Resolution of Policy Uncertainty and Sudden Declines in Volatility

Dante Amengual†
CEMFI

Dacheng Xiu‡
Chicago Booth

Abstract

We introduce downward volatility jumps into a general non-affine modeling framework of the term structure of variance. With variance swaps and S&P 500 returns, we find that downward volatility jumps are associated with a resolution of policy uncertainty, mostly through statements from FOMC meetings and speeches of the Fed chairman. We also find that such jumps are priced with positive risk premia, which reflect the price of the “put protection” offered by the Fed. Ignoring them may lead to an incorrect interpretation of such tail events. Moreover, variance risk premia tend to be insignificant or even positive at the inception of crises. On the modeling side, we explore the structural differences and relative goodness-of-fits of factor specifications, and find that the log-volatility model with two Ornstein-Uhlenbeck factors and double-sided jumps is superior in capturing volatility dynamics and pricing variance swaps, compared to the affine model prevalent in the literature or non-affine specifications without downward jumps.

Keywords: Non-Affine Derivative Pricing Models, Log Volatility Models, Quadratic Volatility Models, Downward Volatility Jumps, Variance Swaps

JEL Codes: G12, G13

---

*We benefited from extensive discussions with Nicolas Polson and Enrique Sentana. We also thank seminar and conference participants at Duke University, the 2014 Annual Meetings of the Western Finance Association, the 41st European Finance Association Annual Meeting, the 7th Annual SoFiE Conference, the 8th World Congress of the Bachelier Finance Society, the 2014 McGill Risk Management Conference, the 2014 China International Conference in Finance, the 2014 Toulouse Financial Econometrics Conference, the XXI Finance Forum in Segovia, and the 2013 Asian Meeting of Econometric Society. This research was funded in part by the Fama-Miller Center for Research in Finance at the University of Chicago Booth School of Business.

†Centro de Estudios Monetarios y Financieros. Address: Casado del Alisal 5 Madrid, 28014 Madrid, Spain. E-mail address: amengual@cemfi.es.

‡Booth School of Business, University of Chicago. Address: 5807 S Woodlawn Avenue, Chicago IL 60637, USA. E-mail address: dacheng.xiu@chicagobooth.edu.


1 Introduction

Volatility responds to news. It rises dramatically and immediately following the occurrence of unexpected bad events.\(^1\) Moreover, volatility not only jumps upward but also moves downward rapidly. Sudden declines in volatility are sometimes related to stock market rallies stimulated by unexpected good news from economic indicators or earning announcements. Yet they are also very often triggered by the resolution of policy uncertainty that shifts investors’ sentiment. Recent news headlines bring this indisputable fact into the spotlight. In particular, as can be seen from Figure 1, the VIX dropped 35% on May 10, 2010, as a result of Europe’s emergency loan plan; another 27% on Aug 9, 2011, due to Federal Reserve’s rate statement on keeping interest rates at a record low through mid-2013; and finally 23% on Dec 31, 2012, in anticipation of lawmakers making a deal to avert the “fiscal cliff.”

While the uncertainty of future fiscal and monetary policies may increase the volatility of asset prices, the government and Federal Reserve often intervene in the midst of hard times, which effectively provides a put protection on asset prices.\(^2\) Our hypothesis is that many downward volatility jumps are ex-post market reactions to these policy measures, and that they are important sources of risk for volatility buyers ex-ante. This type of variance risk should be priced in volatility derivatives, and could be an important part of the total variance risk premia. Many previous studies have reported a negative price of volatility jump risk. Ignoring downward volatility jumps, if they are priced, may lead to an incorrect interpretation of the price of tail events. The goal of this paper is to provide a systematic investigation of where downward volatility jumps originate, how they affect asset prices, and whether they are priced risk factors.

These questions invite us to search for appropriate derivatives to investigate the asset pricing implications of volatility shocks. While the S&P 500 options offer a developed battlefield for volatility trading, volatility derivatives have thrived on the demand for volatility hedging and speculation since their inception. The over-the-counter index variance swap contract is one particular example of these popular derivatives. As with most swaps, the fixed leg of variance swaps pays a predetermined amount at maturity in exchange for the realized variance that the floating leg commits to offer. Despite the path-dependence of realized variance, the payoff structure of variance swaps is appealing for studying the term structure of variance and variance risk premia, as opposed to the exchange-traded VIX derivatives, in that variance swaps directly reflect investors’ expectation on future uncertainty.\(^3\) Moreover, a variance swap can be replicated using a portfolio of S&P 500

---

\(^1\)For instance, the terrorist attack on September 11, 2001 sent the VIX near what had been its historical high.

\(^2\)We use “put protection” to refer to the monetary policy approaches that Alan Greenspan and Ben Bernanke, former Chairman of the Federal Reserve Board, exercised from 1987 to 2000 and during recent financial crisis.

\(^3\)Since 2004 and 2006, the Chicago Board Options Exchange (CBOE) has introduced VIX futures and VIX options, respectively, offering investors additional instruments for volatility trading. However, these contracts are written on the VIX, which is very similar to a 1-month variance swap, hence they are more complicated than variance swaps.
options, which is very similar to the VIX itself. Therefore it is very sensitive to volatility jumps.

Despite their existence, whether and how these volatility jumps are priced by investors remains largely unknown, particularly in the case of the large downward jumps. This is partially due to the absence of derivative pricing models that allow for downward volatility jumps in the mainstream finance literature. Popular affine models such as the square-root volatility models can only incorporate upward jumps in order to ensure the positivity of variance. We incorporate downward volatility jumps and other potentially negative latent factors into a non-affine framework that guarantees the positivity of variance.

With this new and general non-affine framework, we price variance swaps in a closed form, and identify downward volatility jumps along with two latent volatility factors from 17 years of variance swap data and S&P 500 returns. We find that volatility jumps are often triggered by unexpected macro announcements.\(^4\) In particular, sudden declines in volatility are mostly associated with the resolution of policy uncertainty, such as monetary policy changes that are explicit or implicit from Federal Open Market Committee (FOMC) statements or the speeches of the Federal Reserve’s chairman, as well as fiscal policy decisions and compromises made by Congress. Our analysis conforms with the existing model-free estimates in Table 1 that the total variance risk premia are negative most of the time, yet they tend to be insignificant or even positive at the inception of crises. This finding is a puzzle as it is in conflict with a representative agent model widely adopted in the literature. In addition we find that downward volatility jumps are priced with a positive price of risk, providing evidence for the compensation for variance risk from the “Greenspan Put” or “Bernanke Put.” Our regression analysis shows that latent volatility factors can be explained not only by excess market returns, but also by liquidity and credit factors, as well as policy news. In particular, policy news are important for the short-term factor, whereas the default risk is paramount for the long-term. More importantly, we also find that most downward volatility jumps affect the short-term factor, and that the impact of policy measures on long-term uncertainty is not always significant. Among several alternative specifications, we provide compelling evidence in favor of the log-volatility model with two Ornstein-Uhlenbeck factors and double exponential jumps.

There is a growing amount of theoretical and empirical work relating political uncertainty to asset pricing. In particular, Pástor and Veronesi (2013) relate the stock market risk premia, volatility, and correlation to the policy uncertainty index constructed by Baker et al. (2013) which is based on the frequency of newspaper references to economic policy uncertainty and other indicators. The regression results of Pástor and Veronesi (2013) agree with all the predictions of their learning model, see also Pástor and Veronesi (2012) for another related model of government policy choice. Boutchkova et al. (2012) investigate how local and global political risks affect industry return volatility. Kelly

---

\(^4\)While many macro announcements are pre-scheduled, their impact remains unexpected. As a result, the literature resorts to Poisson processes for modeling jumps, with notable distinctions by Maheu and McCurdy (2004), Piazzesi (2005), Dubinsky and Johannes (2006), and Beber and Brandt (2009).
et al. (2016) find evidence for government guarantee premia by examining the basket-index spread from out-of-the-money put options. Bernanke and Kuttner (2005) study stock market reactions to Federal Reserve policy and find that the effects of unanticipated monetary policy actions on expected excess returns account for the largest part of the responses of stock prices. Beber and Brandt (2009) investigate the link between ex-ante macroeconomic uncertainty and ex-post uncertainty resolution in financial markets, using the prices of some options whose underlying is the release of non-farm payroll. They find that higher ex-ante uncertainty leads to a larger reduction in volatility along with a greater increase in trading activity after the news release. While these studies have shed light on the link between political uncertainty and risk in equity markets, we further point out that sudden decreases in volatility are particularly related to the resolution of monetary policy uncertainty, and provide evidence showing that some portion of the variance risk premia is associated with investors’ anticipation of the Federal Reserve’s action.

Our empirical findings on volatility jumps are also relevant to the large literature that investigates the unique role of jumps in asset pricing, which dates back as early as Merton (1976), who introduces jumps to model index returns. Since the seminal work by Duffie et al. (2000), positive volatility jumps, exponentially distributed, have been constantly added to model index volatility dynamics, so that volatility can jump upward but revert back to the mean slowly. Eraker et al. (2003), in particular, point out the role played by such volatility jumps and compare them to the role of jumps in returns. However, models in the literature that discuss the existence and necessity of downward volatility jumps are rare. An exception is Andersen et al. (2015a), who document downward volatility jumps using the constructed intraday corridor implied volatilities. Also, Todorov and Tauchen (2011) investigate the activity of volatility jumps using high-frequency historical returns of the VIX. Their specification allows downward volatility jumps. In contrast, we focus on the asset pricing implications of volatility jumps and their price of risk, which require modeling the risk-neutral and the objective dynamics jointly. Recently, Chernov et al. (2016) have discussed the impact of jumps on exchange rates and the impact of positive jumps on their variances, and they relate these to macroeconomic and political news. They find few positive jumps in variance that respond to such news. We also find positive jumps less responsive to political news, unlike negative jumps.

Previous work in the literature on variance swaps is mostly based on fully specified parametric models using both variance swaps and index values. Egloff et al. (2010) and Amengual (2008) find that single-factor volatility models are incapable of fitting the term structure of variance swap rates. They therefore suggest applying models with two-volatility factors to investigate the term structure of variance. None of their models have volatility jumps. Aït-Sahalia et al. (2014) propose a similar affine model with positive volatility jumps to estimate the liquidity and variance risk premia. They focus on the component of variance risk premia due to price jumps. Li and Zinna (2016) further add a self-exciting jump factor to the same affine model. Fusari and Gonzalez-Perez (2012) consider
a log-affine model with two Ornstein-Uhlenbeck factors but without volatility jumps, in addition to
an affine model. Carr et al. (2012) focus on the pricing and hedging of variance swaps and volatility
derivatives using time-changed Lévy processes. Filipovic et al. (2015) independently propose a class
of quadratic models without volatility jumps. All the aforementioned continuous-time models are
nested within our framework. Recently, Dew-Becker et al. (2016) propose to investigate structural
economic models using variance swaps. Their affine models are cast in discrete-time and only positive
volatility jumps are allowed. While these two-factor volatility models without volatility jumps have
been shown to yield accurate variance swap prices, our empirical results suggest that their dynamics
under the objective measure are likely misspecified.

To study variance risk premia, many papers adopt alternative nonparametric techniques. Among
500 options, whereas Carr and Wu (2009) study variance risk premia using synthetic variance swaps
for individual firms and indexes. The synthetic variance swap price thereafter becomes a popular
proxy of the risk neutral conditional variance. To measure the conditional variance under the ob-
jective measure, Bollerslev et al. (2011) suggest the use of high-frequency five-minute-based realized
volatilities, see also Zhou (2009). Although realized volatilities are model-free estimates, estimating
the objective conditional expectation requires a parametric forecasting model. In this regard, Bekaert
and Hoerova (2014) evaluate a plethora of state-of-the-art forecasting models to produce an accurate
measure of the conditional variance, and point out that a non-linear model may be better equipped
to capture the behavior of conditional variance and variance risk premia in severe crises. We specify
and estimate non-affine dynamic models for the index return and its volatility, hence we are able to
address their conjecture. Moreover, our full-fledged and unified model facilitates the joint statistical
inference on conditional variances under both measures. It is worth emphasizing that unlike these
papers, our focus mainly lies on the economic interpretation and pricing implications of downward
volatility jumps, which require a more structural approach.

Our paper is also related to the specification of models that can capture index volatility dynamics,
one of the central themes in empirical option pricing and financial econometrics. This strand of
literature investigates the volatility dynamics through the lens of S&P 500 options, see, e.g., Bakshi
et al. (1997), Bates (2000), Pan (2002), Eraker (2004), and Broadie et al. (2007) for examples of
affine jump diffusion models with stochastic volatility driven by one square-root factor. Recent
findings by Christoffersen et al. (2009) and Bates (2012) also suggest that models with two square-
root factors are essential for capturing the term structure of variance. Andersen et al. (2015b)
argue for more factors in order to capture the time-varying skewness of the implied volatility. All
these papers focus on affine volatility models, in particular the square-root models. Nevertheless,
ample evidence from historical time series of stock returns supports log-volatility models, including
discrete-time ones by French et al. (1987), Schwert (1990), Nelson (1991), as well as continuous-
time models, potentially with jumps or even comprised purely of jumps, e.g., Barndorff-Nielsen and Shephard (2001), Chernov et al. (2003), and Todorov and Tauchen (2011). Indeed, log-volatility models naturally allow downward volatility jumps since they always guarantee the positivity of variance. Plus, log-volatility models allow Ornstein-Uhlenbeck factors, which are not restricted by a similar Feller’s condition for square-root processes. Empirically, Feller’s condition is often binding for the risk neutral dynamics, even for models with multiple volatility factors, see, e.g., Song and Xiu (2016). Therefore, log-volatility models allow for more persistent volatility factors. The drawback of these log-volatility models lies in their lack of tractability for option pricing. Instead of relying on options, we resort to variance swaps and derive a closed-form pricing formula, using which we can investigate the pricing implications of log-volatility models. There are a couple of papers in the empirical option pricing literature, though, which investigate non-affine risk neutral dynamic models using simulation methods, e.g., Christoffersen et al. (2006) and Durham (2013). However, conducting statistical inference on top of simulated prices is computationally intensive.

This paper is organized as follows. Section 2 presents our framework for variance swap modeling. Section 3 discusses the statistical inference, followed by empirical results in Section 4. Section 5 concludes the paper. The Appendix provides mathematical proofs and additional technical details.

2 Variance Swap Modeling

A variance swap contract is an over-the-counter derivative in which the contract holder pays at maturity $t + \tau$ a fixed amount (variance swap rate) for the realized variance:

$$\frac{1}{\tau} \sum_{i=1}^{\lfloor \tau/\Delta \rfloor} \left( Y_{t+i\Delta} - Y_{t+(i-1)\Delta} \right)^2,$$

where $Y$ is the log-price of the underlying index, i.e. S&P 500 index. By entering long positions in such contracts, investors can hedge against high realized variance. Thus, the differences between the expectation of variance and the swap price, i.e., the variance risk premia investors earn, are typically negative, see, e.g., Bollerslev et al. (2009) and Drechsler (2013).

Variance swap trading has grown rapidly since the aftermath of the LTCM turmoil in late 1990s. For investors using medium- or low-frequency trading strategies, these over-the-counter contracts are more favorable than S&P 500 options for the purpose of volatility trading, since investors can express their views on volatility without having to do labor-intensive delta hedging.

We start by proposing a full-fledged multi-factor non-affine volatility model for which we provide a general pricing formula of variance swaps in Section 2.1. Section 2.2 specifies the risk premia and the dynamics under the objective measure. Section 2.3 discusses canonical forms and identification. We then provide examples of two-factor volatility models in Section 2.4, which we use in the empirical study.
2.1 Risk Neutral Modeling and Pricing

As is well known, realized variance converges (in probability) to the quadratic variation of $Y$, i.e. $[Y, Y]_{t,t+\tau}$, and modeling the quadratic variation is a common practice that facilitates the variance swap pricing. Since there is no money changing hands at the initiation of the trade, i.e., time $t$, the variance swap rate, under some risk neutral measure $Q$, is given by:

$$P(t, \tau) = 100 \times \frac{1}{\tau} \mathbb{E}^Q_t \left\{ \int_t^{t+\tau} \sigma_s^2 ds + \int_t^{t+\tau} \int_{\mathbb{R}} j^2 \nu_s^Q(dj) ds \right\},$$

where the calculation is based on the usual specification of the risk neutral dynamics of $Y$:

$$dY_t = \mu_t^Q dt + \sigma_t dB_t^Q + dJ_t^Q,$$

where $B_t^Q$ is a Brownian motion, $J_t^Q$ is a compensated jump process with compensator $\nu_t^Q(\cdot)$, $\sigma_t$ is a volatility process, and $\mu_t^Q$ is the drift determined by the no-arbitrage condition.

2.1.1 Variance Dynamics

We model the variance as certain non-affine function of some factors summarized in $X$:

$$\sigma_t^2 = \Pi_0 + \Pi_1^T X_t + X_t^T \Pi_2 X_t + \exp \left\{ \Pi_3 + \Pi_4^T X_t \right\},$$

where $\Pi_0 - \frac{1}{\tau} (\Pi_1)^T (\Pi_2)^{-1} \Pi_1 \geq 0$, which warrants a positive variance. This model augments the exponentially affine specification by a quadratic component, hence nesting the common affine cases when $X$ only takes positive values, as well as the quadratic variance swap model by Filipovic et al. (2015).

To ensure the tractability of this general non-affine class of models, we assume that the underlying $N$-dimensional factor $X_t$ follows a multivariate affine process, similar to the affine term structure model discussed in Dai and Singleton (2000), but allowing for jumps, e.g., as in Duffie et al. (2000) and Chen and Joslin (2012). We write the risk neutral model of $X_t$ as:

$$dX_t = (\Lambda^Q + K^Q X_t) dt + \Sigma \sqrt{S_t} dW_t^Q + dZ_t^Q,$$

where $W_t^Q$ is an $N$-dimensional standard Brownian motion, and $S_t$ is a diagonal matrix in $\mathbb{R}^{N \times N}$ with $[S_t]_{i,i} = \alpha_i + \beta_i^T X_t$, and $Z_t^Q$ is another compensated jump process. While there is no need to introduce correlations among $W_t^Q$ because of $\Sigma$, we impose a correlation between $B_t^Q$ and each element of $W_t^Q$ to incorporate the so-called “leverage effect.”

