaussian with a time-varying conditional mean (long-run risk) and conditional volatility. This model builds on Eraker and Shaliastovich (2008) and generalizes many prior models in which consumption growth contains a small predictable persistent component, including the original work of Bansal and Yaron (2004).

---

27 These papers generalize work by Barro (2006) and Barro and Ursua (2008) in which rare disasters are i.i.d.

28 This model builds on Eraker and Shaliastovich (2008) and generalizes many prior models in which consumption growth contains a small predictable persistent component, including the original work of Bansal and Yaron (2004).
Figures 11 and 12 summarize the findings from predictive regressions for future volatility and jump risks as well as equity and variance risk premia in the models of Wachter (2013) and Drechsler and Yaron (2011), using the concurrent level of the relevant state variables in each model as predictors. Not surprisingly, given the one-factor structure, the Wachter (2013) model faces some challenges in accommodating the evidence laid out in Figures 7 and 8. Importantly for our analysis, the predictability of future excess returns and variance risk premia is linked closely to the predictability of the future return variation, and the underlying pattern of significance is identical in all cases (flat line), and the degree of explanatory power rises roughly linearly with maturity. Compared to Figure 7, the predictability in the Wachter (2013) model is inverted, as the return variation is forecast with relatively higher precision over long rather than short horizons. Furthermore, the explanatory power is uniformly too low. Likewise, referencing Figure 8, the model fails to capture the degree and pattern of predictability in the excess returns and variance risk premium.\footnote{A two-factor extension, like in Seo and Wachter (2013), in which rare disaster probability is driven by two factors, can potentially generate dynamic patterns more consistent with the data. However, as we discuss below in more general terms, such an extension will still produce a close link between the predictability of future risks and risk premia unlike what we find in the data.}

Turning to Figure 12, it is evident that the more flexible volatility structure of Drechsler and Yaron (2011) is useful in accommodating some of the stylized features of the data. Nonetheless, it is equally clear that the long-run risk factor helps predict neither the future volatility and jump risks nor the equity and variance risk premia. In this structural setting, essentially all predictability stems from the two volatility factors. They provide the channel through which past variance risk premia generate predictable movements in the equity risk premium. Thus, relative to our empirical finding, captured by Figures 7 and 8, this structural model also ties the predictability of future volatility and jump risks too closely to the predictability of the equity and variance risk premia. Equivalently, the structural model implies a tight relationship between the dynamics of the option panel and the return dynamics of the underlying equity market. In contrast, our empirical results based on model (3.1) – nesting all standard two-factor affine volatility specifications – document a partial, and critical, decoupling between the factors driving the equity return dynamics and those governing the pricing of risk, and thus the equity and variance risk premia.

There is a fundamental reason for the discrepancy between our empirical findings and the

\footnote{One of the state variables that we label volatility factors directly controls the conditional variance of consumption growth, while the other captures the variation in the long run variance of consumption growth.}
implications of the structural models of Wachter (2013), and its extension in Seo and Wachter (2013), as well as Drechsler and Yaron (2011). Although the models generate risk premia through different channels – the presence of rare disasters and uncertainty about their arrival (Wachter (2013) and Seo and Wachter (2013)) versus long-run risk and stochastic volatility in consumption growth (Drechsler and Yaron (2011)) – they share a critical feature in the pricing of jump tail risk. They both imply that the ratio $\frac{\nu_Q(dx,dy)}{\nu_P(dx,dy)}$ is time-invariant. That is, the risk-neutral jump intensity is proportional to that under the actual probability measure, so the two jump intensities are equivalent in terms of their time variation. Therefore, the jump risk premia are generated by changing the distribution of the jump size only. This implies that the variation in the jump intensity is “inherited” under the equivalent change of measure. Consequently, these equilibrium based pricing kernels cannot “generate” new state variables in addition to those that drive the fundamental risks in the economy. In turn, this necessarily generates the tight link between the dynamics of the underlying asset and the option surface within these models.

