Market Liquidity, Variance Dynamics, and Option Prices*

Yoontae Jeon †

Rotman School of Management, University of Toronto

January 9, 2016

Abstract

Market liquidity has been known to have a significant relationship to the return variance dynamics. But, little is known about the role of liquidity in determining option prices. Motivated by this, I build a model in which observed market liquidity affects the return variance dynamics and I study its implications on option prices. The option pricing literature has shown that two-factor latent stochastic volatility models are well-suited for explaining option smirks and variance term structure slopes. My model generalizes the latent two-factor SV model by identifying liquidity as a factor. Hence, it provides a plausible economic explanation for one of the underlying factors in multi-factor models. I also contribute to the recent volatility modelling literature that attempts to identify observable financial variables that drive volatility dynamics. When fitted to S&P500 returns, realized index variance, equity market liquidity, and SPX index options, my model outperforms a standard one-factor model and equity market liquidity is shown to play a crucial role in explaining index option prices.

* The author thanks Peter Christoffersen, Bing Han, Thomas McCurdy, Chayawat Ornthanalai, and Jason Wei for helpful comments.

† Joseph L. Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, Ontario, Canada, M5S 3E6. Email: Yoontae.Jeon12@Rotman.Utoronto.ca
1 Introduction

The relationship between returns volatility and liquidity has been studied extensively in the literature. In particular, motivated by the mixture of distribution hypothesis (MDH) which assumes simultaneous dependence of volatility and volume on a latent information process, much research effort has been devoted studying existing commonality between stock return volatility and volume (Clark (1973), Epps and Epps (1976), Tauchen and Pitts (1983)). The empirical results are mixed and understanding the relationship between an information flow and trading activity is still an active research area.\(^1\) On the other hand, Glosten and Milgrom (1985)'s model implies the positive relationship between bid-ask spreads and volatility. In their model, the friction due to the existence of informed traders creates the positive relationship in the equilibrium. More recently, it has been shown that volume alone is not an adequate measure of liquidity and many other measures that directly measures the trading costs have been proposed.\(^2\) Especially with the recent availability of the intra-day trade quote data, one can obtain more precise measure of liquidity than volume by looking at the individual trade-level prices and quotes (Ait-Sahalia and Yu (2009), Goyenko, Holden, and Trzcinka (2009), Christoffersen et al. (2014)). Many of these papers find non-trivial relationship between the volume and trading cost based liquidity measures. Putting these evidences together, it is not clear whether the volume is the best measure that co-moves with the volatility.

In the standard option valuation literature, an extensive amount of research is devoted to studying a volatility dynamics that can resolve the shortcomings of the option pricing model of Black and Scholes (1973). Particular attention is placed on the models that can explain the phenomenon that out-of-the-money index put options tend to be more expensive than the Black-Scholes implied price. This is coined by the "option smirk", which refers to the implied volatility shape of observed market prices. The most popular approach resolving this issue is the use of stochastic volatility models (Hull and White (1987), Heston (1993)). However, the standard single-factor stochastic volatility model suffers from the fact that the volatility level and volatility term structure slopes are not allowed to move independently of one another. This major drawback becomes problematic when one tries to fit large sample of cross-section of option contracts simultaneously.

Consequently, two-factor stochastic volatility models have been suggested to resolve this issue (Bates (2000), Christoffersen, Heston, and Jacobs (2009), Anderson, Fusari, and

---

\(^1\) For theoretical modifications on MDH, see, for example, He and Velu (2014). Works on information flow and trading activity, see Chae (2005), Tookes (2008), Hendershott, Jones, and Menkveld (2011)

Todorov (2013, 2014)). Including a second factor helps explain time-variation in option smirks as the added factor effectively becomes a slope factor of the implied volatility smirk as well as the slope of the future variance term structure. However, as all these models are written in a latent factor form, they do not provide an explanation for any of the underlying factors.

Motivated by the above two facts, I develop a model in which the second stochastic volatility factor is not latent any more, but instead given by the observable liquidity process measured by effective spreads using intra-day trade level data. In addition, I assume that the return variance is well-measured by the realized variance measure also constructed from high-frequency intra-day data. Using these observables, I can decompose the variance into a component coming from liquidity and a component that is independent from liquidity. Therefore, I provide an economical interpretation of the two-factor model by claiming that one of the factors is market liquidity.

Using stochastic co-variates in derivative pricing has been popular in credit derivatives literature. However, there has been almost no research that uses stochastic co-variates to help option valuation. In this paper, I suggest liquidity as one of the possible candidates for stochastic co-variates of the volatility, thus providing an interpretation to the multi-factor models of stochastic volatility. Specifically, I estimate the proposed model on observables excluding the options data first. Then, I estimate the risk premium parameters only using the options data sequentially. In this way, I can compare each model’s ability to explain observed dynamics of market variables and option prices simultaneously.

The impact of liquidity on option prices is still a growing area of research and more empirical work needs to be done. Cetin, Jarrow, and Protter (2004) and Cetin et al. (2006) model liquidity using a stochastic supply curve and obtains a formula for pricing European options. Cetin et al. (2006) finds empirical evidence that spot liquidity is an important component in determining the option prices. Chou et al. (2011) empirically shows that a more illiquid underlying asset is related to higher implied volatility, hence more expensive option prices. Since the no-arbitrage pricing of options crucially depends on the replication of options with the underlying and the risk-free asset, frictions in the trading of the underlying asset should play an important role in determining option prices. In this sense, my measure of liquidity based on effective spreads is arguably better than the trading volume as I directly capture the magnitude of hedging costs.

provides evidence that market liquidity is a factor significantly related to volatility and its contribution to volatility dynamics helps explain option prices.

In this paper, I use the observed market returns, realized index variance and the average effective spread of S&P500 stocks from 2004 to 2012 to study the impact of liquidity on index volatility dynamics. A significant portion of index volatility is explained by liquidity when fitted using the Kalman filter. I subsequently estimate the risk-neutral dynamics by fitting the model to the European index option data from 2004 to 2012. In both versions of the models I develop, I find that having observed liquidity as a second factor significantly improves the model’s ability to explain option prices. I also demonstrate the ability of a liquidity factor to capture the slopes of the option smirk.

The remainder of the paper proceeds as follows. Section 2 describes the data I use and show statistical evidences that relates liquidity to return volatility. Section 3 introduces the two models I develop and derives some model properties. Section 4 is devoted to the discussion of the estimation procedure using the observed time-series and estimation results. Section 5 first derives the risk-neutral dynamics and closed-from option pricing, then discusses its implications by fitting into option data panels. Section 6 concludes. Proofs and technical details are provided in the appendix.

2 Empirical Analysis

In this section, I introduce the data used in the paper and show the connection between index realized variance and market liquidity empirically. Looking at data, it strongly signals the contemporaneous correlation as well as lagged dependence between realized variance and liquidity.

2.1 Data Description

First, I use daily close-to-close S&P500 index futures returns from Jan 2nd, 2004 to Dec 31st, 2012. This gives me 2,265 days of observations.