Practitioners price variance swaps using a replicating portfolio of options, which relies on the same quadratic variation approximation, see Bossu et al. (2005), hence the discretization error can be ignored, e.g., Jiang and Tian (2007).

As shown in Cheridito et al. (2010), the canonical forms in Dai and Singleton (2000) are not exhaustive, unless $m \leq 1$ or $m \geq N - 1$. While Duffie et al. (2003) and Joslin (2016) propose more general affine processes, we adopt the model used by Dai and Singleton (2000) for its popularity and simplicity.
While the factor $X$ is restricted within the affine class, the volatility dynamics is non-affine, which leads to several differences compared with the usual term structure models. For example, even when $X$ is a homoscedastic Gaussian factor, $\sigma_t^2$ is heteroscedastic and non-Gaussian, as is obvious from Itô’s lemma. Moreover, the volatility of volatility is another (non-affine) function of $X$.\footnote{The volatility of volatility, $[d\sigma_t^2, d\sigma_t^2]^{c}/dt$, can be calculated by Itô’s lemma, where $[\cdot, \cdot]^{c}$ denotes the continuous component of the quadratic variation. Since $\sigma_t^2$ is a non-linear function of $X_t$, $[d\sigma_t^2, d\sigma_t^2]^{c}/dt$ is in general a non-linear function of $X_t$.}

### 2.1.2 Jumps

To specify jumps in both $Y$ and $X$, there are trade-offs that must be considered. First, Poisson type jumps are our preferred choice for modeling daily data, as Lévy type jumps are difficult to identify and disentangle from Brownian shocks generated by stochastic volatility at a daily frequency.\footnote{There is a large literature on jump detection with intraday data, see e.g. Huang and Tauchen (2005), Barndorff-Nielsen and Shephard (2006), Lee and Mykland (2008), Jiang and Oomen (2008), Jacod and Todorov (2009), and Aït-Sahalia and Jacod (2009).} Second, if the intensities of Poisson jumps are independent for $Y$ and $X$, then there would be no co-jumps of $Y$ and $X$ almost surely, which conflicts with the data, see, e.g., Jacod and Todorov (2010). Third, there are many pre-scheduled macro announcements, FOMC meetings, and speeches by the Federal Reserve Chairman, which potentially cause jumps on the market. Hence it may be reasonable to model jumps with deterministic timing, see, e.g., Piazzesi (2005), Maheu and McCurdy (2004), Dubinsky and Johannes (2006), and Beber and Brandt (2009). However, there are many days in our sample with at least one such event, and jumps are not always present. From our empirical analysis below, whether a jump occurs or not on a scheduled event depends on the extent of the news surprise, i.e., the content of the announcement. Most of these events do not lead to jumps. Moreover, from the perspective of risk premia estimation, the major difference between deterministic timing and the random arrival of jumps lies in the risk premia associated with the intensity of the jumps – there are no risk premia for the deterministic arrival of jumps, which may not be the case for the random arrival. Hence, for the sake of parsimony, we conform with the common practice in the literature, e.g., Pan (2002), and model jumps in $Y$ and $X$ as compound Poisson processes driven by the same intensity, with no price of risk associated with the jump intensity. Therefore, we write

$$
\left(J^Q_t, Z^Q_t\right) = \int_0^t \int_R \int_{\mathbb{R}^N} (j, z) \left( N(ds, dj, dz) - \nu^Q_s(dj, dz)ds \right),
$$

where $N$ is the Poisson random measure, $\nu^Q_s(dj, dz)$ denotes its compensator, and the Poisson jump intensity is given by $l_0 + l_1^\top X_t$ with $l_0 \in \mathbb{R}_+$ and $l_1 \in \mathbb{R}_+^N$. $l_1$ only has non-zero and positive loadings on positive factors in $X$.\footnote{There is a large literature on jump detection with intraday data, see e.g. Huang and Tauchen (2005), Barndorff-Nielsen and Shephard (2006), Lee and Mykland (2008), Jiang and Oomen (2008), Jacod and Todorov (2009), and Aït-Sahalia and Jacod (2009).}
2.1.3 Variance Swaps Valuation

We now derive the variance swap rate under the proposed model:

**Proposition 1.** Suppose the risk neutral dynamics follow (1), (2), (3), and (4). The variance swap rate is given in closed form by:

\[
P(t, \tau, X_t) = \frac{100}{\tau} \left\{ \int_t^{t+\tau} \Pi_0^Q + (\Pi_1^Q)^T \nabla_u \Psi(s, t, u, X_t) \bigg|_{u=0} + \nabla_u^T \Pi_2 \nabla_u \Psi(s, t, u, X_t) \bigg|_{u=0} ds \right. \\
+ \left. \int_t^{t+\tau} \exp(\Pi_3) \Psi(s, t, \Pi_4, X_t) ds \right\},
\]

where \( \nabla_u = (\partial/\partial u_1, \ldots, \partial/\partial u_N)^T \) is a derivative operator, \( \Psi(s, t, u, X_t) = \mathbb{E}_t^Q \left[ e^{u^T X_s} \right] \), \( \Pi_0^Q = \Pi_0 + l_0 \int_{\mathbb{R}} j^2 \nu^Q(dj) \), \( \Pi_1^Q = \Pi_1 + l_1 \int_{\mathbb{R}} j^2 \nu^Q(dj) \), and \( \nu^Q(dj) \) is the marginal distribution of the jump size of \( Y \) in (1).

The derivation given in Appendix A is based on the Fourier Transforms of tempered distributions. The technique has been adopted in the mathematical finance literature for pricing options, e.g., Lipton (2001), for affine models, whereas we apply it for non-affine models. In fact, the specification of variance in (2) can be generalized to the so-called tempered distributions.\(^9\) We choose quadratic and exponential functions because they nest the common specification of variance dynamics in the literature. Using a similar technique, Sepp (2008) develops the closed-form VIX option prices for the Heston model, whereas Chen and Joslin (2012) extend Duffie et al. (2000) to obtain a closed form formula for nonlinear moments of affine processes.

Regarding the implementation of the pricing formula, the computational expense is the same as that of any affine model for pricing options. The time-consuming part is the numerical integration over solutions of ordinary differential equations, which are necessary to obtain \( \Psi \) and its first and second order derivatives in general.\(^10\) That said, the numerical integration is over a finite interval, which converges faster than the integration over the typical infinite domain needed for option pricing.

2.2 Risk Premia Specification

In general, risk premia can be chosen as completely affine, e.g., Dai and Singleton (2000), or essentially affine, e.g., Duffee (2002), or can be defined as the scaled difference between \( \mathbb{P} \)- and \( \mathbb{Q} \)-measures. As shown by Cheridito et al. (2007), the last procedure can also ensure lack of arbitrage as long as the

\(^9\)Tempered distribution refers to the distribution of functions in the Schwarz space, a linear space of functions all of whose derivatives are rapidly decreasing, see Stein and Shakarchi (2003). Fourier transform is well-defined for a tempered distribution, see Kanwal (2004). The term “distribution” here should not be confused with the “distribution” in statistics.

\(^10\)Note that each derivative of \( \Psi \) contributes to one additional differential equation. The pricing routine is written in C++, and is available upon request.
existence and boundary non-attainment conditions are satisfied under both measures. They call this extended affine specification. We follow their suggestion, and specify the objective dynamics as

\[
\begin{align*}
    dY_t &= \mu_t^P dt + \sigma_t^P dB_t^P + dJ_P^t, \\
    dX_t &= (\Lambda^P + K^P X_t) dt + \Sigma \sqrt{S_t} dW_t^P + dZ_P^t,
\end{align*}
\]

(5) (6)

where \((J_P^t, Z_P^t)\) is specified in the same form as in (4). Therefore, we leave market prices of risk defined as the scaled differences between the drifts under \(P\) and \(Q\). This market price of risk specification does not offer arbitrage opportunities, provided that the existence and the boundary non-attainment conditions are satisfied under both measures.\(^{11}\) These restrictions are given explicitly in Appendix C for the models we estimate. In the empirical study, we examine whether these restrictions hold for models being fit.

2.3 Extended Canonical Forms

To understand the admissibility and identification of the proposed non-affine models, we study their canonical forms. Canonical forms are unique and maximal representations for a class of models with the same observable implications. The canonical forms of affine term structure models have been discussed in Dai and Singleton (2000) in the absence of jumps.

We can recycle their notation and extend their canonical forms by imposing similar identifying constraints on the model excluding jumps. We then add jumps on top of these canonical forms, with restrictions that the intensity can only load on positive factors, and that jumps of the positive factors can only have positive sizes.

More specifically, we classify a model as \(\mathcal{A}_m(N)\), if \(N\) is the number of state variables, and \(m\) is the number of independent linear combinations of those state variables that appear in the diffusion matrix, i.e., \(m = \text{rank}(\mathcal{B})\), where \(\mathcal{B} = (\beta_1, \ldots, \beta_N)\). The state variables in the diffusion matrix are non-negative. For each \(m\), we partition \(X^\top = (X_{m \times 1}^\top, X_{(N-m) \times 1}^\top)^\top\). We present the extended canonical forms below. For reasons of space, these canonical forms do not allow pure jump factors.\(^{12}\)

**Definition 1.** The extended canonical representation takes a special form of equation (3), where for

\(^{11}\)The existence and uniqueness of \(X_t\) follows from Theorem 2.7 in Duffie et al. (2003), which in turn implies the existence and uniqueness of \(\sigma_t^P\) and \(Y_t\), since they can be written explicitly in terms of \(X_t\) or its stochastic integral defined by (1) and (5). To show the existence of an equivalent probability measure \(Q\), which ensures that our specification precludes arbitrage opportunities, we point out that the semimartingales specified as solutions to \(\{1, (2), (3)\}\) and \(\{(2), (5), (6)\}\) satisfy the assumptions of Corollary 3.68 of Jacod and Shiryaev (2003), since the drifts and the diffusions of \(Y\) and \(X\) are locally Lipchitz, and jumps are locally bounded. The desired absolute continuity between \(P\) and \(Q\) follows from Theorem 2.6(a) in Jacod and Shiryaev (2003), provided that the associated Hellinger process is a.s. finite.\(^{12}\) Extended canonical forms with pure jump factors are available upon request.
m > 0,

\[ K^Q = \begin{pmatrix}
K^Q_{m \times m} & 0_{m \times (N-m)} \\
0_{(N-m) \times m} & K^Q_{(N-m) \times (N-m)}
\end{pmatrix},
\]

and \( K^Q \) is either the upper or lower triangle for \( m = 0 \). In addition,

\[ \Lambda^Q = \begin{pmatrix}
\Lambda^Q_{m \times 1} \\
0_{(N-m) \times 1}
\end{pmatrix}, \quad \Sigma = I_{N \times N}, \quad \alpha = \begin{pmatrix}
0_{m \times 1} \\
1_{(N-m) \times 1}
\end{pmatrix}, \]

\[ B = \begin{pmatrix}
I_{m \times m} & B_{m \times (N-m)} \\
0_{(N-m) \times m} & 0_{(N-m) \times (N-m)}
\end{pmatrix}, \quad l_1 = \begin{pmatrix}
l_{1, m \times 1} \\
0_{(N-m) \times 1}
\end{pmatrix},
\]

with admissibility restrictions (existence conditions) such that for \( 1 \leq i \neq k \leq m \) and \( 1 \leq j \leq N \),

\[ B_{ij} \geq 0, \quad l_{1,i} \geq 0, \quad l_0 \geq 0, \quad \bar{\nu}^Q(R^m_\infty \times \mathbb{R}^{N-m}) = 0. \]

Our specification of jumps leads to the following boundary nonattainment condition:

\[ K^Q_{i,k} \geq 0, \quad \Lambda^Q_{i} - l_0 \int_{\mathbb{R}} z_i \bar{\nu}^Q(dz) \geq \frac{1}{2}, \quad \text{for} \quad 1 \leq i \neq k \leq m,
\]
as well as the stationarity condition:

\[ \text{Re}(\text{Eigen}(\bar{K}^Q)) < 0, \]

where, using \( \text{Diag} \) as an operator that maps a vector to a diagonal matrix,

\[ \bar{K}^Q = \begin{pmatrix}
K^Q_{m \times m} - \text{Diag} \left( l_{i,i} \int_{\mathbb{R}} z_i \bar{\nu}^Q(dz) \right)_{1 \leq i \leq m} & 0_{m \times (N-m)} \\
K^Q_{(N-m) \times m} & K^Q_{(N-m) \times (N-m)}
\end{pmatrix}.
\]

These conditions are similar to those in Aït-Sahalia and Kimmel (2010) when jumps are absent and the entire model is affine.

Similar to Dai and Singleton (2000), we have:

**Proposition 2.** For any process that satisfies (1), (2), (3) and (4), there exists a unique canonical representation that is observationally equivalent to it.

Therefore, the canonical representation \( \Lambda_m(N) \) is not only admissible, but it is also maximal in the sense that for each \( m \) we have imposed minimal known sufficient conditions for admissibility and minimal normalizations for econometric identification.

### 2.4 Examples of Two-Factor Volatility Models

Modeling volatility as a two-factor process is an established approach from the literature. Engle and Rangel (2008) decompose volatility shocks into their short-term and long-term components, and
relate the long-term component to business cycles in a comprehensive international setting. Adrian and Rosenberg (2008) also decompose equity volatility into similar components, and in addition relate the short-term component to market skewness risk with a cross-section of equity returns. Corradi et al. (2013) directly model the market volatility as a combination of business cycle factors and one additional latent factor, and find that their macro-factors explain the majority of volatility fluctuations. Christoffersen et al. (2009) also find a two-factor volatility structure necessary to model S&P 500 options.

Specifically, we write the risk-neutral dynamics of $X$, a special case of (3), as

$$
\begin{bmatrix}
\frac{dX_1}{dt} \\
\frac{dX_2}{dt}
\end{bmatrix} = \left(\begin{bmatrix}
\lambda_1^Q \\
\lambda_2^Q
\end{bmatrix} + \begin{bmatrix}
\kappa_{11}^Q & \kappa_{12}^Q \\
\kappa_{21}^Q & \kappa_{22}^Q
\end{bmatrix} \begin{bmatrix}
X_1 \\
X_2
\end{bmatrix}\right) dt + \begin{bmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{bmatrix} \begin{bmatrix}
\frac{dW_1^Q}{dt} \\
\frac{dW_2^Q}{dt}
\end{bmatrix} + \begin{bmatrix}
\frac{dZ_1^Q}{dt} \\
\frac{dZ_2^Q}{dt}
\end{bmatrix},
$$

(7)

where jumps follow compound Poisson processes with independent jump sizes following double exponential distributions:

- size of $Z_{1t}^Q \sim \begin{cases} 
\exp(\beta_{1+}^Q), & q_1 \\
-\exp(\beta_{1-}^Q), & 1 - q_1 
\end{cases}$, and
- size of $Z_{2t}^Q \sim \begin{cases} 
\exp(\beta_{2+}^Q), & q_2 \\
-\exp(\beta_{2-}^Q), & 1 - q_2 
\end{cases}$.

Their intensity is specified as $l_0 + l_{11}X_{1t} + l_{12}X_{2t}$. In the canonical forms, the specification of jump size distribution is flexible. For parsimony, we employ a simple double-exponential distribution so as to allow for downward jumps as well as the asymmetry in the size of upward and downward jumps. Note that a double-exponential distribution is a natural extension of the exponential distribution typically used in the literature.

Equation (7) nests the three canonical forms we consider in Sections 3 and 4: $A_0(2)$, $A_1(2)$, and $A_2(2)$, with each allowing a two-factor structure, and with the first two allowing for negative volatility jumps. We spell out the details of these models in Appendix C.

For comparison purpose, in addition to these maximal non-affine models, we also fit three special cases in our empirical study, including the special cases of $A_0(2)$ and $A_1(2)$ without negative jumps (i.e., $q_1 = 1$ and $q_2 = 1$), as well as the affine special case of $A_2(2)$ (i.e., $\Pi_2 = 0$, $\Pi_3 = 0$, $\Pi_4 = 0$). We denote them $A_0^+(2)$, $A_1^+(2)$, and $A_2^+(2)$, respectively. $A_2^+(2)$ is a widely used model in the literature, see, e.g., Egloff et al. (2010) and Aït-Sahalia et al. (2014).

While we specify different models for volatility, they share the same return dynamics (1). We assume that the size of $J_t^Q$ follows a Gaussian distribution with mean $\mu_t^Q$ and variance $\sigma_t^2$. As a result, the compensator in $dJ_t^Q$ is $\mu_t^Q (l_0 + l_1^X_t) dt$ and the drift $\mu_t^Q$ can be written down explicitly.
\[ \mu_t^Q = r_t - d_t - \frac{1}{2} \sigma_t^2 - \left( e^{\mu_J^0 + \frac{1}{2} \sigma_J^2} - 1 - \mu_J^0 \right) \left( l_0 + l_1^\top X_t \right). \]

Since the interest rate \( r_t \) and dividend \( d_t \) do not affect variance swap prices, their risk neutral dynamics are not identified, hence they are left unspecified.

As the payoff of variance swaps depends on the underlying index \( Y_t \) only through its risk neutral quadratic variation, variance swaps contain much less information about the dynamics of \( Y_t \), compared to European options. As a result, with variance swaps, we can only identify the risk neutral jumps of \( Y \) up to the expected quadratic variation. For this reason, we impose no market price of risk on the variance of jump sizes in \( Y \), so that this parameter can be identified from the \( \mathbb{P} \)-measure dynamics using the S&P 500 index. The mean of jump sizes in \( Y \) absorbs all the risk premia of the price jumps, which can be identified from the expected quadratic variation under the \( \mathbb{Q} \)-measure. This assumption is also imposed by, e.g., Aït-Sahalia et al. (2014).