The enhanced flexibility in capturing joint movements in the volatility term structure and skew, attained through the inclusion of the $U$ factor, suggests that a key attribute of model (3.1) is the decoupling of (negative) jump risk from volatility risk. One direct way to assess this feature is to compute the fraction of the overall model-implied conditionally expected (risk-neutral) return..
Figure 12: Predictive regressions implied by the Drechsler and Yaron (2011) structural model. The predictive variables are the conditional mean of consumption growth (solid line), stochastic volatility of consumption growth (dashed-dotted line) and central tendency of stochastic volatility (dashed line). The dashed lines in the $R^2$ plots correspond to constrained regressions including only the volatility state variables.

variation that is attributable to negative jumps over the sample. Obviously, as we progress from the one- to the two-factor models, we loosen the links between these risk factors, but only the introduction of $U$ allows for a significant degree of independent jump variation. Thus, if the realized option surface dynamics can be better reconciled through this type of mechanism, we would expect to see a significant degree of discrepancy across the models along this dimension. Figure 13 provides this comparison. The implied (negative) jump variation is, indeed, dramatically distinct in our three-factor model. While the fraction of the quadratic return variation accounted for by negative jumps is around 30% for all models, it is essentially flat for the one-factor Gaussian jump model, it displays only very minor fluctuations in the one- and two-factor exponential jump models, but it varies significantly, albeit quite smoothly, from under 10% to over 50% for the flexible three-factor model. In particular, the negative jump variation is elevated in the latter model from the onset of the Asian crisis in 1997 up to year 2000, and then again in the aftermath of the financial crisis in 2009 through the end of the sample. As noted before, the initial phase of the financial crisis is not characterized by any large movement in the jump risk relative to volatility risk – both are extremely elevated at that point. The relative importance of the jump risk manifests itself clearly
only after volatility starts subsiding. Likewise, the events surrounding the dissolution of LTCM in 1998 and the European sovereign debt crisis in 2010 feature dramatic relative increases in the downside tail risk.

In fact, this tight linkage of the physical and risk-neutral jump intensities is operative for a wide class of popular structural models with representative agents having Epstein-Zin preferences. The part of the density $dQ/dP$ due to the change of the jump measure is characterized by,

$$
E \left( \int_0^t \int_{\mathbb{R}^n} (Y(\omega, s, x) - 1)\tilde{\mu}^P(ds, dx) \right),
$$

(7.16)

where the jump component of the state vector in the economy under the $P$ probability measure is given as a (multivariate) integral of the form $\int_0^t \int_{\mathbb{R}^n} x \tilde{\mu}^P(ds, dx)$; $Y(\omega, t, x) = \frac{\nu_Q(dx)}{\nu_P(dx)}$ is the measure change for the jump intensity, and $E$ is the Doleans-Dade exponential.\footnote{For the expression in (7.16) and the definition of the Doleans-Dade exponential, see Jacod and Shiryaev (2003), Corollary III.5.22 and I.4.59, respectively.} In general, $Y(\omega, t, x)$ will be stochastic. However, within an equilibrium setting, stipulating an affine dynamics for the fundamentals along with a representative agent with Epstein-Zin preferences generates the

---

**Figure 13:** Ratio of model-implied negative jump variation to total return variation. Alternative model-implied ratios of negative jump return variation versus total quadratic return variation across the full sample.
restriction that the expression in equation (7.16) is exponentially affine in the state vector, see, e.g., equation (2.22) of Eraker and Shaliastovich (2008). In turn, this implies that $Y(\omega, s, x)$ must be non-random and time-invariant, i.e., depend only on $x$. Hence, the equilibrium implied statistical and risk-neutral intensities of the price and volatility jumps (which are mixtures of the jumps in the state variables driving the fundamentals in the equilibrium model) will be affine functions of the same state vector. In contrast, for our extended three-factor model, in which $\nu_P^T(\cdot)$ loads, at most, very marginally on $U_{t-}$, there is a natural wedge between the time variation in statistical and risk-neutral jump intensities.