Next, I use 1-min grid of intra-day returns of S&P500 index futures to construct the realized variance (RV) measure. In order to reduce noise due to market micro-structure, I use the 15-min average RV estimator as suggested in Zhang, Mykland, and Ait-Sahalia (2005). Hence, starting from the opening time of the market to 14 minutes after the market opens, I compute 15 different sums of 15-minute squared returns and then average them to produce RV data. Then, I normalize the RV series in order to match the unconditional variance of returns to force the relationship $E[RV_t] = \text{Var}[R_t]$. See Hansen and Lunde (2005)
for further details regarding this adjustment.

Due to the lack of intra-day trade data for S&P500 index futures, I proxy for market liquidity by an equally-weighted average liquidity measure of the 500 firms belonging to S&P500 index. The effective spread computed using intra-day bid-ask spread from TAQ database is used for individual firm’s liquidity measure, as in Goyenko, Holden, and Trzcinka (2009) and Christoffersen, Goyenko, Jacobs, and Karoui (2014). Specifically, the effective spread of each trade is computed as

\[ IL_k^s = \frac{2|S_k^P - S_k^M|}{S_k^M} \]

where \( S_k^P \) is the price of executed trade and \( S_k^M \) is the mid-price of the prevailing market, so that it captures the relative difference between executed trade’s price and the prevailing mid-price. Then, the liquidity measure of each day is computed by taking dollar-volume weighted average of all trades occurred during that day. Finally, the aggregate market liquidity is computed by averaging across all firms belonging to the S&P500 index.

\[ IL^M = \frac{1}{N} \sum_{s=1}^{N} \frac{\sum_k DolVol_k IL_k^s}{\sum_k DolVol_k} \]

I use CBOE’s VIX index as a measure of risk-neutral volatility for each day. Following Bollerslev, Tauchen, and Zhou (2009), I use the difference between VIX and realized volatility as a measure of variance risk premium. This measure has an advantage being directly observable at the end of each day. BTZ shows that forward-looking expected variance risk premium measure is highly correlated to this definition of variance risk premium and it has similar forecasting ability.

The daily time-series of Realized Variance, Effective Spread, and VIX from 2004 to 2012 are plotted in Figure 1. I observe commonality between the effective spread and variance measures. They tend to spike simultaneously, but it is notable to see that the magnitudes of spikes are not always proportional. Especially, realized variance spikes dramatically during the financial crisis of 2007-2008 period, but the effective spread and VIX does not spike as much as realized variance during this time.

2.2 Statistical Evidences

In order to study the inter-dependence between the effective spread and variance measures, I plot the cross-correlation functions with lag orders from -10 to 10 in Figure 2. The horizontal lines denote the Bartlett 2-standard error confidence interval around zero. The most notable
observation is the contemporaneous correlation which corresponds to the lag order of 0. I see that the lagged cross-correlation spikes at the lag order zero with both realized variance and VIX, as well as the variance risk premium. The magnitude of correlation is also significantly high with all three measures.

To remove the plan autoregressive effects from the persistent time-series, I also fit each series to an ARMA(1,1) model and study the cross-correlations among its residuals which are plotted in Figure 3. I see that the contemporaneous correlations remain different from 0 in a statistically significant way, while lagged correlations mostly lose statistical significance. Interestingly, the lagged correlation between effective spread and realized variance residuals remains positive and non-zero for many positive lags while it does not differ statistically from 0 for negative lags. In other words, realized variance has a persistent impact on effective spreads, but not the other way around. However, no such a persistent impact is observed when looking at the risk-neutral variance. The lagged dependence dies out much more rapidly, and more interestingly, the lagged correlation of order -1 remains significant. Hence, I have a two-way causality relationship between the effective spread and the risk-neutral variance.

To study the lagged dependence structure, I also fit a VAR(1) model to the effective spread and variance measure pairs as well as all three of them together. The results of VAR(1) fit is reported in Table 1. As expected, all coefficients turn out to be statistically significant in the bivariate cases. The more interesting case is when effective spreads, realized variance, and variance risk premium are all fit together. Implied volatility is omitted since it is a linear combination of RV and VRP. I see that both the lagged realized variance and variance risk premium have significant impact on the effective spread, but not the other way around. Interestingly, the lagged variance risk premium loses its significance in explaining the current variance risk premium and the lagged realized variance is shown to be the strongest explanatory variable of the current variance risk premium. Lastly, I see an insignificant relationship between lagged effective spreads and both the realized variance and variance risk premium measures. Figure 4 plots the impulse response function implied by the VAR(1) estimation. The shock applied to the RV and VRP has a positive and persistent effect on the ES, which is consistent with the significant coefficient estimates. On the other hand, the shock applied to the ES on the RV and VRP rather dies out quickly and the magnitude is much smaller as expected from the insignificance of the coefficient estimates.

Overall, these tests provide ample evidences on the relationship between the effective spreads and the variance dynamics. The contemporaneous relationship seems to be the strongest, but the lead-lag relationships are not negligible. In particular, the lagged realized variance and variance risk premium have a significant impact on the future effective spread.
With these evidences, now I turn to building a model that incorporates such features.

3 Modelling Variance Dynamics with Liquidity

In this section, I propose two models that uses observed market liquidity when building the dynamics of index returns. The first model is a new extension of the Stochastic Volatility model of Heston (1993) where the liquidity affects the variance dynamics, but not returns. The second model is a simple generalization of two-factor Stochastic Volatility model (SV-2F) of Bates (2000) 3.

3.1 Stochastic Volatility Model with Liquidity embedded Variance dynamics (SV-RVL)

The analysis in section 2.2 suggests a strong contemporaneous relationship between realized variance and liquidity. Motivated by this intuition, I propose the following new model

\[
\frac{dS_t}{S_t} = \mu dt + \sqrt{V_t} dW_t^1 \\
dV_t = \kappa_1(\theta_1 - V_t) dt + \gamma dL_t + \xi_1 \sqrt{V_t} dW_t^2 \\
dL_t = \kappa_2(\theta_2 - L_t) dt + \xi_2 \sqrt{L_t} dW_t^3
\]

where \(\gamma\) controls the contemporaneous impact of a change in liquidity on variance. In this set-up, the liquidity factor does not appear directly in the stock price dynamics, but rather only implicitly through the variance process. The special case \(\gamma = 0\) is equivalent to Heston’s Stochastic Volatility model. There are only three Brownian innovations required for this model. I let the correlation between \(W_t^1\) and \(W_t^2\) capture the conventional leverage effect parameter, \(\rho\), while all other correlations are set to 0. Therefore, liquidity is not a priced risk factor in this model as its contribution is limited to variance dynamics only.