Finally, the \( \mathbb{P} \)-measure dynamics with two volatility factors can be written explicitly as:

\[
\begin{bmatrix}
    dX_{1t}^\mathbb{P} \\
    dX_{2t}^\mathbb{P}
\end{bmatrix} = \left( \begin{bmatrix}
    \lambda_1^\mathbb{P} \\
    \lambda_2^\mathbb{P}
\end{bmatrix} + \begin{bmatrix}
    \kappa_{11}^\mathbb{P} & \kappa_{12}^\mathbb{P} \\
    \kappa_{21}^\mathbb{P} & \kappa_{22}^\mathbb{P}
\end{bmatrix} \begin{bmatrix}
    X_{1t} \\
    X_{2t}
\end{bmatrix} \right) dt + \begin{bmatrix}
    \sqrt{\alpha_1 + \beta_{11} X_{1t} + \beta_{12} X_{2t}} dW_{1t}^\mathbb{P} \\
    \sqrt{\alpha_2 + \beta_{21} X_{1t} + \beta_{22} X_{2t}} dW_{2t}^\mathbb{P}
\end{bmatrix} + \begin{bmatrix}
    dZ_{1t}^\mathbb{P} \\
    dZ_{2t}^\mathbb{P}
\end{bmatrix},
\]

with double-exponentially distributed jumps but different parameters from those under the \( \mathbb{Q} \) measure. For each canonical form under \( \mathbb{Q} \), we adopt a \( \mathbb{P} \)-model within the same category. More specifically, for each of \( A_0(2) \), \( A_1(2) \), and \( A_2(2) \), we adopt the same constraints on \( K^\mathbb{P} \) as that on \( K^\mathbb{Q} \), but leave \( \Lambda^\mathbb{P} \) unconstrained to obtain more flexibility for the market prices of risk, which are implicitly defined as the differences between \( \mathbb{Q} \) and \( \mathbb{P} \). This does not affect the identification of \( \Lambda^\mathbb{P} \), since we set \( \Pi \) to be the same under the two measures, and \( \Pi \) is identified from the variance swap prices alone.\(^{13}\) The \( \mathbb{P} \)-dynamics of \( A_0^\dagger(2) \), \( A_1^\dagger(2) \), and \( A_2(2) \) are in turn determined as special cases.

In the dynamics of returns, (5), we assume that the size of jumps under \( \mathbb{P} \) is Gaussian with mean \( \mu_J^\mathbb{P} \) and variance \( \sigma_J^2 \). The intensity is the same under \( \mathbb{P} \) and \( \mathbb{Q} \), i.e., \( (l_0 + l_1^\top X_t) \). We do not, however, specify \( \mu_I^\mathbb{P} \), as it is poorly estimated from variance swaps or even from options, and the focus of this paper is not equity risk premia. Therefore, in our estimation, we follow Eraker et al. (2003) and Eraker (2004) by treating \( \mu_I^\mathbb{P} \) as a constant. We also confirm in simulations below that such misspecification does not have any noticeable impact on the inference of the remaining parameters.

Overall, the total numbers of parameters equal to 34 for the \( A_0(2) \), \( A_1(2) \), and \( A_2(2) \) models, and 28, 31, and 28 for the \( A_0^\dagger(2) \), \( A_1^\dagger(2) \), and \( A_2(2) \) models, respectively. A full list of parameters is available from the first columns of Tables 5, 6, 7, and 8.

\(^{13}\)Once we specify a \( \mathbb{Q} \)-canonical form, \( \mathbb{P} \)-dynamics is partially determined by the Girsanov Theorem. Therefore, using the same canonical form under \( \mathbb{P} \) imposes restrictions on risk premia implicitly.
3 Likelihood Inference

Our estimation strategy relies on observations of the joint time-series of the underlying S&P 500 index and several variance swap rates with different maturities. Our joint modeling strategy allows us to separately pin down risk premia related to each of the different sources of uncertainty. However, as is common in many financial models with jump diffusions, likelihood functions are not available. In addition, our state variables are latent, and sometimes non-Gaussian. Moreover, as discussed in Section 4.1 below, our panel of data is unbalanced. Therefore, we resort to Markov Chain Monte Carlo (MCMC) methods, see, e.g., Johannes and Polson (2010) for a detailed survey.

3.1 Posterior Simulator

We assume that there are observations available on S&P 500 returns and $k$ different variance swap rates and that observations are recorded at a daily frequency $\Delta = 1/252$, and that the total number of time periods under consideration is $T$. Let $Y$ denote the $T \times 1$ vector of S&P 500 prices, and $P$ denote the $T \times k$ panel of variance swap rates.

For convenience, we introduce $V$ and $\Theta$ to summarize latent variables and parameters. Typically, $V$ will contain the latent factors in $X$ of the model as well as the remaining latent variables such as jump sizes (denoted by $j_t$ and $z_t$) and jump times (denoted by $n_t$), even though they do not enter into the pricing formula. As for $\Theta$, we split it into $(\Theta_M, \Theta_{\Pi}, \Theta_P, \Theta_E)$. $\Theta_M = (\Lambda^Q, \text{vec}(K^Q), \{\alpha_i, \beta_i\}_{i=1}^m, \theta^Q_Z)$, which contains the parameters determining the dynamics of the latent factors under the risk-neutral measure, with $\theta^Q_Z$ denoting the parameters governing the jump processes; $\Theta_{\Pi} = (\Pi_0, \text{vech}(\Pi_1), \text{vech}(\Pi_2), \Pi_3, \text{vech}(\Pi_4))$ includes the parameters defining $\sigma^2_t$; $\Theta_P = (\Lambda^P, \text{vech}(K^P), \rho_1, \rho_2, \theta^P_Z)$ summarizes the remaining $P$-measure parameters; and finally, given that there are more derivatives than sources of uncertainty in the theoretical model we allow pricing errors to avoid stochastic singularity. The pricing errors are also economically important in that they capture the remaining factors that are not captured by our pricing model, such as the illiquidity factor or the counterparty risk factor. Specifically, we assume additive pricing errors $\varepsilon^j_i$ associated with time $i\Delta$ for variance swap with maturity $j$, so that the observed price satisfies

$$P^j_i = P(i\Delta, \tau_j, X_{i\Delta}; \Theta_M, \Theta_{\Pi}) + \varepsilon^j_i,$$

with $\varepsilon^j_i$ following a Gaussian distribution with variance $s^2_{j,i}$, and $\varepsilon^j_i$ is independent of $\varepsilon^h_i$ for $h \neq j$ and across time. Therefore, the pricing errors are heteroscedastic in the cross-section. $s^2_{j,i}$’s are stacked in $\Theta_E$. Notice that all prices except the S&P 500 index are assumed to be observed with error in our framework and, therefore, there is no need to assume that certain combinations of variance swap prices are perfectly observed.

The purpose of MCMC sampling is to obtain a sample of parameters $\Theta$ and latent variables $V$ from their joint posterior density. Specifically, for a given model $M$, the posterior distribution is
given by
\[ p(V, \Theta|Y, P, M) \propto \mathcal{L}(Y, P|V, \Theta, M) \cdot \mathcal{H}(V|\Theta, M) \cdot p(\Theta|M) \] (9)
where \( \mathcal{L}(Y, P|V, \Theta, M) \) denotes the likelihood function, \( \mathcal{H}(V|\Theta, M) \) is the density for the latent variables, and \( p(\Theta|M) \) is the prior density over the parameter vector \( \Theta \).

We use a Gibbs sampling procedure to estimate these models. In essence, this amounts to reducing a complex problem, i.e., sampling from the joint posterior distribution, into a sequence of tractable ones, i.e., sampling from conditional distributions for a subset of the parameters conditional on all the other parameters, for which the literature already provides a solution. The Gibbs sampling procedure involves sampling sequentially from several blocks:

- **Latent factors**: \( p(V_t^{(g)}|V_{<t}^{(g)}, V_{>_t}^{(g-1)}, j_t^{(g-1)}, z_t^{(g-1)}, n_t^{(g-1)}, \Theta^{(g-1)}, Y, P) \)
- **Q-measure parameters**: \( p(\Theta_M^{(g)}|V^{(g-1)}, \Theta^{(g-1)}, Y, P) \)
- **Pricing equation parameters**: \( p(\Theta_P^{(g)}|V^{(g-1)}, \Theta^{(g-1)}, P) \)
- **P measure parameters**: \( p(\Theta_P^{(g)}|V^{(g-1)}, \Theta_M^{(g-1)}, Y) \)
- **Pricing error variances**: \( p(\Theta_E^{(g)}|V^{(g-1)}, \Theta_M^{(g-1)}, \Theta_P^{(g-1)}, P) \)
- **Jump processes**: \( p(j^{(g)}, z^{(g)}, n^{(g)}|V^{(g-1)}, \Theta^{(g-1)}, Y) \)

The supplemental Appendix D contains a detailed description of how we sample the relevant quantities for each of the sampling blocks.

The empirical results shown later are based on 2,000,000 draws. The first 1,200,000 draws are disregarded as burn-in and of the remaining 800,000, one every 80 draws is retained. We also run 4 additional chains for each model to check the convergence of the estimation.

### 3.2 Choice of Priors

In Table 2 we summarize the priors we use by reporting their type of distribution, mean, standard deviation and 95% highest density region for the different elements of \( \Theta \). The priors for most of the elements of \( \Theta_M \) and \( \Theta_P \) are uninformative: Gaussian priors with zero mean and large standard deviations. We choose conjugate Gaussian priors for those parameters in \( \Theta_P \) that can be sampled directly from their conditional posterior. Moreover, we use the same mean and variance for these parameters under \( \mathbb{P} \)- and \( \mathbb{Q} \)-measures to avoid imposing prior information about the sign and magnitude of the risk premia. For convenience, we do not impose the stationarity or boundary nonattainment conditions explicitly through priors. We nonetheless impose admissibility conditions, e.g., sign or range restrictions on parameters that appear in the diffusion, jump sizes, and jump intensities. To disentangle Brownian increments from jumps, i.e., to reflect the nature of jumps as large and infrequent changes in returns and volatility factors, we set slightly more informative priors for jump size parameters such that they place small probability in small jumps, as in Eraker et al. (2003).
Finally, the choice of an Inverse Gamma prior for $s^2$ allows us to sample directly from the conditional posterior of that parameter.

### 3.3 Monte Carlo Simulations

In this subsection we discuss the simulation results for $A_0(2)$, $A_1(2)$ and $A_2(2)$. To save space, we omit the results for their special cases: $A_{i+0}(2)$, $A_{i+1}(2)$, and $A_{i+2}(2)$. For each model, we simulate 20 samples that share the same length and characteristics with the real unbalanced variance swap panel. We also use parameter values close to their estimates.

Tables 3 and 4 report the true parameter values, as well as the bias, standard deviation, and 95% high probability regions of the posteriors based on 40,000 draws, from which one every 100 is retained. The parameters in $\Theta_M$ and $\Theta_P$ are precisely estimated. In contrast, the posterior distributions of the drift parameters in $\Theta_P$ have a large dispersion around their true values, an expected feature given that the sample period is less than two decades. Similarly, the jump parameters in $\Theta_P$ have much lower precision than the corresponding ones in $\Theta_M$ because only a few jumps occur on average per year, whereas the variance swap rates contain substantial information about jump parameters in $\Theta_M$. For the same reason, the estimates of the pricing error variances are also very precise.

### 4 Empirical Results

#### 4.1 Data

We estimate all six models using daily S&P 500 index returns and variance swap rates with six different maturities (2, 3, 6, 9, 12, and 24 months) over the period from January 4, 1996 to January 11, 2013. The number of daily observations is 4,276, excluding weekends and holidays. Due to restrictions from our data source, the sample is constructed as follows: it contains data on variance swap mid-quotes on 5 maturities (2, 3, 6, 12 and 24 months) from an anonymous U.S. bank over the period January 4, 1996 to March 30, 2007, whereas the the second dataset, which belongs to the same source, covers the period starting from January 2, 2001 to January 11, 2013 with 4 maturities (3, 6, 9, and 12 months). Overall, we have an unbalanced panel of variance swaps over the past 17 years.\(^\text{14}\)

Figure 2 presents the variance swap rates for different maturities along with S&P 500 index returns over the whole sampling period. During the first half of the sample, they are characterized by a significantly higher market volatility which is due in part to the Asian, Russian and LTCM crises. After the “quiet” period between the years 2004 and 2007, during which the market witnessed

\(^\text{14}\)While we do not have data on 1-month variance swaps, which may help identify jumps from stochastic volatility and which are informative about risk premia as shown by Andersen et al. (2016), we use the squared-VIX in out-of-sample studies and find very small pricing errors.
a persistently low volatility level and an increasing trend of stock prices, there is a sharp elevation
in volatility due to the 2007-08 financial meltdown, followed by two spikes related to the European
sovereign debt crisis and the U.S. debt-ceiling confrontation.

The bottom panel of Figure 2 highlights the changes in the slope of the variance term struc-
ture. For most of the sample, the variance term structure is upward sloping, whereas in the middle
and aftermath of crises, the term structure switches to a downward sloping shape, suggesting that
volatility is expected to decrease towards its long-term mean level. The fact that the term structure
is not in perfect tandem with the variance level suggests the necessity of incorporating at least one
additional factor that captures the slope of the term structure of variance. In the next section, we
document a few empirical facts that surface from our analysis.

4.2 Model Performance

4.2.1 Choice of Models

We perform principal component analysis for the balanced panel with 1558 observations. The first
three eigenvalues account for 97.80%, 99.69%, and 99.91% of the total variations. The corresponding
eigenvectors suggest that the first principal component is related to level shifts in the variance curve
while the second one captures changes in the slope of the curve. Nevertheless, the convexity effect
seems negligible for variance swap data, as the contribution of the third principal component is tiny.

While our principal component analysis suggests that one factor can explain a good deal of
variation in variance swap prices, Aït-Sahalia et al. (2015) find strong support for two-factor compared
to one-factor models through a more formal comparison based on a model selection criterion using the
likelihood ratio for non-nested models. Moreover, the variation in the term structure also suggests
at least two factors: one for the short-end of the curve and the other one for the long-end or,
intuitively, the slope. For those reasons, in what follows we focus our analysis on alternative models
that include two volatility factors.\footnote{Andersen et al. (2015b) recently propose a three-factor parametric model for S&P 500 options. It would be
interesting to compare the performance of two-factor non-affine models with that of three-factor affine models and
assess whether adding a third factor would improve the pricing performance of affine models. Clearly, this question
should be addressed using S&P 500 options or VIX options. We leave such an exercise for future work.}

In light of the evidence of negative jumps highlighted in Figure 1, we estimate models $A_0(2)$ and $A_1(2)$, each of which allows for negative jumps through at least
one Ornstein-Uhlenbeck factor, as well as model $A_2(2)$, despite its inadequacy of capturing negative
volatility jumps. For additional comparison, we also fit $A_0^+(2)$ and $A_1^+(2)$, which are special cases
of non-affine models without downward jumps, as well as $A_2(2)$, the affine model prevalent in the
literature. This comparison will shed light on the importance of downward volatility jumps and the
advantage of non-affine models over the affine one.
4.2.2 Estimation Results

In Tables 5, 6, 7, and 8, we report the posterior means and standard deviations of the parameter vectors $\Theta_M$, $\Theta_{\Pi}$, $\Theta_P$, and $\Theta_E$ for all six models. Parameters are defined in annual terms following the convention in the empirical option pricing literature.

We first discuss the estimates of $\Theta_M$ and $\Theta_P$. As can be seen from the $\kappa^{\mathbb{Q}}$'s estimates, the first factor $X_1$ mean-reverts much faster than the second factor $X_2$ does. $\kappa^{\mathbb{Q}}_{11}$ is closer to the values found in the option pricing literature under the pricing measure. Also, the mean reversion parameter of $X_2$ under both measures is very low, around 0.2 for both $A_0(2)$ and $A_1(2)$, implying that shocks to $X_2$ have a half life of several years. Moreover, for both $A_0(2)$ and $A_1(2)$ models, positive jump sizes are larger under $\mathbb{Q}$ than under $\mathbb{P}$, while negative jump sizes are smaller in magnitude under $\mathbb{Q}$. This indicates that both types of jumps are priced, and that negative jumps have positive risk premia. The lower panel of Table 5 contains the corresponding summary statistics of the posterior distribution of the pricing equation parameters in $\Theta_{\Pi}$. Not surprisingly, these parameters are estimated with high precision given that they are identified from prices. Moreover, the percentage of volatility explained by the exponential component dominates, accounting for on average over 90% across all models. This provides strong evidence in favor of log-type volatility models against the affine volatility models or the quadratic ones advocated by Filipovic et al. (2015).

The persistence of volatility and its zero-lower bound together impose a substantial barrier for fitting square-root models. As a result, it is not surprising to find that the estimated parameters violate the boundary nonattainment conditions given explicitly in Section 2.3 and Appendix C for both $A_2(2)$ and $\tilde{A}_2(2)$ models. In contrast, the parameter estimates from the other four models satisfy all the required conditions.

Figure 3 provides the time series of the estimated factors for all six models. For $\tilde{A}_2(2)$, we plot the estimated factors in log scales for comparison. Interestingly, the extracted factors (or their logarithms) share similar patterns across all models. Although the levels of these factors are not the same due to the sign restrictions, the similarity suggests that the extracted patterns are very robust. The factors from the $\tilde{A}_2(2)$ model appear noisier over 2004 - 2007 when volatility is persistently small. This is because the variance swap rates are not sensitive to the magnitude of volatility factors, when they are at a extremely low level as required by the fitting and specification of this model.

Overall, the above analysis suggests that models with at least one Ornstein-Uhlenbeck factor are more desirable than those with two square-root factors.