There are several ways in which the link between the asset and option price dynamics may be relaxed within an equilibrium setting to potentially account for our empirical evidence. They all involve generalizing the preferences of the representative agent in some form. One approach is to allow the representative agent’s coefficient of risk aversion to vary over time. Du (2010) proposes a generalization of a habit formation model in which consumption growth is i.i.d. with rare jumps, building on the equilibrium models of Campbell and Cochrane (1999) and Barro (2006). In this model, and consistent with our findings, there is a wedge between the time-variation of the $P$ and $Q$ jump intensities: the former is constant while the latter is time-varying. However, the time variation of the $Q$ jump intensity is driven by the habit formation mechanism and is purely a function of the consumption surplus ratio. The latter, however, determines uniquely the diffusive volatility as well. This is at odds with our findings as the presence of $U$ in the risk-neutral jump intensity drives a wedge between the time variation of the latter and that of the stochastic volatility. Thus, it remains an open question whether a model with external habit formation can decouple the option and asset price dynamics in a manner reminiscent of our empirical findings. It is also unclear whether the frequency and intensity of stock market jumps can be mapped into corresponding jumps in the consumption growth rate as implied by this model of habit formation.

An alternative way to relax the link between the option and asset price dynamics is by recognizing that agents do not directly observe the state vector and therefore need to filter the states from observables. The absence of perfect information plus aversion to ambiguity (particularly about extreme negative risks), like in Hansen and Sargent (2008), or lack of confidence in the estimate of the state vector, like in Bansal and Shaliastovich (2010), can make investors appear more risk

---

32 The restriction on $Y(\omega, s, x)$ is actually stronger. In the equilibrium model, $Y(\omega, s, x)$ is an exponential function of jump size, see Theorem 1 of Eraker and Shaliastovich (2008). Hence, the jump measure change is implemented by exponential tilting (in Laplace transform space), with the degree of tilting determined by the preference parameters of the representative agent.

33 In the models of Wachter (2013), Seo and Wachter (2013) and Drechsler and Yaron (2011), the jump intensities are even more tightly connected, as they are directly proportional.

34 Our empirical findings, unlike the model of Du (2010), suggest strong time variation in the $P$ jump intensity.
averse than they would be in a perfect information setting. The key question is whether the ambiguity aversion or confidence risk variation can generate the required gap between the dynamics of the statistical and risk-neutral jump intensities. For example, the tight link between the jump intensity under $P$ and $Q$ remains within the generalization of the Drechsler and Yaron (2011) model, developed by Drechsler (2013), in which agents are ambiguous about part of the dynamics, including the jumps in the conditional mean and variance of consumption growth. The representative agent’s ambiguity drives an additional wedge between fundamental risks and asset prices and helps explain why variance risk premia have superior predictive power, relative to volatility itself, for future returns. Nonetheless, the linearity of the pricing kernel with respect to the state vector implies that no new state variable is “generated” going from $P$ to $Q$, thus ultimately rendering the model predictions incompatible with our empirical findings.

We finish this section with a brief discussion regarding the robustness of our findings. For generating the predictive regression results in Section 7.2 we relied on a parametric model for the risk-neutral dynamics of $X$ but we were otherwise essentially fully nonparametric about the rest of the features of the model. Thus, the conclusions drawn in this section regarding the structural equilibrium models naturally depend on the correct risk-neutral parametric model. Regarding this, in Section ?? we reported a battery of tests indicating that our risk-neutral model is successful in capturing the dynamics of the option surface. The findings in Section 7.2 underscore the importance of our estimation philosophy of letting the option data to speak for itself and not restricting the pricing kernel, specifically the part concerning the pricing of jump risk. In particular, unlike essentially all earlier empirical parametric option pricing studies we are aware of, which constrain the pricing of jump risk to a change in the jump distribution from $P$ to $Q$ and/or scale its intensity by a constant, we imposed no such restriction on the pricing of jump risk. If we were to constrain the pricing of jump risk to only a change of the jump distribution and/or scaling up the jump intensity by a constant, then by construction the time-variation in the risk-neutral and statistical jump intensities should coincide. Therefore, the cost of imposing such additional structure in the estimation, would be to rule out by construction the possibility of $U$ to play the role of a separate risk premium factor that we find here.

8 Conclusion

We document that the standard exponentially-affine jump-diffusive specifications used in the empirical option pricing literature are incapable of fitting critical features of the option surface dynamics

\footnote{Recall that, according to the estimates for (3.1), $U$ does not covary perfectly with stochastic volatility.}
for the S&P 500 index, especially in scenarios involving significant shifts in the volatility smirk. We extend the risk-neutral volatility model to include a separate state variable which is crucial in capturing the time variation of priced downside tail risk. This new factor has no incremental explanatory power, beyond the traditional volatility factors, for the future evolution of volatility and jump risks. On the other hand, relative to the volatility components, the new factor provides critical, and superior, information for the time variation in the equity and variance risk premia.