Conditional higher moments resulting from this specification are provided below.

\[
\text{Var}_t[dV_t] = (\xi_1^2 V_t + \gamma^2 \xi_2^2 L_t) dt \\
\text{Var}_t[dL_t] = (\xi_2^2 L_t) dt \\
\text{Cov}_t[dV_t, dL_t] = (\gamma \xi_2 \sqrt{L_t}) dt \\
\text{Corr}_t[dV_t, dL_t] = \left( \frac{\gamma \xi_2 \sqrt{L_t}}{\sqrt{\xi_1^2 V_t + \gamma^2 \xi_2^2 L_t}} \right) dt
\]

\footnote{Christoffersen, Heston, and Jacobs (2009) and Anderson, Fusari, and Todorov (2013,2014) also use two-factor model.}
In addition to the observed market liquidity process, I also use an observed realized variance (RV) process to identify the latent factor $V_t$. I assume the following measurement error structure for the observed RV and Liquidity process

$$\log(RV_{t+1}) = \log(E_t[\int_t^{t+1} V_s ds]) + \epsilon_{1,t+1} \quad (3.4)$$

$$\log(\bar{L}_{t+1}) = \log(E_t[\int_t^{t+1} L_s ds]) + \epsilon_{2,t+1} \quad (3.5)$$

where $\epsilon_t$ denotes error due to market micro-structure noise and estimation error, and where $RV_{t+1}$ and $\bar{L}_{t+1}$ are observed realized variance and liquidity processes. Note that both realized variance and observed liquidity are assumed to be the expected value of integrated variance and liquidity in logs plus measurement errors. I also assume that $\epsilon_{1,t+1}$ and $\epsilon_{2,t+1}$ are uncorrelated i.i.d. normal random variables with observation error variance equal to $\sigma_1$ and $\sigma_2$, which are parameters to be estimated as a part of the model estimation. Details of the calibration methodology are provided below.

The following proposition gives the term structure of variance in this model, which is also used to derive the closed-form for the conditional expectation in (3.4).

**Proposition 1.** In SV-RVL model, the expected conditional future variance is given by

$$E_t[V_T] = \theta_1 + (V_t - \theta_1)e^{-\kappa_1(T-t)} + [(L_t - \theta_2)\gamma \frac{\kappa_2}{\kappa_1 - \kappa_2}](e^{-\kappa_1(T-t)} - e^{-\kappa_2(T-t)}) \quad (3.6)$$

**Proof.** See Appendix. 

Note that the long-term mean of the variance process $V_t$ is still equal to $\theta_1$ which can be obtained by letting $t \to \infty$. This expression shows how different speeds of mean reversion of latent variance factor and liquidity factor affect the term structure. Observe that the contribution from mean-reversion part of liquidity $(L_t - \theta_2)$ is always amplified by the negative factor $\gamma \frac{\kappa_2}{\kappa_1 - \kappa_2}(e^{-\kappa_1(T-t)} - e^{-\kappa_2(T-t)})$ in the term structure of variance. As a result, under the assumption that current level of variance $V_t$ is equal to its long-run variance $\theta_1$, if the current level of liquidity $L_t$ is above its long-term mean $\theta_2$, then the term structure of variance is downward sloping first and eventually mean-reverts to its long-term mean $\theta_1$. On the other hand, when $L_t$ is below $\theta_2$, the slope of term structure of variance is upward sloping first, then mean-reverts to $\theta_1$. In other words, liquid current market (lower $L_t$) implies upward sloping first, then downward sloping term structure of variance while illiquid current market (higher $L_t$) implies downward sloping term structure of variance. Figure 5 graphically demonstrates the various term structure this model can produce under different parametrizations. Top panel plots the case where $\kappa_1$ is very close to $\kappa_2$ hence two series have similar speed of
mean-reversion. Second panel plots the case where $\kappa_1$ is much smaller than $\kappa_2$, and last panel plots the opposite case. Expected conditional future variance mean-reverts extremely fast when the speed of mean-reversion is close to each other, while it takes much longer to converge when two speeds of mean-reversion is far apart. In particular, when the speed of mean-reversion of liquidity factor $\kappa_2$ is high, it takes much longer to converge than the opposite case.

3.2 Two-Factor Stochastic Volatility Model with Realized Variance and Liquidity (SV-RVL2F)

Bates(2000) assume the following process for stock price with two latent variance processes plus jump. Here, I exclude the jump component and focus on the diffusive variance components only.

$$
\frac{dS_t}{S_t} = \mu dt + \sqrt{V_{1,t}}dW_{1,t} + \sqrt{V_{2,t}}dW_{2,t}
$$

$$
dV_{1,t} = \kappa_1(\theta_1 - V_{1,t})dt + \xi_1\sqrt{V_{1,t}}dW_{3,t}
$$

$$
dV_{2,t} = \kappa_2(\theta_2 - V_{2,t})dt + \xi_2\sqrt{V_{2,t}}dW_{4,t}
$$

where the Brownian motions $W_{1,t}^1$ and $W_{3,t}^3$ have a correlation of $\rho_1$, $W_{2,t}^2$ and $W_{4,t}^4$ have a correlation of $\rho_2$, and all others are uncorrelated. In this model, the variance processes $V_1$ and $V_2$ are both latent and typically need to be filtered from returns data. Instead, I propose to replace one of the factor, say $V_2$, with a liquidity process while preserving the two-factor structure. Hence, the model dynamics is given by

$$
\frac{dS_t}{S_t} = \mu dt + \sqrt{V_{t}}dW_{1,t} + \gamma\sqrt{L_{t}}dW_{2,t}^{2}
$$

(3.7)

$$
dV_{i,t} = \kappa_i(\theta_i - V_{i,t})dt + \xi_i\sqrt{V_{i,t}}dW_{3,t}^{3}
$$

(3.8)

$$
dL_{t} = \kappa_2(\theta_2 - L_{t})dt + \xi_2\sqrt{L_{t}}dW_{4,t}^{4}
$$

(3.9)

The new parameter $\gamma$ controls the portion of the total variance explained by the liquidity component. Note that when $\gamma = 1$ the new model is equivalent to the 2-factor stochastic volatility model and when $\gamma = 0$ the model is equivalent to the Heston’s stochastic volatility model. Note that the total spot variance can be decomposed as follows

$$
\text{Var}_t\left(\frac{dS_t}{S_t}\right) = (V_t + \gamma^2L_t)dt
$$

(3.10)
hence $\frac{\gamma^2 L_t}{V_t + \gamma^2 L_t}$ can be interpreted as the fraction of the total conditional variance explained by liquidity.