4.2.3 Assessment of the $\mathbb{Q}$-Measure Dynamics

We then analyze the properties of pricing error variances $\Theta_E$, from which we can intuitively learn about the performance of different models. Ideally, better models tend to produce smaller pricing errors, given similar amounts of unknown parameters. It turns out that we find very similar perfor-
mances across the non-affine models. As shown in Tables 7 and 8, the estimated standard deviations of pricing errors are around 0.37, 0.07, 0.21, 0.25, 0.06, and 0.23 for the 6 maturities, respectively, with $A_0(2)$ and $A_1(2)$ being slightly better. For the affine model, the corresponding numbers are larger across the board: 0.38, 0.08, 0.23, 0.26, 0.06, and 0.27. In short, non-affine models achieve a better fitting parsimoniously.

We then compare the out-of-sample performance of model-fitting using the VIX. The out-of-sample study here is cross-sectional, instead of based on time-series forecasting, as is common for models with latent factors, see Piazzesi (2010). The VIX is constructed by the CBOE using option portfolios, which often coincides with how variance swap contract writers hedge their risk exposure. As a result, it is expected that the time-series of the squared VIX (scaled by 100) and the model-predicted 1-month swap rates present very similar patterns. The results are shown in Figure 4. The out-of-sample performance compared to the VIX is almost identical across these models, with correlations as high as 0.89 for all.

### 4.2.4 Assessment of the $\mathbb{P}$-Measure Dynamics

Having witnessed quite similar results across the $\mathbb{Q}$-measure performance of these non-affine models, in particular the $A_0(2)$ and $A_1(2)$ models, we then move on to their $\mathbb{P}$-measure performance by investigating the time series of the estimated spot variance $\sigma_t^2$, which can be decomposed into jumps and Brownian shocks. We decompose changes of estimated spot variances for all six models in Figure 5, respectively, which sheds light on some new evidence on model selection among two-factor volatility models.

The changes of the spot variance are very similar to changes of the squared VIX in Figure 1 across all models (hence we omit the figure of the changes for reasons of space), but their decompositions are strikingly different. We highlight on the figure three downward volatility jumps associated with the three news events mentioned at the beginning of the introduction. Obviously, except for the $A_0(2)$ and $A_1(2)$ models, the rest cannot capture any of these downward volatility jumps, so that they are misidentified as large Brownian shocks. Although the $A_1(2)$ model is able to capture negative jumps in one of its factors, the Ornstein-Uhlenbeck factor, this factor turns out to be slow mean-reverting and highly persistent, which cannot accommodate those jumps that perhaps only affect short-term volatility levels. As a result, several significantly downward volatility changes are attributed to Brownian shocks, as the square-root factor does not permit negative jumps. As previously mentioned, for the $A_2(2)$ model, the Feller constraint is binding for both the $\mathbb{P}$- and $\mathbb{Q}$-measure dynamics. In contrast, the $A_0(2)$ model can accommodate jumps in both the short-term and long-term factors, so that those short-term jumps missed by $A_1(2)$ are captured, and that the two

---

16Strictly speaking, the difference between the squared VIX and variance swap rates is related to the higher order moments of price jumps, as Aït-Sahalia et al. (2014) point out.
components are well-separated. Overall, the $A_0(2)$ model is clearly more desirable from the evidence in Figure 5.

We hence employ the $A_0(2)$ model in the following empirical analysis. While we can further improve the $A_0(2)$ model by augmenting it with a time-varying intensity factor, we choose not to do so for parsimony and in order to avoid overfitting.

4.3 Economic Interpretation of Volatility Components

4.3.1 Volatility Factors

We now provide an interpretation for the latent volatility factors, before addressing how jumps are related to them. To do so, we conduct a regression analysis trying to link the identified latent factors with factors of economic fundamentals. We select two credit variables, including the daily TED spread, calculated as the difference between the three-month LIBOR and the three-month T-Bill interest rate and the default spread (DEF), calculated as the difference between the monthly Moody’s AAA and BAA corporate bond yield. We also obtain two monthly macroeconomic factors from the Federal Reserve’s website: the Chicago Fed National Activity Index (CFI), constructed from 85 monthly indicators of economic activity; and industrial production growth (IPG), as suggested by Pástor and Veronesi (2013) and Adrian and Rosenberg (2008). We also include the daily term spread (TERM), i.e. the difference between the yields on the 10-year and 3-month Treasury securities. We also add one monthly liquidity factor (LIQ), the innovation of the aggregate liquidity from Pástor and Stambaugh (2003). To identify potential policy risk that may be related to volatility jumps, we add the policy news index (POL) constructed by Baker et al. (2013). Finally, we construct the market skewness factor as it is shown to be important for the short-term component by Adrian and Rosenberg (2008).

We consider one-by-one simple regressions of the posterior mean of each factor $X_i$ of $A_0(2)$, sampled at the end of each month from 1996 to 2012, on the innovation of each covariate given above, as well as the lagged value of the posterior mean of $X_i$ from the past month:

$$X_{i,t} = \beta_0 + \beta_1 Z_{j,t} + \beta_2 X_{i,t-1} + \varepsilon_t,$$

with $Z_{j,t}$ being the innovation of the $j$th covariate. For POL and TERM, we use ARIMA(1,1,0) innovations, as the Dickey Fuller tests fail to reject the unit-roots in our sample period. For IPG, we use the AR(3) innovation, following Adrian and Rosenberg (2008). For the rest of the covariates, we use AR(1) innovations. The results are identical when using other regression specifications.

We also consider a multiple time-series regression for all the innovations of the covariates plus the lagged posterior mean of $X_i$:

$$X_{i,t} = \beta_0 + \beta_1 \text{DEF}_t + \beta_2 \text{TED}_t + \beta_3 \text{TERM}_t + \beta_4 \text{LIQ}_t + \beta_5 \text{POL}_t + \varepsilon_t,$$
\[ + \beta_6 \text{SKEW}_t + \beta_7 \text{ExM}_t + \beta_8 \text{IPG}_t + \beta_9 \text{CFI}_t + \beta_{10} X_{i,t-1} + \varepsilon_t. \]  

(11)

Tables 9 and 10 provide regression results for \(X_1\) and \(X_2\), respectively. Table 9 suggests that the time variation of short-term volatility factor \(X_1\) can be explained by credit risk, liquidity risk, and policy news, in addition to the excess returns. The signs of each coefficient agree with the intuition that short-term volatility rises if risk or uncertainty increases. When stacking these covariates into the multiple regression, policy news, excess market returns, and lagged values of \(X_1\) subsume the rest of the covariates. As for the second volatility factor \(X_2\), default risk, term premia, and excess market return become significant with all covariates included. The AR(1) coefficient reported in Table 10 confirms that \(X_2\) is much more persistent than \(X_1\). It is worth mentioning that our business cycle variables are not significant, for potentially two reasons. First, the sample period is as short as 17 years, which does not accommodate many business cycles. Secondly, the longest maturity of our variance swaps is 2 years, so that the “long” term factor extracted here may be regarded as the “median” term in macroeconomics, so that business cycle variables are less important. The results are very similar for the other non-affine models, an expected feature in light of Figure 3.

### 4.3.2 Volatility Jumps

Regarding volatility jumps, we find that downward volatility jumps are as common as positive ones, and that they are often associated with a resolution of policy uncertainty. Apart from the three news headlines mentioned in the introduction, we highlight 29 additional days in Table 11, which are clearly related to some policy news, out of the 40 days with largest downward volatility jumps. From this table, we find that the majority of large downward volatility jumps are caused by changes of current monetary policy or clear indications about future monetary policy, despite few jumps being relevant to fiscal policy, all of which may help comfort investors.

To understand how volatility jumps originate, we construct measures of news surprises based on surveys of economists’ expectations on 18 economic indicators from Bloomberg. The detailed information about the categories, the announcement time, and the frequency of these news events are given in Table 12. We proxy news surprise as the differences between the actual news release and the survey expectations:

\[ \text{News Surprise}_t = \text{Announced Quantity}_t - \text{Median of Expectations}_t. \]  

(12)

The news surprises of economic indicators are treated as the control variables since they are expected to produce jumps in S&P 500 returns and other markets, e.g. Beechey and Wright (2009) and Faust and Wright (2009). To proxy the resolution of policy uncertainty, we use the schedules of FOMC

\footnote{Our dataset does not contain standard errors of the survey estimates. It can be expected that the regression results would become more significant if standard errors were used to scale the differences.}
and ECB meetings, as well as the speech schedules of the Federal Reserve’s Chairman. Although the schedules are usually pre-announced and the target interest rates do not change often, the minutes, statements or press conferences after the FOMC meetings are informative about monetary policy decisions, and this information could be unpredictable.

We regress the magnitudes of positive and negative volatility jumps onto the magnitude of macroeconomic news surprises for each volatility factor, respectively: \[ |\text{Positive/Negative jump size of } X_{j,t}| = \beta_{j,0}^{+/-} + \sum_{i=1}^{21} \beta_{j,i}^{+/-} |s_{i,t}| + \varepsilon_{j,t}^{+/-}, \quad \text{for } j = 1, 2. \] (13)

where coefficients with +/− correspond to regressions with positive and negative jumps, respectively, and \( s_{i,t} \) is the \( i \)-th news surprise at time \( t \). If there is no such news event on day \( t \), then \( s_{i,t} \) is set to 0. All news surprises are rescaled so that they all have time-series standard deviation equal to 1.

In addition, we also run two similar regressions for jumps in \( \sigma_t^2 \):

\[ |\text{Positive/Negative jump size of } \sigma_t^2| = \beta_0^{+/-} + \sum_{i=1}^{21} \beta_i^{+/-} |s_{i,t}| + \varepsilon_t^{+/-}. \] (14)

The results are provided in Table 13. We find that negative volatility jumps in \( X_1 \) are mainly caused by FOMC meetings and Federal Reserve Chairman’s speeches, whereas other volatility jumps are the result of surprising news about employment, consumer spending, and national output. This conforms with our conjecture and earlier event studies suggesting that negative volatility jumps are highly correlated with the resolution of policy uncertainty. In addition, such jumps mostly affect the short-term volatility level, suggesting that not all policy measures have a significant impact on long-term uncertainty.

To further analyze the short-term versus long-term impact of policy news, we conduct event studies by investigating the identified jumps in \( X_1 \) and \( X_2 \) separately, for the three events mentioned in the introduction. It turns out that not all of the three policy news we highlighted have a strong long-term impact on volatility, despite their significant influences on the short-term volatility level with magnitudes (posterior medians) as high as 0.37, 0.24, and 0.29, respectively. Regarding Europe’s Debt Crisis, the unprecedented emergency loan plan unveiled on May 10, 2010 hit the long-term volatility level by -0.09, although a larger downward jump of magnitude 0.29 in \( X_2 \) came two days later, as investors digested the details of the $1 trillion European aid package. Another long-term volatility jump (-0.24) came more than one year later, after European Union leaders agreed to expand Europe’s bailout fund and take major losses on Greek bonds at the end of marathon talks on October 27, 2011. Also, the Federal Reserve’s FOMC statement on August 9, 2011 decreases the long-term uncertainty level by -0.10, partially because of the additional “forward guidance” information on how long the Committee expects to keep the target for the federal funds rate exceptionally low.

\[ ^{18} \text{We use the posterior medians of the identified jumps so as to obtain a more sparse time series of jumps.} \]
In contrast, the news about the fiscal cliff do not show a significant impact, as investors remained cautious about the deal. Indeed, Congress failed to reach an agreement on spending cuts and the sequestration was delayed until March 2013 as part of the American Taxpayer Relief Act of 2012, passed by Congress on January 1, 2013.

4.4 Variance Risk Premia

Now we investigate the pricing implications of downward volatility jumps. Comparing the estimates in Table 5, positive volatility jumps have larger magnitudes under the \( \mathbb{Q} \)-measure than under the \( \mathbb{P} \)-measure, whereas negative jumps have smaller magnitudes under the \( \mathbb{Q} \)-measure. This indicates that positive volatility jumps have a negative price of risk, whereas negative jumps have a positive price of risk. As a result, the total variance risk premia would be overestimated if downward volatility jumps are excluded.

To gauge the economic importance of the bias, we compare the total variance risk premia implied from \( A_0(2) \) with the estimates from \( A_0^+(2) \) model that does not include downward volatility jumps. The difference in the amount of variance risk premia can be interpreted as the bias due to model misspecification. We define variance risk premia in the same way as was introduced in Bollerslev et al. (2009), Carr and Wu (2009), and Todorov (2010):

\[
VRP(t, \tau) = \frac{1}{\tau} \left\{ \mathbb{E}_t^\mathbb{P} \left[ (Y_t, Y_{t+\tau}) \right] - \mathbb{E}_t^\mathbb{Q} \left[ (Y_t, Y_{t+\tau}) \right] \right\}.
\]

We plot in Figures 6 the time series of the term structure of variance risk premia implied from the \( A_0(2) \) and \( A_0^+(2) \) models, respectively. The plots suggest that variance risk premia are mostly negative, with confidence bands not including zero, and countercyclical, (i.e., they become even more negative in bad times). For example, the lower troughs in the figure are associated with the 1997-98 Asian crisis, the dot-com bubble, the recent financial meltdown, and the European and U.S. debt crises, suggesting that investors require more compensation for bearing variance risk during difficult times.

However, what we find more interesting is that at the inception of crises, the estimates of variance risk premia become positive or at least insignificantly different from zero for short periods of time based on \( A_0(2) \), whereas for \( A_0^+(2) \) the estimates are always negative. The finding based on \( A_0(2) \) agrees with most model-free estimates, see Table 1.\(^19\) Our result suggests that \( A_0(2) \) model produces better estimates of the \( \mathbb{P} \)-measure expectations than the \( A_0^+(2) \) model, and that the downward volatility jumps allow for more flexible risk premia specification.

The risk premia should have been more negative, rather than being statistical insignificant or even positive, if the variance risk were priced by the representative agent in a rational expectation.

\(^{19}\)It is worth mentioning that model-free estimates still rely on affine volatility forecasting models to estimate the conditional variance under the objective measure.
framework. To resolve this puzzle, one may resort to a model in which investors have heterogeneous beliefs, so that the variance risk premia could be either positive or negative depending on the prevalent view of the market. In that regard, Bakshi et al. (2015) suggest a U-shape volatility pricing kernel by exploring the link between the monotonicity of the pricing kernel and returns on VIX option portfolios. They further build a stylized model with heterogeneity in beliefs to account for the U-shape. On the other hand, using traders’ position data from CFTC, Cheng (2015) finds evidence indicating that time-varying demand from heterogeneous investors affects premia embedded in VIX futures, and that the low demand from dealers and unlevered asset managers help explain the low premia during these periods.

4.5 Robustness Checks

For robustness, on the first panel of Figure 7 we compare the estimated daily volatility path using the $A_0(2)$ model, with realized volatility computed from high frequency data using the square-root of the truncated sum of squared 5-minute returns of SPY, the SPDR S&P 500 ETF. A truncation is necessary to get rid of those jumps in returns, see, e.g., Aït-Sahalia and Jacod (2014). We proxy the realized profit and loss (P&L) earned by investing in a hypothetical variance swap with 1-month time-to-maturity, using the VIX$^2$ minus the realized variance of the corresponding trading days. This P&L proxy, a model-free measure of variance risk premia, as shown from the bottom panel of Figure 7 also tends to be positive at the inception of crises.

5 Conclusion

Motivated by recent news headlines about the dramatic changes of the VIX following the announcements of policy makers, our systematic investigation examines the sudden declines of market volatility. We find downward volatility jumps to be as common as positive ones, and that the majority of them are caused by FOMC announcements and the speeches of the Federal Reserve Chairman, showing the impact of Central Bank intervention, whereas only a small portion of downward volatility jumps are responses to surprising news about employment, consumer spending, and national output. This conforms with earlier event studies suggesting that negative volatility jumps are highly correlated with the resolution of policy uncertainty. Moreover, we find that while such jumps affect the short-term volatility level, not all of them have a significant impact on long-term volatility.

Our results indicate that positive volatility jumps have a negative price of risk, whereas negative jumps have a positive price of risk. In other words, the protection offered by the Federal Reserve is indeed priced by investors. While a model without downward volatility jumps may be able to capture part of the jump risk premia through other components in the model, the interpretation would be entirely different. Unlike the premia for persistent shocks, jump risk premia are compensation for
tail events. In particular, downward volatility jumps are associated with important policy measures, reflecting the protection offered by the Federal Reserve.

In order to model downward volatility jumps, this paper introduces a new non-affine modeling framework which extends the classification and characterization of term structure models to allow for jumps. Our canonical models nest square-root factors, Ornstein-Uhlenbeck factors, pure-jump factors with state-dependent intensity, self-exciting jumps, Lévy jumps, etc. We find that the log-type volatility model, which has been favored by financial econometricians in the past, with two Ornstein-Uhlenbeck factors and double exponential jumps yields the best performance in fitting variance swap prices and the volatility dynamics. Such a model can also be used to investigate S&P 500 options, which we leave for future work.

References


Baumohl, B. (2010), *The Secrets of Economic Indicators,* Prentice Hall.