Our findings demonstrate that the pricing in the option market is closely integrated with the underlying asset market. Moreover, the option panel embodies critical information regarding equity risk pricing that cannot be extracted directly from the underlying asset price dynamics. The wedge between the two probability measures arises primarily from the varying degree of compensation for downward tail jump risk. Our results suggest that time-varying risk aversion and/or ambiguity aversion, driven in part by the presence of large shocks, must be incorporated into structural asset pricing models if they are to explain the joint dynamics of the equity and option markets.

References


9 Appendix

9.1 The Nonparametric Truncated Realized Variance Estimators

For ease of notation we normalize the time unit to be a day. Each day is then divided into a trading and a non-trading part and by convention a day starts with the close of trading on the previous day and ends at the closing of the following day trading period. The resulting daily interval $[t-1, t]$ is divided into $[t-1, t-1+\pi]$ overnight period and $[t-1+\pi, t]$ active part of the trading day. Over the trading part, we observe the futures price at $n+1$ equidistant times, resulting in $n$ intraday increments, each over a time interval of length $\Delta_n \equiv \frac{1-\pi}{n}$. The intraday increments are given by

$$\Delta_i^{n,t} f = f_{t-1+\pi+i\Delta_n} - f_{t-1+\pi+(i-1)\Delta_n} \text{ for } i = 1, \ldots, n \text{ and } t = 1, \ldots, T.$$ 

$\widehat{V}_t^{(n,m_n)}$ and $\widehat{V}_t^{(n,m_n)}$ are nonparametric estimators of the diffusive return variation constructed from the intraday record of the log-futures price of the underlying asset, $f$, as follows,

$$\widehat{V}_t^{(n,m_n)} = \frac{n}{m_n} \sum_{i=n-m_n+1}^{n} (\Delta_i^{n,t} f)^2 1_{\{\|\Delta_i^{n,t} f\| \leq \alpha \Delta_n \varpi \}},$$

(9.17)

where $\alpha > 0$, $\varpi \in (0, 1/2)$, and $m_n$ denotes a deterministic sequence with either $m_n/n \to 0$ or $m_n = n$. For $m_n = n$, $\widehat{V}_t^{(n,m_n)}$ is the truncated variation (see, e.g., Mancini (2009)), which provides a consistent estimate of the integrated diffusive variance over $[t-1, t]$. For $m_n/n \to 0$, $\widehat{V}_t^{(n,m_n)}$ is a consistent estimator of the spot variance at $t$ and corresponds to the truncated variation computed over an (asymptotically shrinking) fraction of the day just prior to the option quote. In our implementation, we sample every minute over a 6.75 hours trading day, excluding the initial five minutes, resulting in $n = 400$. We employ $m_n = 300$ for $\widehat{V}_t^{(n,m_n)}$.

Theoretically any $\alpha > 0$ and $\varpi \in (0, 1/2)$ will work. We fix $\varpi = 0.49$. The computation of $\alpha$ is more involved and takes into account for the fact that volatility changes over time and it also
shows a strong diurnal patter over the trading part of the day. To account for the latter diurnal patter we estimate nonparametrically a time-of-the-day factor $TOD_i$, $i = 1, ..., n$,

$$TOD_i = NOI_i \frac{\sum_{t=1}^{T}(\Delta_i^{n,t}f)^21_{\{\Delta_i^{n,t}f \leq \bar{\alpha}\Delta_n^{0.49}\}}}{\sum_{t=1}^{T}1\sum_{j=1}^{n}(\Delta_j^{n,t}f)^21_{\{\Delta_j^{n,t}f \leq \bar{\alpha}\Delta_n^{0.49}\}}}, \quad NOI_i = \frac{\sum_{t=1}^{T}1\sum_{j=1}^{n}1_{\{\Delta_j^{n,t}f \leq \bar{\alpha}\Delta_n^{0.49}\}}}{\sum_{t=1}^{T}1\sum_{i=1}^{n}1_{\{\Delta_i^{n,t}f \leq \bar{\alpha}\Delta_n^{0.49}\}}}.$$ \hspace{1cm} (9.18)