Model implied moments between total variance and liquidity are straightforward to compute, we get

$$\text{Var}_t[dL_t] = (\xi^2 L_t)dt$$
$$\text{Cov}_t[dS_t, dL_t] = (\gamma \xi \rho_2 L_t)dt$$
$$\text{Corr}_t[dS_t, dL_t] = \left(\frac{-\gamma \rho_2 \sqrt{V_t}}{\sqrt{V_t + \gamma^2 L_t}}\right)dt$$

I again assume the measurement error structure from equations (3.4)-(3.5), namely

$$\log(RV_{t+1}) = \log(E_t[\int_t^{t+1} (V_s + \gamma^2 L_s)ds]) + \epsilon_{1,t+1}$$  (3.11)
$$\log(\bar{L}_{t+1}) = \log(E_t[\int_t^{t+1} L_s ds]) + \epsilon_{2,t+1}$$  (3.12)

In this model, the conditional expected future variance is simply the sum of two individual factor’s expected variance,

$$E_t[\tilde{V}_T] = [\theta_1 + (V_t - \theta_1)e^{-\kappa_1(T-t)}] + \gamma^2 [\theta_2 + (L_t - \theta_2)e^{-\kappa_2(T-t)}]$$  (3.13)

where $\tilde{V}_t$ denotes total variance. From this, the conditional expectation of the integrated variance is straightforward to compute. As shown in Christoffersen, Heston, and Jacobs (2009), the term structure of future variance is mainly determined by the decomposition of current variance into the two latent factors. As a result of this decomposition, the model can deliver flexible term structure for the same current total spot variance level.
3.3 Uni-variate Benchmark Models

To study the impact of having liquidity as the second factor, I compare my models to the following two uni-variate cases.

\[ \frac{dS_t}{S_t} = \mu dt + \sqrt{V_t} dW_t^1 \]
\[ dV_t = \kappa(\theta - V_t) dt + \xi \sqrt{V_t} dW_t^2 \]
\[ \log(RV_{t+1}) = \log(E_t[\int_t^{t+1} V_s ds]) + \epsilon_{1,t+1} : \text{SV-RV Model} \]
\[ \log(L_{t+1}) = \log(E_t[\int_t^{t+1} V_s ds]) + \epsilon_{1,t+1} : \text{SV-L Model} \]

The first model is the basic Heston’s SV model where the only difference being that RV is variance plus an error term. The second model has the same structure but now I assume that liquidity measure is a proxy for observed variance. In the second model, the liquidity measure has been normalized to have the same sample mean as the realized variance to ensure the fair comparison. If observed liquidity does not have a meaningful contribution to the variance dynamics, then the first uni-variate (SV-RV) model will perform as well as the two-factor models introduced in section 3.1 and 3.2. On the other hand, if liquidity alone can explain the variance dynamics, then the second uni-variate (SV-L) model will perform as well as the two-factor models. Thus, these two benchmarks will provide insight into how the observed liquidity measure contributes to the variance dynamics.

4 Estimation Methods and Fit

In the previous section, I developed the two-factor models designed to capture the impact of liquidity on variance dynamics, as well as the benchmark models for comparison. Now, the task is the estimation of these models on observed realized variance and liquidity. I first describe the estimation procedures, then discuss estimation results and its implications.

4.1 Extended Kalman Filtering Method

Traditional estimation of stochastic volatility models focus on extracting the latent variance dynamics from the observed time-series of returns. Several methods have been proposed including MCMC (Jacquier, Polson, and Rossi (1994)), GMM (Pan (2002)), and the particle filter (Johannes and Polson (2009)). However, I differ from the previous approaches because I have additional observed data, namely realized variance and liquidity. Consequently, my
estimation relies on the extended Kalman Filter (EKF) applied to the discretized system of state-space equations. The proposed EKF procedure has advantage over the other methods in at least two ways. First, it filters the latent states in the physical probability measure using observables thus does not require the specification of the pricing kernel, or the risk premiums, in the filtering process. Thus it allows to compare different models based on how well it explains the observables in the physical probability measure without using the risk neutral observables such as options. Next, the computational time of filtering latent states are dramatically faster than all other methods as it does not rely on any simulation or numerical integrations. However, it needs to be noted that the purpose of these methods are to understand the relationship between the market liquidity and variance dynamics, not to provide precise estimation of the latent state vectors and option prices. I rather focus on the implications of the reduced form model specifications to see how much and to what degree the market liquidity impacts the variance dynamics under the both physical and risk neutral measures.

First, I need to discretize the underlying dynamics given by (3.1)-(3.3) and (3.7)-(3.9). I rely on a simple Euler scheme which is the standard approach in the literature (Eraker (2001)). The complete system of state transition equations for the SV-RVL2F model is given by

\[ V_{t+1} = V_t + \kappa_1(\theta_1 - V_t)\Delta t + \xi_1\sqrt{V_t}\sqrt{\Delta t}u_{1,t} \]  
\[ L_{t+1} = L_t + \kappa_2(\theta_2 - L_t)\Delta t + \xi_2\sqrt{L_t}\sqrt{\Delta t}u_{2,t} \]  

and for the SV-RVL model as

\[ V_{t+1} = V_t + \kappa_1(\theta_1 - V_t)\Delta t + \gamma(L_{t+1} - L_t) + \xi_1\sqrt{V_t}\sqrt{\Delta t}u_{1,t} \]  
\[ L_{t+1} = L_t + \kappa_2(\theta_2 - L_t)\Delta t + \xi_2\sqrt{L_t}\sqrt{\Delta t}u_{2,t} \]  

where \( u_{i,t} \) are independent normal innovations with mean 0 and unit variance. The covariance structure of the system is straightforward to compute from the initial model specifications given by (3.1)-(3.3) and (3.7)-(3.9).

The measurement equation consists of three observables - returns, realized variances, and market liquidity measured by effective spread. To avoid having state transition innovations and measurement errors being correlated arising from the leverage effect parameter \( \rho \), I rewrite the returns equation as follows. First, I apply Ito’s lemma to the index price dynamics
(3.1) and (3.7) to get the equation for the instantaneous returns.

\[
d\log(S_t) = (\mu - \frac{1}{2}V_t - \frac{1}{2}\gamma^2 L_t)dt + \sqrt{V_t}dW_t^1 + \gamma\sqrt{L_t}dW_t^2
\]

(4.5)

\[
d\log(S_t) = (\mu - \frac{1}{2}V_t)dt + \sqrt{V_t}dW_t^1
\]

(4.6)

Next, I re-write the equation in terms of independent Brownian motions.

\[
d\log(S_t) = (\mu - \frac{1}{2}V_t - \frac{1}{2}\gamma^2 L_t)dt + \sqrt{V_t}(\rho_1dW_t^3 + \sqrt{1 - \rho_1^2}dW_t^1) + \gamma\sqrt{L_t}(\rho_2dW_t^4 + \sqrt{1 - \rho_2^2}dW_t^2)
\]

(4.7)

\[
d\log(S_t) = (\mu - \frac{1}{2}V_t)dt + \sqrt{V_t}(\rho dW_t^2 + \sqrt{1 - \rho^2}dW_t^1)
\]

(4.8)

Last, I substitute the Brownian motions that can be written in terms of the latent state dynamics as below.

\[
d\log(S_t) = (\mu - \frac{1}{2}V_t - \frac{1}{2}\gamma^2 L_t)dt + \sqrt{V_t}\sqrt{1 - \rho_1^2}dW_t^1 + \gamma\sqrt{L_t}\sqrt{1 - \rho_2^2}dW_t^2 + \rho_1(\xi_1 dV_t - \kappa_1(\theta_1 - V_t)dt) + \rho_2\gamma(\xi_2 dL_t - \kappa_2(\theta_2 - L_t)dt)
\]