<table>
<thead>
<tr>
<th>Article</th>
<th>Data</th>
<th>Period</th>
<th>Frequency</th>
<th>Model</th>
<th>VRP 1M</th>
<th>VRP 2M</th>
<th>VRP 1Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aït-Sahalia et al. (2014)</td>
<td>VS</td>
<td>1996 - 2010</td>
<td>Daily</td>
<td>AJD</td>
<td>[-7%, 0%]</td>
<td>[-10%, -0.5%]</td>
<td></td>
</tr>
<tr>
<td>Amengual (2008)</td>
<td>VS</td>
<td>1996 - 2007</td>
<td>Daily</td>
<td>AJD</td>
<td>[-5%, 0%]</td>
<td>[-7%, -0.5%]</td>
<td></td>
</tr>
<tr>
<td>Bollerslev et al. (2011)</td>
<td>SPO</td>
<td>1990 - 2003</td>
<td>Intraday</td>
<td>MF</td>
<td>[-20%, 5%] *</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corradi et al. (2013)</td>
<td>VIX</td>
<td>1950 - 2006</td>
<td>Monthly</td>
<td>AD</td>
<td>[-30%, 8%] *</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fan et al. (2016)</td>
<td>VIX</td>
<td>2006 - 2011</td>
<td>Intraday</td>
<td>MF</td>
<td>[-30%, 20%] *</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fusari and Gonzalez-Perez (2012)</td>
<td>SPO</td>
<td>1996 - 2010</td>
<td>Daily</td>
<td>Log-OU</td>
<td>[-20%, 3%]</td>
<td></td>
<td>[-23%, 1%]</td>
</tr>
<tr>
<td>Zhou (2009)</td>
<td>VIX</td>
<td>1990 - 2008</td>
<td>Intraday</td>
<td>MF</td>
<td>[-200%, 40%]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 1: Selection of Literature Reporting Variance Risk Premia Estimates**

Note: This table collects estimates from papers that report either the point estimates or time series plots of variance or volatility risk premia. In the Data column, “VS” denotes variance swaps, “SPO” denotes S&P 500 options, and “VIX” is the CBOE volatility index. In the Model column, “AD” refers to affine diffusion, “AJD” denotes affine-jump diffusion, “MF” means model-free, and “Log-OU” is a log-affine process with two Ornstein-Uhlenbeck factors. The columns of VRP provide the approximate bounds of the estimated time series of risk premia, with 1M, 2M and 1Y referring to the 1-month, 2-month, and 1-year time-to-maturities. Most MF methodologies provide positive estimates of variance risk premia for certain periods of time, so that their upper bounds are positive.

* provides volatility risk premia.

** gives a point estimate with the standard error provided in the brackets.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Distribution</th>
<th>Mean</th>
<th>Stdev</th>
<th>HPD 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{21}$</td>
<td>Gamma(2, 2)</td>
<td>4.000</td>
<td>(2.828)</td>
<td>[0.484, 11.144]</td>
</tr>
<tr>
<td>$\lambda_1$, $\lambda_2$, $\mu_1$, $\mu_2$</td>
<td>N(0, 20^2)</td>
<td>1.000</td>
<td>(10.00)</td>
<td>[-18.607, 20.607]</td>
</tr>
<tr>
<td>$\beta_{1+}$, $\beta_{1-}$, $\beta_{2+}$, $\beta_{2-}$</td>
<td>Inv.Gamma(3, 2)</td>
<td>0.250</td>
<td>(0.250)</td>
<td>[0.069, 0.807]</td>
</tr>
<tr>
<td>$q_1$, $q_2$</td>
<td>Uniform(0, 1)</td>
<td>0.500</td>
<td>(0.289)</td>
<td>[0.025, 0.975]</td>
</tr>
<tr>
<td>$\rho_1$, $\rho_2$</td>
<td>Uniform(B(0, 1))</td>
<td>0.000</td>
<td>(0.580)</td>
<td>[0.050, 0.950]</td>
</tr>
<tr>
<td>$l_0$, $l_{11}$, $l_{12}$</td>
<td>Gamma(1, 1)</td>
<td>1.000</td>
<td>(1.000)</td>
<td>[0.025, 3.687]</td>
</tr>
<tr>
<td>$\Pi_0$, $\Pi_1$‘s, $\Pi_2$‘s, $\Pi_3$, $\Pi_4$‘s</td>
<td>N(0, 5^2)</td>
<td>0.000</td>
<td>(5.000)</td>
<td>[-9.797, 9.797]</td>
</tr>
<tr>
<td>$s^2$</td>
<td>Inv.Gamma(1, 1)</td>
<td>1.000</td>
<td>(1.000)</td>
<td>[0.287, 1.385]</td>
</tr>
<tr>
<td>$\sigma_J^2$</td>
<td>Gamma(0.02, 0.5)</td>
<td>0.010</td>
<td>(0.071)</td>
<td>[0.000, 0.096]</td>
</tr>
</tbody>
</table>

**Table 2: Prior Distributions**

Note: This table presents the mean, standard deviation, and 95% highest prior density region for the priors we use to implement our estimation procedure. $\text{Gamma}(\alpha, \beta)$ denotes a gamma random variate with parameters $\alpha$ and $\beta$, such that its mean is $\alpha \beta$ and variance $\alpha \beta^2$; $\text{Inv.Gamma}(\alpha, \beta)$ denotes an inverse gamma random variate with parameters $\alpha$ and $\beta$ such that if $X \sim \text{Inv.Gamma}(\alpha, \beta)$ then $X^{-1} \sim \text{Gamma}(\alpha, \beta^{-1})$; $\text{Uniform}(a, b)$ denotes a random variable which is uniformly distributed on the interval $[a, b]$; and $N(\mu, \sigma^2)$ denotes a Gaussian random variate with mean $\mu$ and variance $\sigma^2$. $B(0, 1) = \{\rho \in \mathbb{R}^2 : ||\rho|| \leq 1\}$. Details on models and parameterizations are given in Appendix C.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>$A_0(2)$</th>
<th>$A_1(2)$</th>
<th>$A_2(2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1^o$</td>
<td>$\lambda_2^o$</td>
<td>$\kappa_{11}^o$</td>
<td>$\kappa_{12}^o$</td>
</tr>
<tr>
<td>True</td>
<td>Bias</td>
<td>Stdev</td>
<td>HPD 95%</td>
</tr>
<tr>
<td>6.000</td>
<td>0.000</td>
<td>0.006</td>
<td>[5.983, 6.009]</td>
</tr>
</tbody>
</table>

Table 3: Simulation Results for $\Theta_M$ and $\Theta_{\Pi}$

Note: This table provides a summary of a Monte Carlo simulation exercise with 100 replications for the two examples of two-factor volatility models that we introduce in Appendix C. We report true values, bias and standard deviations across the simulations for $\Theta_M$, the parameters determining the dynamics of the latent factors under the risk-neutral measure, and $\Theta_{\Pi}$, the parameters defining $f$. 
<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\lambda_1^*$</th>
<th>Bias</th>
<th>Stdev</th>
<th>HPD 95%</th>
<th>$\lambda_2^*$</th>
<th>Bias</th>
<th>Stdev</th>
<th>HPD 95%</th>
<th>$\lambda_3^*$</th>
<th>Bias</th>
<th>Stdev</th>
<th>HPD 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa_{11}$</td>
<td>-3.600</td>
<td>-0.260</td>
<td>0.789</td>
<td>[−5.653, −2.547]</td>
<td>-4.000</td>
<td>-0.087</td>
<td>0.728</td>
<td>[−5.611, −2.757]</td>
<td>-5.500</td>
<td>0.012</td>
<td>0.893</td>
<td>[−7.298, −3.801]</td>
</tr>
<tr>
<td>$\kappa_{12}$</td>
<td>4.800</td>
<td>0.011</td>
<td>0.877</td>
<td>[2.965, 6.451]</td>
<td>4.000</td>
<td>-0.209</td>
<td>1.055</td>
<td>[1.574, 5.805]</td>
<td>4.500</td>
<td>-0.505</td>
<td>1.258</td>
<td>[1.447, 6.527]</td>
</tr>
<tr>
<td>$\beta_{1+}$</td>
<td>0.200</td>
<td>-0.002</td>
<td>0.044</td>
<td>[0.119, 0.293]</td>
<td>0.100</td>
<td>0.005</td>
<td>0.018</td>
<td>[0.076, 0.145]</td>
<td>0.100</td>
<td>0.004</td>
<td>0.015</td>
<td>[0.077, 0.137]</td>
</tr>
<tr>
<td>$\beta_{1-}$</td>
<td>0.180</td>
<td>0.009</td>
<td>0.041</td>
<td>[0.122, 0.284]</td>
<td>0.160</td>
<td>0.004</td>
<td>0.043</td>
<td>[0.084, 0.254]</td>
<td>0.120</td>
<td>-0.002</td>
<td>0.022</td>
<td>[0.080, 0.163]</td>
</tr>
<tr>
<td>$\beta_{2+}$</td>
<td>0.200</td>
<td>0.008</td>
<td>0.044</td>
<td>[0.136, 0.309]</td>
<td>0.270</td>
<td>-0.004</td>
<td>0.051</td>
<td>[0.175, 0.373]</td>
<td>0.100</td>
<td>0.000</td>
<td>0.003</td>
<td>[0.003, 0.016]</td>
</tr>
<tr>
<td>$\beta_{2-}$</td>
<td>0.200</td>
<td>0.005</td>
<td>0.032</td>
<td>[0.148, 0.274]</td>
<td>0.010</td>
<td>0.000</td>
<td>0.002</td>
<td>[0.006, 0.015]</td>
<td>0.010</td>
<td>0.000</td>
<td>0.003</td>
<td>[0.005, 0.015]</td>
</tr>
<tr>
<td>$\mu_1^+$</td>
<td>-0.700</td>
<td>0.017</td>
<td>0.041</td>
<td>[−0.745, −0.578]</td>
<td>-0.600</td>
<td>-0.009</td>
<td>0.023</td>
<td>[−0.652, −0.564]</td>
<td>-0.450</td>
<td>-0.007</td>
<td>0.017</td>
<td>[−0.490, −0.424]</td>
</tr>
<tr>
<td>$\mu_2^+$</td>
<td>-0.450</td>
<td>0.018</td>
<td>0.040</td>
<td>[−0.510, −0.354]</td>
<td>-0.500</td>
<td>0.033</td>
<td>0.036</td>
<td>[−0.535, −0.390]</td>
<td>-0.750</td>
<td>0.002</td>
<td>0.012</td>
<td>[−0.772, −0.724]</td>
</tr>
<tr>
<td>$\mu_3^+$</td>
<td>0.070</td>
<td>0.000</td>
<td>0.046</td>
<td>[−0.013, 0.169]</td>
<td>0.150</td>
<td>-0.010</td>
<td>0.077</td>
<td>[−0.008, 0.293]</td>
<td>0.120</td>
<td>-0.021</td>
<td>0.059</td>
<td>[−0.021, 0.206]</td>
</tr>
</tbody>
</table>

Table 4: Simulation Results for $\Theta_P$ and $\Theta_E$

Note: This table provides a summary of a Monte Carlo simulation exercise with 100 replications for the two examples of two-factor volatility models that we introduce in Appendix C. We report true values, bias and standard deviations across the simulations for the additional parameters that characterize the $P$-dynamics, $\Theta_P$, and the pricing error variances, summarized in $\Theta_E$. 
<table>
<thead>
<tr>
<th>Parameter</th>
<th>$A_0(2)$</th>
<th>$A_1(2)$</th>
<th>$A_2(2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1^0$</td>
<td>Mean</td>
<td>Stdev</td>
<td>HPD 95%</td>
</tr>
<tr>
<td>$\lambda_2^1$</td>
<td>7.519</td>
<td>0.117</td>
<td>[7.369, 7.717]</td>
</tr>
<tr>
<td>$\kappa_1^0$</td>
<td>-4.322</td>
<td>0.048</td>
<td>[-4.390, -4.227]</td>
</tr>
<tr>
<td>$\kappa_2^1$</td>
<td>2.001</td>
<td>0.063</td>
<td>[1.918, 2.116]</td>
</tr>
<tr>
<td>$\kappa_2^2$</td>
<td>-0.225</td>
<td>0.005</td>
<td>[-0.233, -0.213]</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>0.971</td>
<td>0.062</td>
<td>[0.896, 1.086]</td>
</tr>
<tr>
<td>$\beta_{21}^0$</td>
<td>0.171</td>
<td>0.004</td>
<td>[0.164, 0.178]</td>
</tr>
<tr>
<td>$\beta_{21}^1$</td>
<td>0.056</td>
<td>0.010</td>
<td>[0.037, 0.071]</td>
</tr>
<tr>
<td>$\beta_{21}^2$</td>
<td>0.598</td>
<td>0.037</td>
<td>[0.580, 0.616]</td>
</tr>
<tr>
<td>$\beta_{21}^3$</td>
<td>0.310</td>
<td>0.018</td>
<td>[0.297, 0.324]</td>
</tr>
<tr>
<td>$\beta_{21}^4$</td>
<td>0.046</td>
<td>0.004</td>
<td>[0.038, 0.054]</td>
</tr>
<tr>
<td>$\beta_{21}^5$</td>
<td>0.179</td>
<td>0.035</td>
<td>[0.168, 0.190]</td>
</tr>
<tr>
<td>$\mu_0$</td>
<td>7.263</td>
<td>0.572</td>
<td>[6.228, 8.326]</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>3.310</td>
<td>0.151</td>
<td>[0.095, 0.553]</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>0.013</td>
<td>0.001</td>
<td>[0.010, 0.015]</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.013</td>
<td>0.001</td>
<td>[0.010, 0.015]</td>
</tr>
</tbody>
</table>

Table 5: Posterior Estimates of $\Theta_M$ and $\Theta_\Pi$ for $A_0(2)$, $A_1(2)$, and $A_2(2)$

Note: This table presents the posterior estimates for $\Theta_M$, the parameters determining the dynamics of the latent factors under the risk-neutral measure, and $\Theta_\Pi$, the parameters defining $f$ across all models. We report the mean, the standard deviation, and the 95% highest posterior density intervals for the $A_0(2)$, $A_1(2)$ and $A_2(2)$. We use daily data on variance swaps from January 4, 1996 to January 11, 2013. The number of daily observations is 4,276, excluding weekends and holidays.
Table 6: Posterior Estimates of $\Theta_M$ and $\Theta_H$ for $A_0^+$ (2), $A_1^+$ (2), and $\tilde{A}_2$ (2)

Note: This table presents the posterior estimates for $\Theta_M$, the parameters determining the dynamics of the latent factors under the risk-neutral measure, and $\Theta_H$, the parameters defining $f$ across all models. We report the mean, the standard deviation, and the 95% highest posterior density intervals for the $A_0^+$ (2), $A_1^+$ (2), and $\tilde{A}_2$ (2). We use daily data on variance swaps from January 4, 1996 to January 11, 2013. The number of daily observations is 4,276, excluding weekends and holidays.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>$A_0(2)$ Mean</th>
<th>$A_0(2)$ Stdev</th>
<th>$A_0(2)$ HPD 95%</th>
<th>$A_1(2)$ Mean</th>
<th>$A_1(2)$ Stdev</th>
<th>$A_1(2)$ HPD 95%</th>
<th>$A_2(2)$ Mean</th>
<th>$A_2(2)$ Stdev</th>
<th>$A_2(2)$ HPD 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_0^P$</td>
<td>-0.780</td>
<td>0.249</td>
<td>[-1.278, -0.295]</td>
<td>8.550</td>
<td>1.004</td>
<td>[6.601, 10.530]</td>
<td>8.842</td>
<td>0.962</td>
<td>[7.000, 10.703]</td>
</tr>
<tr>
<td>$\lambda_1^P$</td>
<td>-0.105</td>
<td>0.288</td>
<td>[-0.671, 0.455]</td>
<td>-12.650</td>
<td>1.664</td>
<td>[-15.949, -9.421]</td>
<td>-6.668</td>
<td>1.168</td>
<td>[-8.986, -4.363]</td>
</tr>
<tr>
<td>$\kappa_{11}$</td>
<td>-3.804</td>
<td>0.469</td>
<td>[-4.693, -2.869]</td>
<td>-5.377</td>
<td>0.613</td>
<td>[-6.565, -4.183]</td>
<td>-5.335</td>
<td>0.554</td>
<td>[-6.415, -4.266]</td>
</tr>
<tr>
<td>$\kappa_{12}$</td>
<td></td>
<td></td>
<td></td>
<td>-0.632</td>
<td>0.147</td>
<td>[-0.917, -0.344]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\kappa_{21}$</td>
<td>5.167</td>
<td>0.688</td>
<td>[3.836, 6.509]</td>
<td>5.859</td>
<td>0.922</td>
<td>[4.035, 7.653]</td>
<td>5.671</td>
<td>0.781</td>
<td>[4.142, 7.217]</td>
</tr>
<tr>
<td>$\kappa_{22}$</td>
<td>-0.529</td>
<td>0.147</td>
<td>[-0.815, -0.241]</td>
<td>-0.737</td>
<td>0.177</td>
<td>[-1.077, -0.387]</td>
<td>-0.099</td>
<td>0.165</td>
<td>[-0.432, 0.220]</td>
</tr>
<tr>
<td>$\beta_{1+}$</td>
<td>0.151</td>
<td>0.016</td>
<td>[0.122, 0.186]</td>
<td>0.107</td>
<td>0.009</td>
<td>[0.090, 0.127]</td>
<td>0.148</td>
<td>0.014</td>
<td>[0.123, 0.178]</td>
</tr>
<tr>
<td>$\beta_{2+}$</td>
<td>0.129</td>
<td>0.018</td>
<td>[0.100, 0.168]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_{1}$</td>
<td>0.184</td>
<td>0.021</td>
<td>[0.148, 0.230]</td>
<td>0.302</td>
<td>0.047</td>
<td>[0.224, 0.406]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_{1}$</td>
<td>-0.003</td>
<td>0.002</td>
<td>[-0.006, 0.000]</td>
<td>-0.009</td>
<td>0.001</td>
<td>[-0.010, -0.007]</td>
<td>-0.016</td>
<td>0.002</td>
<td>[-0.019, -0.013]</td>
</tr>
<tr>
<td>$\rho_{2}$</td>
<td>-0.763</td>
<td>0.009</td>
<td>[-0.780, -0.745]</td>
<td>-0.664</td>
<td>0.014</td>
<td>[-0.691, -0.636]</td>
<td>-0.520</td>
<td>0.013</td>
<td>[-0.546, -0.494]</td>
</tr>
<tr>
<td>$\mu_{ij}$</td>
<td>-0.397</td>
<td>0.016</td>
<td>[-0.427, -0.365]</td>
<td>-0.551</td>
<td>0.017</td>
<td>[-0.582, -0.518]</td>
<td>-0.701</td>
<td>0.012</td>
<td>[-0.723, -0.676]</td>
</tr>
<tr>
<td>$s_{M}^2$</td>
<td>0.089</td>
<td>0.034</td>
<td>[0.024, 0.157]</td>
<td>0.156</td>
<td>0.034</td>
<td>[0.090, 0.225]</td>
<td>0.133</td>
<td>0.039</td>
<td>[0.058, 0.210]</td>
</tr>
<tr>
<td>$s_{M}^2$</td>
<td>0.143</td>
<td>0.004</td>
<td>[0.135, 0.151]</td>
<td>0.146</td>
<td>0.004</td>
<td>[0.138, 0.154]</td>
<td>0.147</td>
<td>0.004</td>
<td>[0.139, 0.156]</td>
</tr>
<tr>
<td>$s_{M}^2$</td>
<td>0.004</td>
<td>0.000</td>
<td>[0.004, 0.005]</td>
<td>0.004</td>
<td>0.000</td>
<td>[0.004, 0.005]</td>
<td>0.005</td>
<td>0.000</td>
<td>[0.004, 0.005]</td>
</tr>
<tr>
<td>$s_{M}^2$</td>
<td>0.045</td>
<td>0.001</td>
<td>[0.043, 0.048]</td>
<td>0.045</td>
<td>0.001</td>
<td>[0.043, 0.047]</td>
<td>0.046</td>
<td>0.001</td>
<td>[0.044, 0.048]</td>
</tr>
<tr>
<td>$s_{M}^2$</td>
<td>0.063</td>
<td>0.002</td>
<td>[0.059, 0.066]</td>
<td>0.063</td>
<td>0.002</td>
<td>[0.060, 0.067]</td>
<td>0.064</td>
<td>0.002</td>
<td>[0.061, 0.068]</td>
</tr>
<tr>
<td>$s_{M}^2$</td>
<td>0.004</td>
<td>0.000</td>
<td>[0.003, 0.004]</td>
<td>0.003</td>
<td>0.000</td>
<td>[0.003, 0.004]</td>
<td>0.004</td>
<td>0.000</td>
<td>[0.003, 0.004]</td>
</tr>
<tr>
<td>$s_{M}^2$</td>
<td>0.055</td>
<td>0.002</td>
<td>[0.052, 0.059]</td>
<td>0.053</td>
<td>0.002</td>
<td>[0.050, 0.057]</td>
<td>0.057</td>
<td>0.002</td>
<td>[0.054, 0.061]</td>
</tr>
</tbody>
</table>