We compute $\bar{\alpha}$ as:

$$\bar{\alpha} = 3\sqrt{\frac{\pi}{2}}\sqrt{\frac{1}{T} \sum_{t=1}^{T} \sum_{i=2}^{n} |\Delta_i^{n,t}f||\Delta_i^{n,t}f|}.$$ 

The truncation level $\bar{\alpha}$ is based on the average volatility in the sample as measured by the so-called bipower variation measure defined in ?. The $NOI_i$ factor puts the numerator and denominator in equation (9.18) in the same unit. The time-of-the-day factor is plotted in Figure 9.1.

![Time-of-Day Factor $TOD_i$](image)

**Figure 14**: Time-of-the-day factor.

To account for time-varying volatility across days, we use the estimated continuous volatility for the previous day (for the fist day we use $\bar{\alpha}$). Finally, our time-varying threshold can be expressed as

$$\alpha_{t,i} = 3\sqrt{\frac{\hat{\nu}_{t-1}}{1 - \frac{1}{\pi}}} \times TOD_i \times \Delta_n^{0.49},$$

46
where the estimate $\hat{\pi}$ is naturally given by

$$
\hat{\pi} = \frac{\sum_{t=1}^{T} (f_{t+\pi} - f_t)^2}{\sum_{t=1}^{T} (f_{t+1} - f_t)^2}.
$$
Panel A: Overall \( RMSE_{IV} \) and ratios

<table>
<thead>
<tr>
<th></th>
<th>1FGSJ</th>
<th>2FGSJ</th>
<th>2FESJ</th>
<th>3Fvol</th>
<th>3F</th>
</tr>
</thead>
<tbody>
<tr>
<td>( RMSE_{IV} ) (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.48</td>
<td>2.66</td>
<td>2.26</td>
<td>2.27</td>
<td>2.01</td>
</tr>
</tbody>
</table>

Panel B: Sorting by moneyness and maturity

<table>
<thead>
<tr>
<th></th>
<th>1FGSJ</th>
<th>2FGSJ</th>
<th>2FESJ</th>
<th>3Fvol</th>
<th>3F</th>
</tr>
</thead>
<tbody>
<tr>
<td>( RMSE_{IV} ) ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \tau \leq 60 )</td>
<td>( \tau &gt; 60 )</td>
<td>( \tau \leq 60 )</td>
<td>( \tau &gt; 60 )</td>
<td>( \tau \leq 60 )</td>
<td>( \tau &gt; 60 )</td>
</tr>
<tr>
<td>( m \leq -3 )</td>
<td>0.35</td>
<td>0.17</td>
<td>0.45</td>
<td>0.42</td>
<td>-0.01</td>
</tr>
<tr>
<td>( -3 &lt; m \leq -1 )</td>
<td>0.95</td>
<td>0.58</td>
<td>0.28</td>
<td>0.18</td>
<td>0.04</td>
</tr>
<tr>
<td>( -1 &lt; m \leq 1 )</td>
<td>1.28</td>
<td>1.27</td>
<td>0.23</td>
<td>0.12</td>
<td>0.17</td>
</tr>
<tr>
<td>( m &gt; 1 )</td>
<td>1.06</td>
<td>0.93</td>
<td>0.32</td>
<td>0.28</td>
<td>0.44</td>
</tr>
</tbody>
</table>

Table 6: Out-of-sample Pricing Performance. 1FGSJ refers to the One-factor Gaussian jump model with time-varying jump intensity; 2FGSJ refers to the Two-factor Gaussian jump model with time-varying jump intensity; 2FESJ refers to the Two-factor Exponential jump model with time-varying jump intensity; 3Fvol refers to the Three-factor Exponential jump model with time-varying jump intensity; 3F refers to the Three-factor Exponential jump model independent jump factor. Finally, moneyness \( m \) is defined as \( (m = \log(K/F) / (\sigma_{ATM} \sqrt{T})) \) and \( \tau \) represents time-to-maturity in days. \( RMSE_{IV} \) ratio is the ratio between \( RMSE_{IV} \) of a given model over the \( RMSE_{IV} \) ratio of the 3F model, minus one.