(4.9)

\[
d\log(S_t) = (\mu - \frac{1}{2}V_t)dt + \sqrt{V_t}\sqrt{1 - \rho^2}dW_t^1 + \rho(\xi (dV_t - \kappa(\theta - V_t)dt)
\]

(4.10)

Finally, discretizing the above gives the final measurement equation that involves lagged latent states. As commonly done in the filtering literature, I add two lagged states \(V_t\) and \(L_t\) that are updated without an error to facilitate the filtering process.

\[
\begin{align*}
  r_{t+1} &= (\mu - \frac{1}{2}V_t - \frac{1}{2}\gamma^2 L_t)\Delta t + \sqrt{V_t}(1 - \rho_1^2) + \gamma^2 L_t(1 - \rho_2^2)\epsilon_{1,t+1} + \frac{\rho_1}{\xi_1}(V_{t+1} - V_t - \kappa_1(\theta_1 - V_t)\Delta t) + \frac{\rho_2\gamma}{\xi_2}(L_{t+1} - L_t - \kappa_2(\theta_2 - L_t)\Delta t) \\
  r_{t+1} &= (\mu - \frac{1}{2}V_t)\Delta t + \sqrt{V_t}\sqrt{1 - \rho^2}\epsilon_{1,t+1} + \rho (V_{t+1} - V_t - \kappa(\theta - V_t)\Delta t)
\end{align*}
\]

(4.11)

(4.12)

The observation equations for RV and liquidity were already provided in equations (3.4)-(3.5) and (3.11)-(3.12) in section 3. Given the complete system of transition and observation equations, I estimate the model parameters by Maximum Likelihood Estimation using the Extended Kalman Filtering (EKF) algorithm. This gives me the optimal parameter set that
describes $\mathbb{P}$-dynamics. The variance of the observation error is not easy to calibrate and it is therefore estimated as a separate parameter. At each iteration, the observation error is implied by the unbiased estimator from the previous iteration and I continue iterating until the error variance converges. The error variance converges rather quickly. It typically stabilizes after 5 to 6 iterations.

Unlike the previous literatures in which the parameters and underlying processes are often re-estimated every year, I only estimate the parameter once for the whole sample which is a much more stringent of the model. Once the optimal parameters are obtained, I can filter out the true $V_t$ and $L_t$ process that will be used as the spot variance and liquidity, respectively, when valuing options.

4.2 Maximum Likelihood Estimation

I use daily returns, realized variances, and liquidity from Jan 2\textsuperscript{nd}, 2004 to Dec 31\textsuperscript{st}, 2012. All data are constructed using the method described in section 2. The results from maximum likelihood estimation are summarized in Table 2. The drift of return dynamics $\mu$ is set at unconditional mean of returns for all models.

First, note that the estimate of the speed of mean-reversion parameter $\theta$ is 4.0064 and 2.2024 for the uni-variate models using RV and liquidity data, respectively. This reflects the fact that the observed series of RV and liquidity are very persistent. However, the parameter values for the remaining latent variance factor in the two-factor models are 5.1449 and 5.8908 for SV-RVL and SV-RVL2F model, respectively. This implies that the portion of the variance that is not explained by liquidity is less persistent than the one factor case.

The estimate of $\rho$ is always negative capturing the well documented leverage effect in index returns. Note that the magnitude of $\rho$ is much smaller for the two-factor models at around -0.5 while the one-factor model has a value around -0.63. This can be interpreted as the liquidity factor explaining more of the leverage effect than the non-liquidity factor. In fact, the $\rho$ estimate implied by the liquidity factor in the SV-RVL2F model is very large in magnitude at -0.7250.

The resulting log-likelihood value from MLE shows the superior performance of the two-factor models in explaining observed returns. The log-likelihood value is 17,215 under the SV-RV model. It increases to 17,350 under the SV-RVL2F model, and it increases even further to 17,610 under the SV-RVL model. Recall that the estimation is done by jointly fitting the three observed time-series of returns, realized variances, and liquidity. Hence, I am not trying to solely maximize the fit on returns even in the two-factor models. This superior performance confirms the ability of observed liquidity in explaining the index return
variance dynamics. In addition, the log-likelihood resulting from the SV-L model is quite low at 15,704 indicating that observed liquidity alone is not capable of explaining the index return variance dynamics.

Figure 6 plots the filtered volatility from different models. The first two panels which plot the one-factor models versus each of the two factor models confirm that liquidity cannot explain the variation needed to fit index returns. Also, filtered volatility in the SV-RVL2F model is less volatile than in the SV-RV model, but filtered volatility in the SV-RVL model is noisier than in the SV-RV model. The last panel in Figure 6 confirms this by comparing the two-factor models only.

Table 3 provides summary statistics of filtered volatility from different models. Overall, the moments turn out to be quite similar to each other across models except for the SV-L model. The two-factor models also imply higher skewness and kurtosis than the SV-RV model while simultaneously implying a lower standard deviation.

To verify the validity of the normality assumption behind the measurement equations (3.4)-(3.5) and (3.11)-(3.12), Figure 7 contains QQ-plots of the filtered measurement error residuals from different models. They show that -as assumed- the residuals are reasonably close to normally distributed with the exception of the SV-L model. This confirms the inability to explain variance from the observed liquidity alone.

The SV-RVL2F model enables a straight-forward decomposition of variance as shown in equation (3.10). It provides a decomposition of spot variance into its liquidity and non-liquidity component. Figure 8 plots the percentage of spot variance explained by the liquidity component over time. The figure is perhaps surprising in that liquidity explains a larger portion during normal times than recession periods. A huge drop in the percentage during the financial crisis of 2007-2008 is observed where the liquidity factor can only explain around 10% of spot variance during this period. This could be partly due to the fact that liquidity measure I am using is the effective spread which does not spike as much as realized volatility. Perhaps, the liquidity of stock markets was not as much of an issue as funding markets in the crisis period. Regardless, it confirms the intuition that significant portion of index return variance is explained by the liquidity component.

5 Option Valuation

Results from the previous section suggests an importance of liquidity factor in explaining variance dynamics. As variance dynamics is a crucial factor in option valuation, implications of liquidity on options is the next task I describe. In this section, I derive the risk-neutral dynamics of returns for each model and discuss the estimation of corresponding risk-neutral
parameters. Then, each model is fitted to option data panels and its implications are discussed.