**Table 7: Posterior Estimates of $\Theta_P$ and $\Theta_E$ for $A_0(2)$, $A_1(2)$, and $A_2(2)$**

Note: This table presents the posterior estimates for the parameters that characterize the $\mathbb{P}$-dynamics, $\Theta_P$, and the pricing error variances $\Theta_E$. We report the mean, the standard deviation, and the 95% highest posterior density intervals for the $A_0(2)$, $A_1(2)$, and $A_2(2)$. We use daily data on variance swaps from January 4, 1996 to January 11, 2013. The number of daily observations is 4,276, excluding weekends and holidays.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>$A_{0}^{+}(2)$</th>
<th>$A_{1}^{+}(2)$</th>
<th>$\tilde{A}_{2}(2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Stdev</td>
<td>HPD 95%</td>
<td>Mean</td>
</tr>
<tr>
<td>$\lambda_{1}^{0}$</td>
<td>-0.692</td>
<td>0.228</td>
<td>[-1.130, -0.237]</td>
</tr>
<tr>
<td>$\lambda_{2}^{0}$</td>
<td>0.134</td>
<td>0.255</td>
<td>[-0.373, 0.630]</td>
</tr>
<tr>
<td>$\kappa_{11}^{0}$</td>
<td>-4.818</td>
<td>0.492</td>
<td>[-5.778, -3.865]</td>
</tr>
<tr>
<td>$\kappa_{12}^{0}$</td>
<td>0.109</td>
<td>0.025</td>
<td>[0.061, 0.159]</td>
</tr>
<tr>
<td>$\kappa_{21}^{0}$</td>
<td>4.052</td>
<td>0.647</td>
<td>[2.780, 5.328]</td>
</tr>
<tr>
<td>$\kappa_{22}^{0}$</td>
<td>-0.032</td>
<td>0.005</td>
<td>[-0.043, -0.022]</td>
</tr>
<tr>
<td>$\beta_{1+}^{0}$</td>
<td>0.172</td>
<td>0.024</td>
<td>[0.130, 0.227]</td>
</tr>
<tr>
<td>$\beta_{2+}^{0}$</td>
<td>0.161</td>
<td>0.024</td>
<td>[0.120, 0.213]</td>
</tr>
<tr>
<td>$\rho_{1}^{0}$</td>
<td>-0.725</td>
<td>0.010</td>
<td>[-0.742, -0.705]</td>
</tr>
<tr>
<td>$\rho_{2}^{0}$</td>
<td>-0.428</td>
<td>0.016</td>
<td>[-0.461, -0.399]</td>
</tr>
<tr>
<td>$\mu_{1}^{0}$</td>
<td>0.991</td>
<td>0.035</td>
<td>[0.023, 0.162]</td>
</tr>
<tr>
<td>$\mu_{2}^{0}$</td>
<td>0.024</td>
<td>0.016</td>
<td>[0.003, 0.056]</td>
</tr>
<tr>
<td>$s_{1M}^{2}$</td>
<td>0.142</td>
<td>0.004</td>
<td>[0.135, 0.150]</td>
</tr>
<tr>
<td>$s_{2M}^{2}$</td>
<td>0.005</td>
<td>0.000</td>
<td>[0.004, 0.005]</td>
</tr>
<tr>
<td>$s_{3M}^{2}$</td>
<td>0.045</td>
<td>0.001</td>
<td>[0.043, 0.047]</td>
</tr>
<tr>
<td>$s_{4M}^{2}$</td>
<td>0.062</td>
<td>0.002</td>
<td>[0.050, 0.065]</td>
</tr>
<tr>
<td>$s_{1Y}^{2}$</td>
<td>0.005</td>
<td>0.000</td>
<td>[0.004, 0.005]</td>
</tr>
<tr>
<td>$s_{2Y}^{2}$</td>
<td>0.053</td>
<td>0.002</td>
<td>[0.050, 0.056]</td>
</tr>
</tbody>
</table>

Table 8: Posterior Estimates of $\Theta_{P}$ and $\Theta_{E}$ for $A_{0}^{+}(2)$, $A_{1}^{+}(2)$, and $\tilde{A}_{2}(2)$

Note: This table presents the posterior estimates for the parameters that characterize the $P$-dynamics, $\Theta_{P}$, and the pricing error variances $\Theta_{E}$. We report the mean, the standard deviation, and the 95% highest posterior density intervals for the $A_{0}^{+}(2)$, $A_{1}^{+}(2)$, and $\tilde{A}_{2}(2)$. We use daily data on variance swaps from January 4, 1996 to January 11, 2013. The number of daily observations is 4,276, excluding weekends and holidays.
<table>
<thead>
<tr>
<th>Variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEF</td>
<td>0.278***</td>
<td>(0.107)</td>
<td>0.091</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TED</td>
<td></td>
<td>0.184*</td>
<td>(0.099)</td>
<td>0.078</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TERM</td>
<td></td>
<td></td>
<td>-0.003</td>
<td>(0.076)</td>
<td>-0.008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIQ</td>
<td></td>
<td></td>
<td>-0.544**</td>
<td>(0.217)</td>
<td>0.126</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POL</td>
<td></td>
<td></td>
<td></td>
<td>0.004***</td>
<td>(0.001)</td>
<td>0.002**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SKEW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ExM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>adj. $R^2$</td>
<td>0.608</td>
<td>0.615</td>
<td>0.6</td>
<td>0.611</td>
<td>0.64</td>
<td>0.601</td>
<td>0.716</td>
<td>0.601</td>
<td>0.601</td>
<td>0.73646</td>
</tr>
</tbody>
</table>

Table 9: Regression Results on Factor $X_1$

Note: In this table we report results from regression analysis to relate our volatility factors to several variables on economic fundamentals at a monthly frequency. Each column reports the results of estimating a linear regression of the posterior mean of the volatility factor $X_1$ on its lagged value (AR) and the innovation of the corresponding variable for each row. The last column corresponds to the multiple regression that includes all the explanatory variables we consider: TED spread, default spread (DEF), Chicago Fed National Activity Index (CFI), industrial production growth (IPG), term spread (TERM), monthly liquidity factor (LIQ), policy news index (POL), market skewness (SKEW), and excess market returns (ExM), see Section 4.1 for more information on their definitions. The coefficients corresponding to the constants are omitted from the regressions.
<table>
<thead>
<tr>
<th>Variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEF</td>
<td>0.818**</td>
<td>(0.336)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.527*</td>
<td>(0.277)</td>
</tr>
<tr>
<td>TED</td>
<td>0.159</td>
<td>(0.172)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.125</td>
<td>(0.167)</td>
</tr>
<tr>
<td>TERM</td>
<td>-0.200**</td>
<td>(0.099)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.194*</td>
<td>(0.099)</td>
</tr>
<tr>
<td>LIQ</td>
<td>-1.446**</td>
<td>(0.604)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1.053</td>
<td>(0.715)</td>
</tr>
<tr>
<td>POL</td>
<td>0.002</td>
<td>(0.002)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.001</td>
<td>(0.002)</td>
</tr>
<tr>
<td>SKEW</td>
<td></td>
<td></td>
<td>-0.022</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.009</td>
<td>(0.045)</td>
</tr>
<tr>
<td>ExM</td>
<td></td>
<td></td>
<td></td>
<td>-0.035***</td>
<td>(0.050)</td>
<td></td>
<td></td>
<td></td>
<td>-0.029***</td>
<td>(0.009)</td>
</tr>
<tr>
<td>IPG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.056</td>
<td>0.025</td>
<td>(0.062)</td>
</tr>
<tr>
<td>CFI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.067</td>
<td>0.097</td>
<td>(0.050)</td>
</tr>
<tr>
<td>AR</td>
<td>0.955***</td>
<td>(0.018)</td>
<td>0.954***</td>
<td>(0.018)</td>
<td>0.953***</td>
<td>(0.019)</td>
<td>0.949***</td>
<td>(0.018)</td>
<td>0.953***</td>
<td>(0.018)</td>
</tr>
<tr>
<td>adj. $R^2$</td>
<td>0.925</td>
<td>0.922</td>
<td>0.923</td>
<td>0.925</td>
<td>0.922</td>
<td>0.921</td>
<td>0.931</td>
<td>0.922</td>
<td>0.922</td>
<td>0.93757</td>
</tr>
</tbody>
</table>

Table 10: Regression Results on Factor $X_2$

Note: In this table we report results from regression analysis to relate our volatility factors to several variables on economic fundamentals at a monthly frequency. Each column reports the results of estimating a linear regression of the posterior mean of the volatility factor $X_2$ on its lagged value (AR) and the innovation of the corresponding variable for each row. The last column corresponds to the multiple regression that includes all the explanatory variables we consider: TED spread, default spread (DEF), Chicago Fed National Activity Index (CFI), industrial production growth (IPG), term spread (TERM), monthly liquidity factor (LIQ), policy news index (POL), market skewness (SKEW), and excess market returns (ExM), see Section 4.1 for more information on their definitions. The coefficients corresponding to the constants are omitted from the regressions.
<table>
<thead>
<tr>
<th>Date</th>
<th>∆Var</th>
<th>News</th>
</tr>
</thead>
<tbody>
<tr>
<td>08/18/98</td>
<td>-0.365</td>
<td>FOMC’s Decision to Leave Interest Rates Unchanged</td>
</tr>
<tr>
<td>09/01/98</td>
<td>-0.664</td>
<td>Fed Adds Money to the Banking System with Repo</td>
</tr>
<tr>
<td>09/08/98</td>
<td>-0.455</td>
<td>Fed Chairman A. Greenspan’s Statement that a Rate Cut might be Forthcoming</td>
</tr>
<tr>
<td>09/14/98</td>
<td>-0.185</td>
<td>President Clinton Advocated a Coordinated Global Policy for Economic Growth in NYC</td>
</tr>
<tr>
<td>09/23/98</td>
<td>-0.280</td>
<td>Fed Chairman A. Greenspan Testimony Before the Committee on the Budget, U.S. Senate</td>
</tr>
<tr>
<td>08/11/99</td>
<td>-0.276</td>
<td>Fed Beige Book Release Shows that US Economy Remains Strong</td>
</tr>
<tr>
<td>04/17/00</td>
<td>-0.296</td>
<td>Treasury Secretary L. Summers Statement that Fundamentals of Economy are in Place</td>
</tr>
<tr>
<td>01/03/01</td>
<td>-0.179</td>
<td>Fed’s Announcement of a Surprise, Inter-Meeting Rate Cut</td>
</tr>
<tr>
<td>05/17/05</td>
<td>-0.303</td>
<td>John Snow Call on China to Take An Intermediate Step in Revaluing its Currency</td>
</tr>
<tr>
<td>05/19/05</td>
<td>-0.297</td>
<td>Fed Chairman A. Greenspan Steps up Criticism of Fannie Mae and Freddie Mac</td>
</tr>
<tr>
<td>06/15/06</td>
<td>-0.625</td>
<td>Fed Chairman B. Bernanke’s Speech on Inflation Expectations within Historical Ranges</td>
</tr>
<tr>
<td>06/29/06</td>
<td>-0.325</td>
<td>FOMC Statement to Raise Its Target for the Federal Funds Rate by 25 Basis Points</td>
</tr>
<tr>
<td>07/19/06</td>
<td>-0.272</td>
<td>Fed Chairman B. Bernanke Warned that the Fed Must Guard Against Rising Prices Taking Hold</td>
</tr>
<tr>
<td>02/28/07</td>
<td>-0.396</td>
<td>Fed Chairman B. Bernanke Told a House Panel that Markets Seemed Working Well</td>
</tr>
<tr>
<td>03/06/07</td>
<td>-0.217</td>
<td>Henry Paulson in Tokyo Said the Global Economy was As Strong As He’s Ever Seen</td>
</tr>
<tr>
<td>03/21/07</td>
<td>-0.244</td>
<td>Fed Policy Makers Concluded their Two-Day Policy Meeting by Keeping the Fed Fund Rate</td>
</tr>
<tr>
<td>06/27/07</td>
<td>-0.271</td>
<td>FOMC Announcement Generated Market Rebound the Previous Day</td>
</tr>
<tr>
<td>08/21/07</td>
<td>-0.188</td>
<td>Senator Dodd said the Fed to Deal with the Turmoil after Meeting with Paulson and Bernanke</td>
</tr>
<tr>
<td>09/18/07</td>
<td>-0.353</td>
<td>FOMC Decided to Lower its Target for the Federal Funds Rate by 50 Basis Points</td>
</tr>
<tr>
<td>03/18/08</td>
<td>-0.216</td>
<td>Fed Cut the Fed Funds Rate by Three-Quarters of a Percentage Point</td>
</tr>
<tr>
<td>10/14/08</td>
<td>-0.304</td>
<td>President Bush and Henry Paulson Separately Announced Revisions to the TARP Program</td>
</tr>
<tr>
<td>10/20/08</td>
<td>-0.413</td>
<td>Fed Chairman B. Bernanke Testimony on the Budget, U.S. House of Representatives</td>
</tr>
<tr>
<td>10/28/08</td>
<td>-0.230</td>
<td>Fed to Cut the Rate Following the Two-Day FOMC Meeting is Expected by the Market</td>
</tr>
<tr>
<td>11/13/08</td>
<td>-0.240</td>
<td>President Bush’s Speech on Financial Crisis</td>
</tr>
<tr>
<td>12/19/08</td>
<td>-0.244</td>
<td>President Bush Declared that TARP Funds to be Spent on Programs Paulson Deemed Necessary</td>
</tr>
<tr>
<td>01/21/09</td>
<td>-0.206</td>
<td>T. Geithner Testified about Nomination as Treasury Secretary before the Senate Finance Committee</td>
</tr>
<tr>
<td>02/24/09</td>
<td>-0.261</td>
<td>President Obama’s First Speech as the President to Joint Session of U.S. Congress</td>
</tr>
<tr>
<td>05/10/10</td>
<td>-0.601</td>
<td>European Policy Makers Unveiled An Unprecedented Emergency Loan Plan</td>
</tr>
<tr>
<td>03/21/11</td>
<td>-0.277</td>
<td>Japanese Nuclear Reactors Cooled Down and Situations in Libya Tamed by Unilateral Forces</td>
</tr>
<tr>
<td>08/09/11</td>
<td>-0.370</td>
<td>FOMC Statement Explicitly Stating A Duration for An Exceptionally Low Target Rate</td>
</tr>
<tr>
<td>10/27/11</td>
<td>-0.205</td>
<td>European Union Leaders Made a Bond Deal to Fix the Greek Debt Crisis</td>
</tr>
<tr>
<td>01/02/13</td>
<td>-0.427</td>
<td>President Obama and Senator McConnell’s Encouraging Comments on the “Fiscal Cliff” Issue</td>
</tr>
</tbody>
</table>