5.1 Risk-Neutralization

In order to use the new model for option valuation, I first need to risk-neutralize the physical process introduced in Section 3. I follow the standard approach that assumes a pricing kernel that is log-linear in the Brownian innovations.

\[
\frac{dQ_t}{dP_t} = \mathcal{E}(\int_0^t \gamma_s dW_s) \tag{5.1}
\]

where \(\mathcal{E}\) denotes the stochastic exponential, \(\gamma_s\) denotes the vector of market prices of risk, and \(W_s\) is vector of Brownian motions. As in common in the literature, I assume the market prices of risk are parametrized as \(\lambda_V \sqrt{V_t}\) and \(\lambda_L \sqrt{L_t}\). I can interpret them, respectively, as the market price of the non-liquidity component and liquidity component of volatility. When using this linear parametrization it is straightforward to compute the risk-neutral process in each model. Lastly, I do not impose the risk premium placed on the liquidity in the SV-RVL model as the liquidity innovation \(dW_t^3\) does not affect the returns in the model specification.

**Proposition 2.** Given the Radon-Nikodym derivative given by (5.1), the two models introduced in Section 3 have the following dynamics under the risk-neutral measure. The SV-RVL2F Model is given by

\[
\frac{dS_t}{S_t} = r dt + \sqrt{V_t} dW_t^1 + \gamma \sqrt{L_t} dW_t^2
\]

\[
dV_t = \kappa_1^* (\theta_1^* - V_t) dt + \xi_1 \sqrt{V_t} dW_t^3
\]

\[
dL_t = \kappa_2^* (\theta_2^* - L_t) dt + \xi_2 \sqrt{L_t} dW_t^4
\]

The SV-RVL Model is given by

\[
\frac{dS_t}{S_t} = r dt + \sqrt{V_t} dW_t^1
\]

\[
dV_t = \kappa_1^* (\theta_1^* - V_t) dt + \gamma dL_t + \xi_1 \sqrt{V_t} dW_t^2
\]

\[
dL_t = \kappa_2 (\theta_2 - L_t) dt + \xi_2 \sqrt{L_t} dW_t^3
\]
where the parameter mappings are

\[
\begin{align*}
\kappa_1^* &= \kappa_1 + \xi_1 \lambda V, \\
\theta_1^* &= \frac{\kappa_1 \theta_1}{\kappa_1^*} \\
\kappa_2^* &= \kappa_2 + \xi_2 \lambda L, \\
\theta_2^* &= \frac{\kappa_2 \theta_2}{\kappa_2^*}
\end{align*}
\]

**Proof.** It is a straightforward computation applying Girsanov theorem to each Brownian motion. \qed

In order to value an option in closed form under the risk-neutral dynamics above, I need to derive the corresponding characteristic function of the log-spot index price. Since the SV-RVL2F model is identical in structure to the two-factor stochastic volatility model of Christoffersen, Heston, and Jacobs (2009), the resulting characteristic function is similar. For the SV-RVL model, the solution is similar to Heston (1993)’s case as the liquidity factor is not a priced risk factor.

**Proposition 3.** Let’s denote the risk-neutral characteristic function of log-spot price by

\[
E_t[\exp(iu \log(S_{t+\tau}))] = S_t^u f(u, \tau, V_t, L_t)
\]

Then function \( f \) is given by

\[
f(u, \tau, V_t, L_t) = \exp(A(u, \tau) + B_1(u, \tau)V_t + B_2(u, \tau)L_t)
\]

For SV-RVL2F model, \( A, B_1, \) and \( B_2 \) are given below.

\[
\begin{align*}
A(u, \tau) &= ru\tau + \frac{\kappa_1 \theta_1}{\xi_1^2}[(\kappa_1 - \rho_1 \xi_1 u\bar{u} + d_1)\tau - 2\log(\frac{1 - g_1 \exp(d_3 \tau)}{1 - g_1})] \\
&\quad + \frac{\kappa_2 \theta_2}{\xi_2^2}[(\kappa_2 - \rho_2 \xi_2 u\bar{u} + d_2)\tau - 2\log(\frac{1 - g_2 \exp(d_2 \tau)}{1 - g_2})] \\
B_j(u, \tau) &= \frac{\kappa_j - \rho_j \xi_j u\bar{u} + d_j}{\xi_j^2} \left( \frac{1 - \exp(d_j \tau)}{1 - g_j \exp(d_j \tau)} \right) \\
g_j &= \frac{\kappa_j - \rho_j \xi_j u\bar{u} + d_j}{\kappa_j - \rho_j \xi_j u\bar{u} - d_j} \\
d_j &= \left( \rho_j \xi_j u\bar{u} - \kappa_j \right)^2 + \xi_j^2 (u\bar{u} + u^2)
\end{align*}
\]

for \( j = 1, 2 \). And for SV-RVL model, \( A, B_1, \) and \( B_2 \) are given as the solution to the following
Ricatti ODE with the initial conditions $A(0) = B_1(0) = B_2(0) = 0$.

\[
\begin{align*}
B_0(u, \tau) &= iu \\
\frac{dA}{d\tau} &= riu + (\kappa_1 \theta_1 + \gamma \kappa_2 \theta_2)B_1 + \kappa_2 \theta_2 B_2 \\
\frac{dB_1}{d\tau} &= -\frac{1}{2} u(i + u) - (\kappa_1 - \rho \xi_1 iu)B_1 + \frac{1}{2} \xi_1^2 B_1 \\
\frac{dB_2}{d\tau} &= \frac{1}{2} \gamma^2 \xi_2 \xi_1^2 B_1 - \gamma \kappa_2 B_1 - (\kappa_2 - \gamma \xi_2 B_1)B_2 + \frac{1}{2} \xi_2^2 B_2 
\end{align*}
\]

Proof. See Appendix. \hfill \Box

Note that in SV-RVL model, the resulting Ricatti ODE cannot be solved analytically thus I rely on 4th order Runge-Kutta algorithm to numerically solve the equation. Since the ODE needs to be solved only once at the beginning at the coefficients does not depend on time, two models do not differ a lot in terms of numerical complexity.

Following Heston (1993), European call options can be valued using

\[
C_t = S_t P_1 - K e^{-r \tau} P_2
\]

where the $P_1$ and $P_2$ probabilities are computed using Fourier inversion:

\[
\begin{align*}
P_1 &= \frac{1}{2} + \frac{1}{\pi} \int_0^\infty Re\left[\frac{e^{iu \log(S_t)}}{iu S_t e^{r \tau}} f(u + 1, \tau, V_t, L_t)\right] du \\
P_2 &= \frac{1}{2} + \frac{1}{\pi} \int_0^\infty Re\left[\frac{e^{iu \log(S_t)}}{iu} f(u, \tau, V_t, L_t)\right] du
\end{align*}
\]

The integrands in the above expression vanish quickly and can be computed effectively using a numerical integration scheme.

5.2 Parameter Estimation with Options

Following Christoffersen, Heston, and Jacobs (2014), I will rely on sequential estimation to infer the parameters of $Q$ dynamics. In other words, I will keep the physical parameters estimated from EKF and will fit to the models to the observed option prices only to infer the market price of risk parameters $\lambda_V$ and $\lambda_L$ only. In this way, I can fully explore the explanatory power of the observable variables as I do not rely on any of the $Q$ estimated latent factors. Hence, I have an advantage over the previous approaches in the sense that I start with the daily observed time series of the underlying factors. Conventional joint estimation of returns and options suffer from the over-fitting on options thus resulting risk-
neutral dynamics dominating the parameter estimates. Therefore, sequential estimation procedure provides a way to study a joint dynamics of physical and risk-neutral measures without such an over-fitting problem.

For computational efficiency, I minimize Vega-Weighted Root Mean Squared Error which has been shown (Trolle and Schwartz (2009)) to proxy for the more numerically involved Implied Volatility Root Mean Squared Error (IVRMSE). When I use VWRMSE, I do not need to compute Black-Scholes implied volatility for the model prices each iteration. Also, I only need to estimate $\lambda$ parameters which makes the computation extremely efficient and it converges quickly. Numerically, I thus solve

$$[\hat{\lambda}_V, \hat{\lambda}_L] = \arg \min_{\lambda_V, \lambda_L} \sum_{m=1}^{N} \frac{(C_{m,M} - C_{m}(\Theta^p, \lambda_V, \lambda_L, V_t, L_t))^2}{\text{Vega}_m^2}$$

(5.2)

where $C_{m,M}$ is the observed market mid-price of $m^{th}$ call option and $C_{m}(\Theta^p, \lambda_V, \lambda_L, V_t, L_t)$ is the model-implied call option price, and Vega$_m$ is the Black-Scholes Vega that measures the sensitivity of the call option price with respect to volatility.

5.3 Option-based Estimation Results

I use closing mid-prices of European S&P500 index options from Jan 7$^{th}$, 2004 to Dec 26$^{th}$, 2012 downloaded from OptionMetrics database. For ease of interpretation, all put options are converted into calls using the put-call-parity relationship. To ensure the contracts I use have enough trading activity, I only pick out-of-the-money options with maturity from 15 to 180 days. In addition, only options from each Wednesday are selected for estimation to minimize the computational burden while maintaining a representative dataset. Standard filters from Bakshi, Cao, and Chen (1997) are applied and all contracts that does not satisfy no-arbitrage conditions are also removed. This yields 11,734 option contracts.

Table 4 summarizes the option data by its moneyness, maturity, and corresponding VIX level. Average implied volatility from the first panel confirms the well-documented volatility smirk effect. Out-of-the-money options have much larger implied volatility then in-the-money options.

As documented in Christoffersen, Heston, and Jacobs (2009), a major advantage of the two-factor stochastic volatility model is its flexibility in generating variance term-structure slopes. Equation (3.6) and (3.13) can be used to compute the conditional expectations of future variances where I plot the term structure of variance resulting from each model in Figure 9. The first panel is from the SV-RVL2F model with equal total spot variance of 0.04, but different percentage explained by the liquidity component. It verifies a role played
by the second liquidity factor. It can generate different slope structure, both upward and downward, by having different decomposition of the spot variance. On the other hand, it is less obvious from equation (3.6) how the liquidity component affects the term structure in the SV-RVL model. Since the spot variance is no longer decomposed into two additive non-liquidity and liquidity components, I plot the term structure with a fixed spot variance of 0.04, but with different spot liquidity levels ranging from 0.01 to 0.04. The resulting plot is quite similar to the SV-RVL2F model which again generates different term structure slopes. Hence, I conclude that both models have enough flexibility to capture different variance term-structure slopes.

Table 5 reports the option-based estimation result. Only $\lambda$ parameters are estimated and others come from EKF estimation. The $\lambda_V$ parameter estimated from the SV-RV model is significantly negative which is consistent with previous findings (Bollerslev, Gibson, and Zhou (2011)). Similarly, the $\lambda_L$ parameter estimated from the SV-L model is also negative, but the magnitude is much lower. As expected, the $\lambda_V$ parameter estimated from the SV-RVL model is also significantly negative resulting in long run mean parameter $\theta_1^*$ of 0.0482 that is similar to the value 0.0492 of SV-RV model.

The $\lambda$ estimates from the SV-RVL2F are informative about the risk-neutral dynamics implied by the options. The $\lambda_V$ is significantly negative as it represents the volatility risk premium of the latent volatility factor. However, $\lambda_L$ representing the market price of risk is strongly positive and very high in magnitude. Consequently, the risk-neutral long-run average of liquidity process $\theta_2^*$ becomes 0.0035 being very close to the perfectly liquid market value of 0. Also the speed of mean reversion for liquidity process under the risk neutral measure $\kappa_2^*$ is around 34 which indicates the risk neutral measure expects the any form of market illiquidity to disappear very quickly.

Log-likelihoods obtained by minimizing VWRMSE from each model shows that the SV-RV model and the SV-RVL model performs similar while the SV-RVL2F model shows superior performance. As expected, the SV-RVL model performs the worst, reflecting the inadequacy of liquidity alone to explain the variance. The last panel of Table 5 also reports more commonly used metric of option fit, IVRMSE. The SV-RVL model outperforms the SV-RV model by 5% while their VWRMSE values are similar. In both the VWRMSE and IVRMSE metric, the SV-RVL2F model outperforms the SV-RV model by around 10&. The relatively small improvements in overall fit are not surprising to see as my models are imposing more structures on observed liquidity in comparison to the models which imposes no restrictions to fit observed variables.

Table 6 and 7 summarize the performance measured by IVRMSE and IV Bias, respectively, of each model sorted by moneyness, maturity, and VIX level. Notably, the two-factor
models outperform the one-factor models for out-of-the-money options, but the SV-RV model outperforms the two-factor models for deep in-the-money options as shown in the top panel of Table 6. The first finding is consistent with the result of Cetin et al (2006) where they find that liquidity helps explaining out-of-the-money options, but less so for in-the-money options. IV Bias is perhaps more informative than IVRMSE as a performance criteria across moneyness which is reported in Table 7. The SV-RV model over-prices in-the-money options while it under-prices out-of-the-money options. The SV-RVL model now under-prices deep in-the-money options, over-prices in-the-money options, but then it over-prices out-of-the-money options. Lastly, the SV-RVL2F model slightly under-prices in all dimensions. Note that the magnitude of IV Bias is much lower for the two-factor models than the SV-RV model as plotted in Figure 10. Figure 10 shows the advantage of having the second liquidity factor in my models. The SV-RV model fails to capture the level and slope of implied volatility simultaneously, hence displaying an increasing pattern of fitting error by moneyness. My two-factor models can partially resolve this problem as both models outperform in the out-of-the-money dimension while the fitting error is more evenly distributed across all moneyness categories. Hence, the two-factor models do not suffer from the systematic patterns of fitting error as the one-factor models do.

6 Summary and Conclusions

Market liquidity and variance have been shown to have a significant relationship in the literature. In this paper, I use the aggregate effective spreads as a proxy for market liquidity and show the significant relationships between observed realized variance and liquidity. A strong contemporaneous relationship as well as lead-lag relationships are found between the two time-series.