Table 11: Policy News Potentially Associated with Estimated Volatility Jumps

Note: In this table, we report the 32 potential events in the last column that may lead to the 40 largest negative volatility jumps in sample. The first column is the date of the event, and the second column shows changes in estimated spot variance. The remaining dates on which policy related news could not be related to downward jumps are May 28, Oct 15, Oct 20, and Oct 30 of 1998, Jan 7 of 2000, June 1 of 2005, July 30 of 2007, and Nov 13 of 2007.
<table>
<thead>
<tr>
<th>Indicator</th>
<th>Category</th>
<th>Frequency</th>
<th>Release Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unemployment Rate</td>
<td>Employment</td>
<td>Monthly</td>
<td>8:30 am First Friday of each month</td>
</tr>
<tr>
<td>ADP Employment Change</td>
<td>Employment</td>
<td>Monthly</td>
<td>8:15 am - Two days before Employment situation</td>
</tr>
<tr>
<td>Initial Jobless Claims</td>
<td>Employment</td>
<td>Weekly</td>
<td>8:30 am every Thursday</td>
</tr>
<tr>
<td>Personal Income</td>
<td>Consumer Spending and Confidence</td>
<td>Monthly</td>
<td>8:30 am 4 weeks after end of reported month</td>
</tr>
<tr>
<td>Personal Spending</td>
<td>Consumer Spending and Confidence</td>
<td>Monthly</td>
<td>8:30 am 4 weeks after end of reported month</td>
</tr>
<tr>
<td>Advance Retail Sales</td>
<td>Consumer Spending and Confidence</td>
<td>Monthly</td>
<td>8:30 am 2 weeks after end of reported month</td>
</tr>
<tr>
<td>Consumer Confidence</td>
<td>Consumer Spending and Confidence</td>
<td>Monthly</td>
<td>10:00 am - Last Tuesday of month being surveyed</td>
</tr>
<tr>
<td>GDP</td>
<td>National Output and Inventories</td>
<td>Quarterly</td>
<td>8:30 am - Final week of Jan Apr Jul Oct</td>
</tr>
<tr>
<td>Durable Goods Orders</td>
<td>National Output and Inventories</td>
<td>Monthly</td>
<td>8:30 am three to four weeks after the end of reporting month</td>
</tr>
<tr>
<td>ISM Manufacturing</td>
<td>National Output and Inventories</td>
<td>Monthly</td>
<td>10:00 am First Business day after reporting month</td>
</tr>
<tr>
<td>Chicago PMI</td>
<td>National Output and Inventories</td>
<td>Monthly</td>
<td>10:00 am First Business day of month being covered</td>
</tr>
<tr>
<td>Empire State Manufacturing</td>
<td>National Output and Inventories</td>
<td>Monthly</td>
<td>8:30 am around 15th of month being reported</td>
</tr>
<tr>
<td>Business Inventories</td>
<td>National Output and Inventories</td>
<td>Monthly</td>
<td>10:00 am released six weeks after the month ends</td>
</tr>
<tr>
<td>Production and Utilization</td>
<td>National Output and Inventories</td>
<td>Monthly</td>
<td>9:15 am released the 15th of the following month</td>
</tr>
<tr>
<td>New Residential Sales</td>
<td>Housing and Construction</td>
<td>Monthly</td>
<td>8:30 am released two to three weeks following month being covered</td>
</tr>
<tr>
<td>FOMC Meeting</td>
<td>Central Bank</td>
<td>Eight Times</td>
<td>2:15 pm of day of conclusion of FOMC meeting</td>
</tr>
<tr>
<td>Federal Reserve Chairman Speech</td>
<td>Central Bank</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>ECB Governing Council Meeting</td>
<td>Central Bank</td>
<td>Monthly</td>
<td>N.A.</td>
</tr>
<tr>
<td>CPI</td>
<td>Prices, Productivity, Wages</td>
<td>Monthly</td>
<td>8:30 am second or third week following month being covered</td>
</tr>
<tr>
<td>PPI</td>
<td>Prices, Productivity, Wages</td>
<td>Monthly</td>
<td>8:30 am two or three weeks after month ends</td>
</tr>
<tr>
<td>Employment Cost Index</td>
<td>Prices, Productivity, Wages</td>
<td>Quarterly</td>
<td>8:30 am - Last Thursday of Jan Apr Jul Oct</td>
</tr>
</tbody>
</table>

**Table 12: Economic Indicators**

Note: In this table, we report the details of the 21 macroeconomic news announcements or central bank events used in Section 4.1. All times are reported in Eastern Standard Time. Source: Bloomberg and the book by Baumohl (2010) on economic indicators.
Table 13: Regression Results on Volatility Jumps

Note: In this table, we report the regressions of the magnitudes of the jumps in each volatility factor as well as jumps in total volatility, onto the magnitudes of news shocks, defined in Section 4.1. All the shocks are standardized to have variance equal to 1. The regressands are taken from the posterior median of the identified jumps in $X_1$, $X_2$, and $\sigma^2$, respectively, based on the $k_0(2)$ model.
Figure 1: Negative Jumps in the VIX

Note: In this figure, we highlight three days corresponding to the following media headlines: VIX, Vstoxx Drop by Records as Stocks Soar on Europe’s Emergency Loan Plan. - Bloomberg, Monday May 10, 2010; VIX Index Driven to Second-Biggest Percentage Drop (- 27%) on Fed’s Rate Statement. - Bloomberg, Tuesday Aug 09, 2011; The CBOE Volatility Index, or the VIX, Wall Street’s Favored Measure of Anxiety, Posted its Biggest One-Day Decline since August 2011, as Lawmakers Closed in on a Deal to Avert the “Fiscal Cliff.” - Reuters, Monday Dec 31, 2012.
Figure 2: The S&P 500 Index and Variance Swap Rates

Note: The top panel plots the time series of the S&P 500 index and its returns from January 4, 1996 to January 11, 2013. The second panel shows the variance swap rates with 6 different maturities. The maximum number of daily observations is 4,276, excluding weekends and holidays. Since we have an unbalanced panel of variance swaps, different maturities may have different number of observations, which are reported in the legend. The bottom panel plots the slopes of corresponding variance swap rates, i.e.

\[ P(t, 1/2) - P(t, 1/4) \]
\[ P(t, 1) - P(t, 1/4) \]

respectively, where \( P(t, \tau) \) denotes the variance swap rate at time \( t \) of a contract with time to maturity \( \tau \). Positive values reflect an upward sloping term structure while the opposite slope is implied by negative values.
Figure 3: Volatility Factors

Note: This figure reports the posterior means (in blue) of the two latent factor estimates for $A_0(2)$, $A_1(2)$, $A_2(2)$, $A_0^+(2)$, and $A_1^+(2)$, as well as the logarithm of the factors for the $A_2(2)$ model. The red areas around the blue curves mark the 95%-credible sets. We use daily data on variance swaps from January 4, 1996 to January 11, 2013. The number of daily observations is 4,276, excluding weekends and holidays.
Figure 4: Out of Sample Performance

Note: This figure compares the estimated 1-month variance swap rates with the VIX across models over the entire sample period. The red solid line denotes the VIX from the CBOE, whereas the blue dash-dotted line is calculated based on the $Q$-parameters estimated from the variance swap rates with time-to-maturity of at least 2 months.
Figure 5: Decomposition of Spot Variances

Note: In this figure we decompose the spot variance $\sigma_t^2$ into its continuous and jump components for all six models, respectively. The left panel plots the identified jump components from the percent changes of the estimated spot variances, and the remaining Brownian shocks are plotted on the right panel. All these components are extracted based on the corresponding parametric models. The red circles correspond to the three aforementioned events in Figure 1.
Figure 6: Term Structure of Variance Risk Premia

Note: In this figure we plot the term structure of variance risk premia for $A_0(2)$ and $A_{0+}(2)$, respectively. The green solid lines plot the risk premia for the 2-month contracts, the blue dash-dotted line for the 6-month contracts, and the red dashed line for the 1-year contracts. The shaded areas around the lines plot the 95% confidence intervals.
Figure 7: Robustness Check

Note: On the upper panel, we compare the realized volatility estimated from the $\mathcal{A}_0(2)$ model (blue solid) with the realized volatility estimated from 5-minute returns of the SPY (red dotted). On the lower panel, we plot the monthly realized profit and loss of holding a hypothetical 1-month variance swap contract each month.
Appendix

A Variance Swap Pricing

Proof of Proposition 1. Recall that since $X$ is affine, the generalized conditional characteristic function (GCCF) of $X_s$ is defined below, for any $s \geq t$ with $t$ fixed:

$$\Psi(s, t, u, X_t) = \mathbb{E}^Q_t [e^{u^\top X_s}],$$

where $u \in \mathbb{C}^N$. There exists a closed-form formula for the GCCF function given by Duffie et al. (2000):

$$\log \left( \Psi(s, t, u, X_t) \right) = A(s - t, u) + B(s - t, u)^\top X_t,$$

where $A$ and $B$ satisfy the following ordinary differential equations (ODEs):

$$\dot{B} = (K^Q)^\top B + \frac{1}{2} \sum_{i=1}^m \left[ \Sigma^\top B \right]_i^2 \beta_i + l_1 \phi(B),$$

$$\dot{A} = (\Lambda^Q)^\top B + \frac{1}{2} \sum_{i=1}^m \left[ \Sigma^\top B \right]_i^2 \alpha_i + l_0 \phi(B),$$

where $B(t) = u$, $A(t) = 0$, and for any $h \in \mathbb{C}^N$,

$$\phi(h) = \int_{\mathbb{R}^N} (e^{h^\top z} - 1 - h^\top z) \nu^Q(dz).$$

Under our risk neutral specification, we have

$$\mathbb{E}^Q_t \left\{ \int_t^{t+\tau} \sigma_s^2 ds + \int_t^{t+\tau} \int_{\mathbb{R}^N} j^2 \nu^Q_s(dj) \right\} = \mathbb{E}^Q_t \left\{ \int_t^{t+\tau} f^Q(X_s)ds \right\},$$

where $f^Q(X_s) = \Pi_0^Q + \Pi_1^Q X + X^\top \Pi_2 X + \exp \{ \Pi_3 + \Pi_4^\top X \}$, $\Pi_0^Q = \Pi_0 + l_0 \int_{\mathbb{R}} j^2 \nu^Q(dj)$, $\Pi_1^Q = \Pi_1 + l_1 \int_{\mathbb{R}} j^2 \nu^Q(dj)$, and $\nu^Q(dj)$ is the marginal distribution of jumps in $Y$.

Denote the transition density of the process $X$ as $p(X_s|s - t, X_t)$, and let $u = -iv$ in $\Psi$ with $v \in \mathbb{R}^N$, we have

$$\mathbb{E}^Q_t \left( f^Q(X_s) \bigg| X_t = x \right) = \int_{\mathbb{R}^N} f^Q(x')p(x'|s - t, x)dx'.$$

We then utilize Fourier Transform of the tempered distributions to simplify the integral with respect to $x'$. Consider the quadratic part first. Note that

$$\int_{\mathbb{R}^N} (\Pi_0^Q + (\Pi_1^Q)^\top x' + x'^\top (\Pi_2) x') e^{iv^\top x'} dx' = (2\pi)^N \left( \Pi_0^Q - i(\Pi_1^Q)^\top \nabla_v - \nabla_v(\Pi_2) \nabla_v \right) \delta(v),$$

52
where \( \delta(\cdot) \) is a Dirac delta that satisfies \( \int_{\mathbb{R}^N} \delta(v)dv = 1 \), and \( \int_{\mathbb{R}^N} \delta(v)g(v)dv = g(0) \) for any test function \( g \). Therefore, by direct calculations we obtain

\[
\mathbb{E}^Q \left( \Pi_0^Q + (\Pi_1^Q)^\top X_s + X_s^\top (\Pi_2) X_s \mid X_t = x \right) = \int_{\mathbb{R}^N} \left( \Pi_0^Q - i(\Pi_1^Q)^\top \nabla_v - \nabla_v \Pi_2 \nabla_v \right) \delta(v) \Psi(s, t, -iv, x)dv
\]

\[
= \Pi_0^Q + (\Pi_1^Q)^\top \nabla_v \Psi(s, t, u, x) \bigg|_{u=0} + \nabla_v \Pi_2 \nabla_v \Psi(s, t, u, x) \bigg|_{u=0}.
\]

For the exponential part, similarly we have

\[
\int_{\mathbb{R}^N} e^{\Pi_3 + (\Pi_4)^\top x'} e^{iv^\top x'} dx' = (2\pi)^{N} e^{\Pi_3} \delta(v - i\Pi_4),
\]

so that we can derive

\[
\mathbb{E}^Q \left( e^{\Pi_3 + (\Pi_4)^\top X_s} \mid X_t = x \right) = \int_{\mathbb{R}^N} e^{\Pi_3 \delta(v - i\Pi_4)} \Psi(s, t, -iv, x)dv = e^{\Pi_3} \Psi(s, t, \Pi_4, x).
\]

The pricing formula for variance swaps follows immediately. Note that we have applied properties of tempered distributions to simplify the calculations, all of which can be found in Kanwal (2004).

\[ \blacksquare \]

### B Invariant Transformations and Extended Canonical Forms

**Proof of Proposition 2.** To prove the existence, we extend Dai and Singleton (2000) and Ahn et al. (2002) to provide invariant transformations of the general model. These transformations lead to alternative specifications without altering the price of variance swaps (or more generally, the likelihood of the observables). We summarize the state factors, Brownian motions, jumps, and parameter vectors in \( \theta \):

\[
\theta = \left( X_t, W_t^Q, Z_t^Q, \Lambda^Q, K^Q, \Sigma, \{\alpha_i, \beta_i\}_{1 \leq i \leq m}, \bar{\nu}^Q(\cdot, dz), \Pi_0, \Pi_1, \Pi_2, \Pi_3, \Pi_4 \right).
\]

There are 4 classes of admissible transformations that ensure the transformed process to follow (1), (2), (3), and (4) (with different set of parameters), while maintaining the same observable implications:

**An Affine Transformation** \( \mathcal{T}_A \) refers to \( \mathcal{T}_A X_t = \mathcal{V} + \mathcal{L} X_t \), where \( \mathcal{V} \) is an \( N \times 1 \) vector and \( \mathcal{L} \) is an \( N \times N \) nonsingular matrix. As a result, \( \mathcal{T}_A \theta \) is defined below.

\[
\mathcal{T}_A \theta = \left( \begin{array}{c}
\mathcal{V} + \mathcal{L} X_t, W_t^Q, \mathcal{L} Z_t^Q, \mathcal{L} \Lambda - \mathcal{L} K \mathcal{L}^{-1} \mathcal{V}, \mathcal{L} K \mathcal{L}^{-1}, \mathcal{L} \Sigma,
\{\alpha_i - \beta_i^\top \mathcal{L}^{-1} \mathcal{V}, \mathcal{L} \Sigma^{-1} \beta_i\}_{1 \leq i \leq m}, \bar{\nu}^Q(\mathcal{L}^{-1}(\cdot + \mathcal{V}), \mathcal{L} dz),
\Pi_0 - (\Pi_1)^\top \mathcal{L}^{-1} \mathcal{V} + \mathcal{V}^\top \mathcal{L}^{-1}(\Pi_2) \mathcal{L}^{-1} \mathcal{V}, \mathcal{L}^{-1} \Pi_1 - 2 \mathcal{L}^{-1} \Pi_2 \mathcal{L}^{-1} \mathcal{V}, \mathcal{L}^{-1} \Pi_2 \mathcal{L}^{-1},
\Pi_3 - (\Pi_4)^\top \mathcal{L}^{-1} \mathcal{V}, \mathcal{L}^{-1} \Pi_4
\end{array} \right).
\]

**An Orthonormal Rotation** \( \mathcal{T}_O \) refers to an affine transformation on the Brownian factor \( W_t^Q \) such that \( \mathcal{T}_O W_t^Q = OW_t^Q \), where \( O \) is an orthonormal matrix satisfying \( O^\top O = OO^\top = I_{N \times N} \).

\[
\mathcal{T}_O \theta = \left( X_t, OW_t^Q, Z_t^Q, \Lambda^Q, K^Q, \Sigma O^\top, \{\alpha_i, \beta_i\}_{1 \leq i \leq m}, \bar{\nu}^Q(\cdot, dz), \Pi_0, \Pi_1, \Pi_2, \Pi_3, \Pi_4 \right).
\]
A Diffusion Rescaling $\mathcal{T}_D$ rescales the diagonal elements of $S_t$ by a nonsingular diagonal matrix $D$ in $\mathbb{R}^{N \times N}$. That is,

$$\mathcal{T}_D \theta = \left( X_t, W_t^Q, Z_t^Q, \Lambda^Q, K^Q, \Sigma D^{-1}, \{D_{ii}^Q \alpha_i, D_{ii}^Q \beta_i\}_{1 \leq i \leq m}, \nu^Q(\cdot, dz), \Pi_0, \Pi_1, \Pi_2, \Pi_3, \Pi_4 \right).$$

**A Permutation** $\mathcal{T}_P$ alters the order of state variables, which has no observable effect.

Using these transformations, we can impose normalizations on the process to achieve its canonical representation. Following exactly the same procedure as in Appendix C of Dai and Singleton (2000), we normalize the parameters in the dynamics of (2), while leaving the parameters in (3) unrestricted (barring from their positivity constraints). Once we have transformed any model of (2) into its canonical form, no restrictions can be imposed on the parameters in (3) without affecting this canonical form, so that the procedure achieves the maximal model.

To show the uniqueness, by the existence result, it is equivalent to prove that canonical forms of different types are not observationally equivalent under these invariant transformations. This is obvious from Dai and Singleton (2000), because the number of positive factors remains unchanged under admissible transformations.

## C Summary of Two Factor Volatility Models

### C.1 $A_0(2)$ Model

The $A_0(2)$ model specifies the dynamics of $X$ as:

$$\begin{bmatrix}
  dX_{1t} \\
  dX_{2t}
\end{bmatrix} = \left( \begin{bmatrix}
  \kappa_{11}^Q & 0 \\
  \kappa_{21}^Q & \kappa_{22}^Q
\end{bmatrix} \begin{bmatrix}
  X_{1t} \\
  X_{2t}
\end{bmatrix} \right) dt + \begin{bmatrix}
  dW_{1t}^Q \\
  dW_{2t}^Q
\end{bmatrix} + \begin{bmatrix}
  dZ_{1t}^Q \\
  dZ_{2t}^Q
\end{bmatrix}.$$ 

Jumps follow compound Poisson processes with independent jump sizes following double exponential distributions:

- size of $Z_{1t}^Q \sim \begin{cases}
  \exp(\beta_{1+}^Q), & q_1 \\
  -\exp(\beta_{1-}^Q), & 1 - q_1
\end{cases}$, and

- size of $Z_{2t}^Q \sim \begin{cases}
  \exp(\beta_{2+}^Q), & q_2 \\
  -\exp(\beta_{2-}^Q), & 1 - q_2
\end{cases}$.

Their intensity is specified as $l_0$.

For this model, we specify the dynamics under $\mathbb{P}$ as

$$\begin{bmatrix}
  dX_{1t} \\
  dX_{2t}
\end{bmatrix} = \left( \begin{bmatrix}
  \lambda_{11}^P & 0 \\
  \lambda_{21}^P & \lambda_{22}^P
\end{bmatrix} \begin{bmatrix}
  X_{1t} \\
  X_{2t}
\end{bmatrix} \right) dt + \begin{bmatrix}
  dW_{1t}^P \\
  dW_{2t}^P
\end{bmatrix} + \begin{bmatrix}
  dZ_{1t}^P \\
  dZ_{2t}^P
\end{bmatrix},$$

where jumps in $Z_{1t}^P$ and $Z_{2t}^P$ are specified with the same mixture probabilities but in different sizes $\beta_{1,+/-}^P$ and $\beta_{2,+/-}^P$.
The parameter constraints in this model are given by:

\[ \kappa_{11}^Q < 0, \quad \kappa_{22}^Q < 0, \quad l_0 \geq 0, \quad \kappa_{11}^p < 0, \quad \kappa_{22}^p < 0. \]

C.2 \( \mathbb{A}_1(2) \) Model

Another model that incorporates negative jumps can be specified as

\[
\begin{bmatrix}
\frac{dX_{1t}}{dX_{2t}}
\end{bmatrix} = \left( \begin{bmatrix}
\lambda_1^Q \\
0
\end{bmatrix} + \begin{bmatrix}
\kappa_{11}^Q & 0 \\
\kappa_{21}^Q & \kappa_{22}^Q
\end{bmatrix} \begin{bmatrix}
X_{1t} \\
X_{2t}
\end{bmatrix} \right) dt + \begin{bmatrix}
\sqrt{X_{1t}} \\
0
\end{bmatrix} dt + \begin{bmatrix}
0 \\
\sqrt{1 + \beta_{21} X_{1t}}
\end{bmatrix} \begin{bmatrix}
dW_{11t}^Q \\
dW_{22t}^Q
\end{bmatrix} + \begin{bmatrix}
dZ_{11t}^Q \\
dZ_{22t}^Q
\end{bmatrix},
\]

where \( X_1 \) is a square-root factor, and \( X_2 \) is an Ornstein-Uhlenbeck factor. Jumps of \( X_1 \) and \( X_2 \) follow compound Poisson processes with independent jump sizes satisfying the exponential or double exponential distributions:

\[
\text{size of } Z_{1t}^Q \sim \exp(\beta_{1+}^Q), \quad \text{and} \quad \text{size of } Z_{2t}^Q \sim \begin{cases} 
\exp(\beta_{2+}^Q) & \text{with probability } q_2 \\
-\exp(\beta_{2-}^Q) & \text{with probability } 1 - q_2.
\end{cases}
\]

Their intensity is specified as \( l_0 + l_{11} X_{1t} \).

For this model, we specify the dynamics under the objective measure \( \mathbb{P} \) as

\[
\begin{bmatrix}
\frac{dX_{1t}}{dX_{2t}}
\end{bmatrix} = \left( \begin{bmatrix}
\lambda_1^p \\
\lambda_2^p
\end{bmatrix} + \begin{bmatrix}
\kappa_{11}^p & 0 \\
\kappa_{21}^p & \kappa_{22}^p
\end{bmatrix} \begin{bmatrix}
X_{1t} \\
X_{2t}
\end{bmatrix} \right) dt + \begin{bmatrix}
\sqrt{X_{1t}} \\
0
\end{bmatrix} dt + \begin{bmatrix}
0 \\
\sqrt{1 + \beta_{21} X_{1t}}
\end{bmatrix} \begin{bmatrix}
dW_{11t}^p \\
dW_{22t}^p
\end{bmatrix} + \begin{bmatrix}
dZ_{11t}^p \\
dZ_{22t}^p
\end{bmatrix}.
\]

Jumps are of the same type with the same intensity and mixture probability but different sizes \( \beta_{1+}^p, \beta_{2+}^p, \) and \( \beta_{2-}^p \).

The parameter constraints in this model are given by:

\[ \kappa_{11}^Q < l_{11} \beta_{1+}^Q, \quad \kappa_{22}^Q < 0, \quad \lambda_1^Q - l_0 \beta_{1+}^Q \geq \frac{1}{2}, \quad \beta_{21} \geq 0, \quad l_0 \geq 0, \quad l_{11} \geq 0, \]

\[ \kappa_{11}^p < l_{11} \beta_{1+}^p, \quad \kappa_{22}^p < 0, \quad \lambda_1^p - l_0 \beta_{1+}^p \geq \frac{1}{2}. \]

C.3 \( \mathbb{A}_2(2) \) Model

The dynamics of the state variables in the \( \mathbb{A}_2(2) \) model is specified as

\[
\begin{bmatrix}
\frac{dX_{1t}}{dX_{2t}}
\end{bmatrix} = \left( \begin{bmatrix}
\lambda_1^Q \\
0
\end{bmatrix} + \begin{bmatrix}
\kappa_{11}^Q & 0 \\
\kappa_{21}^Q & \kappa_{22}^Q
\end{bmatrix} \begin{bmatrix}
X_{1t} \\
X_{2t}
\end{bmatrix} \right) dt + \begin{bmatrix}
\sqrt{X_{1t}} \\
0
\end{bmatrix} dt + \begin{bmatrix}
0 \\
\sqrt{X_{1t}}
\end{bmatrix} \begin{bmatrix}
dW_{11t}^Q \\
dW_{22t}^Q
\end{bmatrix} + \begin{bmatrix}
dZ_{11t}^Q \\
dZ_{22t}^Q
\end{bmatrix},
\]

where jumps in \( Z_{1t} \) and \( Z_{2t} \) cannot be negative. The intensity of jumps is \( l_0 + l_{11} X_{1t} + l_{12} X_{2t} \).

The corresponding \( \mathbb{P} \) measure dynamics is specified as:

\[
\begin{bmatrix}
\frac{dX_{1t}}{dX_{2t}}
\end{bmatrix} = \left( \begin{bmatrix}
\lambda_1^p \\
0
\end{bmatrix} + \begin{bmatrix}
\kappa_{11}^p & 0 \\
\kappa_{21}^p & \kappa_{22}^p
\end{bmatrix} \begin{bmatrix}
X_{1t} \\
X_{2t}
\end{bmatrix} \right) dt + \begin{bmatrix}
\sqrt{X_{1t}} \\
0
\end{bmatrix} dt + \begin{bmatrix}
0 \\
\sqrt{X_{1t}}
\end{bmatrix} \begin{bmatrix}
dW_{11t}^p \\
dW_{22t}^p
\end{bmatrix} + \begin{bmatrix}
dZ_{11t}^p \\
dZ_{22t}^p
\end{bmatrix},
\]
with exponentially distributed jumps and different mean parameters.

The parameter constraints in this model are given by:

\[
\begin{align*}
\text{Re} \left( \text{Eigen} \left( \begin{bmatrix}
\kappa_{11}^\Omega - l_{11}^\beta 1_+ \\
\kappa_{21}^\Omega \\
\kappa_{21}^\beta - l_{22}^\beta 2_+
\end{bmatrix} \right) \right) &< 0, \quad \lambda_1^\Omega - l_0 \beta_1^+ \geq \frac{1}{2}, \quad \lambda_2^\Omega - l_0 \beta_2^+ \geq \frac{1}{2}, \\
\text{Re} \left( \text{Eigen} \left( \begin{bmatrix}
\kappa_{11}^\nu - l_{11}^\beta 1_+ \\
\kappa_{21}^\nu \\
\kappa_{21}^\beta - l_{22}^\beta 2_+
\end{bmatrix} \right) \right) &< 0, \quad \lambda_1^\nu - l_0 \beta_1^+ \geq \frac{1}{2}, \quad \lambda_2^\nu - l_0 \beta_2^+ \geq \frac{1}{2}, \\
\kappa_{12}^\Omega &\geq 0, \quad \kappa_{21}^\Omega \geq 0, \quad \kappa_{12}^\nu \geq 0, \quad \kappa_{21}^\nu \geq 0, \quad l_0 \geq 0, \quad l_{11} \geq 0, \quad l_{12} \geq 0.
\end{align*}
\]

D Likelihood Inference in Detail

Below we give a more detailed description of the Gibbs blocks used in the posterior simulator. For the purpose of concreteness in this section we focus on the \( A_1(2) \) model, which contains one Ornstein-Uhlenbeck factor and one square-root factor.

D.1 Time Discretization and Joint Likelihood

A time discretization of the model with time interval \( \Delta \) yields

\[
\begin{align*}
y_i := Y_i - Y_{(i-1)} &= \mu \Delta + \sigma_{i-1} \sqrt{\Delta} \left[ \sqrt{1 - \rho_1^2 - \rho_2^2} \epsilon_{0i} + \rho_1 \epsilon_{1i} + \rho_2 \epsilon_{2i} \right] + j_i n_i, \\
X_{1i} - X_{1(i-1)} &= \left[ \lambda_1 + \kappa_{11} X_{1(i-1)} \right] + \sqrt{X_{1(i-1)} \Delta} \epsilon_{1i} + z_{1,i} n_i, \\
X_{2i} - X_{2(i-1)} &= \left[ \lambda_2 + \kappa_{21} X_{1(i-1)} + \kappa_{22} X_{2(i-1)} \right] \Delta + \sqrt{X_{2(i-1)} \Delta} \epsilon_{2i} + z_{2,i} n_i,
\end{align*}
\]

where \( n_i \) denotes the jump time indicator that takes the value one if there is a jump on that day, and \( \epsilon_{0i}, \epsilon_{1i}, \) and \( \epsilon_{2i} \) are standard normal variates with zero correlations, \( j_i, z_{1,i} \) and \( z_{2,i} \) are Gaussian, Gamma, and mixture of Gammas, respectively. Note that \( \mu = \mu^\nu - l_0 \mu^\beta, \quad \lambda_1 = \lambda_1^\nu - l_0 \beta_1^+, \quad \lambda_2 = \lambda_2^\nu - l_0 (q_2 \beta_2^+ - (1 - q_2) \beta_2^-), \quad \kappa_{11} = \kappa_{11}^\nu - l_{11} \beta_1^+, \quad \kappa_{21} = \kappa_{21}^\nu - l_{11} (q_2 \beta_2^+ - (1 - q_2) \beta_2^-), \) and \( \kappa_{22} = \kappa_{22}^\nu - l_{12} (q_2 \beta_2^+ - (1 - q_2) \beta_2^-). \)

The joint likelihood of the observables is then given by:

\[
\mathcal{L}(Y, P | V, \Theta, A_1(1)) = \prod_{i=1}^{T} p(y_i | X_i | X_{i-1}, j_i, z_i, n_i) \times p(j_i, z_i, n_i | X_{i-1}) \times p(P_i | X_i, \Theta),
\]

which includes both the likelihood from the Euler discretization of the process and the likelihood of the variance swap rates. Eraker et al. (2003) show that discretization performs well with daily data. Alternatively, one could introduce a set of auxiliary data points in between of each pair of sampled latent variables and integrate them out of the likelihood function by MCMC.
D.2 Jump Times and Sizes

In our application the jump indicator $n_i$ is a binary random variable (taking on 0 or 1). To compute the Bernoulli probability, we use the conditional density of increments to volatility and returns to get that $\Pr(n_i = 1|V, \Theta, Y, P)$, which is equal to

$$
p(y_i, X_i|X_{(i-1)}, j_i, z_1,i, z_2,i, n_i = 1, \Theta) \times \Pr(n_i = 1|X_{1(i-1)})
\sum_{s=0,1} p(y_i, X_i|X_{(i-1)}, j_i, z_1,i, z_2,i, n_i = s, \Theta) \times \Pr(n_i = s|X_{1(i-1)})^{-1},$$

where $p(y_i, X_i|X_{(i-1)}, j_i, z_1,i, z_2,i, n_i = s, \Theta)$ is trivariate normal with mean and covariance matrix that can be easily obtained from (D.1) and $\Pr(n_i = 1|X_{1(i-1)}) = (l_0 + l_1X_{1(i-1)})\Delta$. Not surprisingly, the conditional posterior of jump times does not depend on the option prices directly since option prices do themselves not depend on the jump indicator.

To sample $j_i$, we note from (9) that $p(j_i|y_i, X_i, X_{(i-1)}, n_i = 1)$ is proportional to

$$p(y_i|X_i, X_{(i-1)}, j_i, n_i = 1) \times p(j_i|n_i = 1).$$

Completing the square in the previous expression we can easily obtain the mean and variance for the conditional posterior of $j_i$ which is normal.

Analogous computations allow us to sample $z_{1,i}$ and $z_{2,i}$, which have a discrete scale mixture of truncated normals (TN) with a mixing variate that takes a positive (negative) value with mean $\mu_{k,+i}$ ($\mu_{k,-i}$) that can be easily obtained for $k = 1, 2$ by completing the squares. That is, if $s_{k,i} \in \{0, 1\}$, with $\Pr(s_{k,i} = 1|y_i, X_i, \Theta) = q_k$, then

$$z_{k,i} = s_{k,i} \cdot \text{TN}(\mu_{k,+i}^*, \sigma_{k,i}^2; z_{k,i} > 0) + (1 - s_{k,i}) \cdot \text{TN}(\mu_{k,-i}^*, \sigma_{k,i}^2; z_{k,i} < 0),$$

where $\sigma_{k,i}^2$ denotes the corresponding conditional posterior variance of the jump size in $z_{k,i}$.

Finally, when $n_i = 0$, the conditional posteriors of $j_i$, $z_{1,i}$ and $z_{2,i}$ are the priors implied by the model assumptions, as the data provide no information about them.

D.3 Latent Factors

The conditional posterior for latent factors is not known in closed form. To sample from it, we collect terms in (9) where $X_i$ is included, which is proportional to

$$p(X_{i+1}|X_i, z_{1,i+1}, z_{2,i+1}, n_{i+1}) \times p(X_i|X_{i-1}, z_1,i, z_2,i, n_i)$$
$$\times p(y_{i+1}|X_i, X_{i+1}, j_{i+1}, n_{i+1}) \times p(y_i|X_{i-1}, X_i, j_i, n_i) \times p(P_i|X_i, \Theta) \times p(n_{i+1}|X_i),$$

where the first five densities are Gaussian and the last term is binomial. At the $g$-th iteration of the sampler, we then draw from its conditional posterior using a random-walk Metropolis algorithm with the Gaussian proposal density with mean and variance computed as in Proposition 2 of Eraker (2001) but taking into account the presence of jumps. The acceptance rate of this step is in the 20-30% range for all models.
D.4 \(\Theta_M\) and \(\Theta_I\)

Conditional on jump sizes, jump times, spot variance, short-term variance level, and remaining parameter vectors, the posterior of \(\Theta_M\) is proportional to (9). Since this conditional distribution is nonstandard, it is sampled using a Metropolis step with a normal source density centered at the current draw and covariance matrix proportional to the Hessian of \(L(Y, P|V, \Theta, M)\cdot H(V|\Theta, M)\) at the peak of \(\Theta_M\). The Hessian was computed by concentrating the latent variables and remaining parameters on their posterior means from a preliminary run of the algorithm. An analogous but simpler procedure, since \(H(V|\Theta, M)\) does not appear in the conditional posterior, allows us to draw \(\Theta_I\). The acceptance rate of this step is around 20% for all three models. The priors are relatively uninformative but still impose the relevant constraints.

D.5 \(\Theta_P\) and \(\Theta_E\)

A similar procedure to the one mentioned above can be used to sample \(\Theta_P\). In practice, however, since those parameters do not depend on variance swap rates once we condition on \(V\), it is often the case that the conditional posterior distribution is available and therefore one can sample from it directly. The same comment applies to the variances of pricing errors as long as one chooses appropriately both the pricing error distributions and priors.

As for \(\beta_{1+}^P\), recall \(z_{1,i} \sim \text{Exponential}(\beta_{1+}^P)\), so that conditional on \(z_{1,i}\), and setting a conjugate prior for \(\beta_{1+}^P\), say \(\pi_{\beta_{1+}^P(\beta_{1+}^P)} \sim \text{InvGam}(\delta_{\beta_{1+}^P1}, \delta_{\beta_{1+}^P2})\), we have that \(\beta_{1+}^P|... \sim \text{invGam}(\delta_{\beta_{1+}^P1}, \delta_{\beta_{1+}^P2})\) with \(\delta_{\beta_{1+}^P1} = N_J + \delta_{\beta_{1+}^P1}\) and \(\delta_{\beta_{1+}^P2} = \delta_{\beta_{1+}^P2} + \sum_{i=1}^{N_J} z_{1,i}\) and where \(N_J = \sum_{i=1}^{T} n_i\). Similarly, we proceed with \(\beta_{2+}^P\) and \(\beta_{2-}^P\), but using the appropriate sample sizes \(N_{J+}^+ = \sum_{i=1}^{T} n_i1\{z_{2,i}>0\}\) and \(N_{J-}^+ = N_J - N_{J+}^+\) and with \(\pi_{\beta_{2+}(\beta_{2+})} \sim \text{invGam}(\delta_{\beta_{2+}1}, \delta_{\beta_{2+}2})\) and \(\pi_{\beta_{2-}(\beta_{2-})} \sim \text{invGam}(\delta_{\beta_{2-}1}, \delta_{\beta_{2-}2})\) being the corresponding priors.

Conditional on \(X\), \(\beta_{1+}^P\), \(\beta_{2+}^P\), \(\beta_{2-}^P\), \(q_2\), \(\Theta_M\) and jump times and sizes, using the jump adjusted processes \(\tilde{Y}_i = Y_i - j_in_i\), \(\tilde{X}_{1i} = X_{1i} - z_{1,in_i}\) and \(\tilde{X}_{2i} = X_{2i} - z_{2,n_i}\), we can sample \(\mu^P\), \(\kappa_{11}^P\), \(\kappa_{22}^P\), and \(\kappa_{21}^P\) using the standard normal multivariate regression model with known variance. In this context, prior information for those parameters can be easily introduced through Gaussian conjugate priors. Finally, we use a Metropolis step to compute the conditional posterior of \(\rho_1\) and \(\rho_2\), which is proportional to \(p(y_i, X_i|X_{i-1}, j_i, z_i, n_i)\).