Motivated by these findings, I develop a two-factor stochastic volatility model in which the observed liquidity factor affects the index return variance. I propose two different models, one being the straight-forward generalization of the two-factor stochastic volatility model while the other model is designed to focus on the impact of liquidity on variance alone. I also use the observed realized variance to infer a latent spot variance factor using a state-space representation. Standard Extended Kalman Filter estimation reveals the significance of the liquidity component in explaining the return dynamics in both models.

As both models are of the affine-form, I derive the closed-form option pricing formula under each specification. By using observed time-series, I directly filter out the underlying state vector which is then used as an input for the pricing of options. Sequential estimation of the risk premium parameter by minimizing option error reveals a superior performance
of the two-factor model with liquidity in explaining the option prices, especially out-of-the-money call options. Consequently, it is shown that observed liquidity indeed plays a role in capturing the slope factor in the implied volatility term structure.

Overall, my findings suggest that liquidity is an important determinant of return volatility as well as of the implied volatility term structure of options. This provides a plausible economic interpretation of the previous two-factor models where all factors were assumed to be latent. It also adds value to the volatility modelling literature that attempts to explain volatility using observed financial variables. Hence, it would be interesting to explore further on other explanatory factors of volatility dynamics and its implications on option prices. Lastly, using the a liquidity measure constructed from the intra-day S&P500 futures contract data would be of interest.

A  Proof of Proposition 1

Denote the conditional expectation of variance as \( x_T = E_t[V_t] \). Taking expectations on the integral form of the equation (3.2), then taking derivative with respect to \( T \) gives:

\[
    x_T = x_t + \int_t^T (\kappa_1(\theta_1 - x_s) + \gamma \kappa_2(\theta_2 - E_t[L_s])) ds
\]

\[
    \frac{dx_T}{dT} = \kappa_1(\theta_1 - x_T) + \gamma \kappa_2(\theta_2 - E_t[L_T])
\]

where \( E_t[L_T] \) can be computed using the integrating factor. Substituting above equation to the integrating factor technique again gives:

\[
    \frac{dx_T e^{\kappa_1 T}}{dT} = \theta_1 \kappa_1 e^{\kappa_1 T} + (\theta_2 - L_t) \gamma \frac{\kappa_2}{\kappa_1 - \kappa_2} e^{(\kappa_1 - \kappa_2)(T-t)}
\]

Then, solving above ODE with the initial condition \( x_t = V_t \) gives the equation (3.6).

B  Proof of Proposition 3

I follow Duffie et al. (2000) to derive the characteristic function. Re-writing as \( dW_t^1 = dB_t^1 \), \( dW_t^2 = \rho dB_t^1 + \sqrt{1-\rho^2} dB_t^2 \), and \( dW_t^3 = dB_t^3 \), where all \( dB_t^i \) are independent Brownian
motions, I get the following affine expression:

\[
\begin{align*}
\frac{d\ln(S_t)}{dt} &= (r - \frac{1}{2}V_t)dt + \sqrt{V_t}dB_t^1 \\
\frac{dV_t}{dt} &= [(\kappa_1\theta_1 + \gamma\kappa_2\theta_2) + (-\kappa_1V_t - \gamma\kappa_2L_t)]dt + \xi_1\rho\sqrt{V_t}dB_t^1 + \xi_1\sqrt{1 - \rho^2}\sqrt{V_t}dB_t^2 + \gamma\xi_2\sqrt{L_t}dB_t^3 \\
\frac{dL_t}{dt} &= (\kappa_2\theta_2 - \kappa_2L_t)dt + \xi_2\sqrt{L_t}dB_t^3
\end{align*}
\]

Under the following notation

\[
\begin{align*}
\mu(x_t) &= K_0 + K_1\ln(S_t) + K_2V_t + K_3L_t \\
\sigma(x_t)\sigma(x_t)^T &= H_0 + H_1\ln(S_t) + H_2V_t + H_3L_t
\end{align*}
\]

I have

\[
\begin{align*}
K_0 &= \begin{pmatrix} r \\ \kappa_1\theta_1 + \gamma\kappa_2\theta_2 \\ \kappa_2\theta_2 \end{pmatrix} \\
K_1 &= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\
K_2 &= \begin{pmatrix} -\frac{1}{2} \\ -\kappa_1 \\ 0 \end{pmatrix} \\
K_3 &= \begin{pmatrix} 0 \\ -\gamma\kappa_2 \\ -\kappa_2 \end{pmatrix} \\
H_0 &= H_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\
H_2 &= \begin{pmatrix} 1 & \xi_1\rho & 0 \\ \xi_1\rho & \xi_1^2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\
H_3 &= \begin{pmatrix} 0 & 0 & 0 \\ \gamma\xi_2^2 & \gamma\xi_2^2 & \xi_2^2 \\ 0 & \gamma\xi_2^2 & \xi_2^2 \end{pmatrix}
\end{align*}
\]

Duffie et al. (2000) states that coefficients of the characteristic function is then given by the following system of Ricatti equations.

\[
\begin{align*}
\frac{dA}{dt} &= -K_0^TB - \frac{1}{2}B^TH_0B \\
\frac{dB_0}{dt} &= -K_1^TB - \frac{1}{2}B^TH_1B \\
\frac{dB_1}{dt} &= -K_2^TB - \frac{1}{2}B^TH_2B \\
\frac{dB_2}{dt} &= -K_3^TB - \frac{1}{2}B^TH_3B
\end{align*}
\]
where \( B \) denotes a vector \([B_0, B_1, B_2]\). Plugging in the values from the above and re-writing the differential equation with respect to \( \tau = (T-t) \) for computational convenience, we have

\[
\begin{align*}
\frac{dA}{d\tau} &= rB_0 + (\kappa_1 \theta_1 + \gamma \kappa_2 \theta_2)B_1 + \kappa_2 \theta_2 B_2 \\
\frac{dB_0}{d\tau} &= 0 \\
\frac{dB_1}{d\tau} &= -\frac{1}{2} B_0 - \kappa_1 B_1 + \frac{1}{2} [B_0^2 + 2 \rho \xi_1 B_0 B_1 + \xi_1^2 B_1^2] \\
\frac{dB_2}{d\tau} &= -\gamma \kappa_2 B_1 - \kappa_2 B_2 + \frac{1}{2} [\gamma^2 \xi_2^2 B_1^2 + 2 \gamma \xi_2 B_1 B_2 + \xi_2^2 B_2^2]
\end{align*}
\]

with the initial condition \( A(0) = B_1(0) = B_2(0) = 0 \) and \( B_0(0) = iu \). Solving for \( B_0 \) and arranging terms gives the final form

\[
B_0(u, \tau) = iu \\
\frac{dA}{d\tau} = riu + (\kappa_1 \theta_1 + \gamma \kappa_2 \theta_2)B_1 + \kappa_2 \theta_2 B_2 \\
\frac{dB_1}{d\tau} = -\frac{1}{2} u(i + u) - (\kappa_1 - \rho \xi_1 iu)B_1 + \frac{1}{2} \xi_1^2 B_1^2 \\
\frac{dB_2}{d\tau} = \frac{1}{2} \gamma^2 \xi_2^2 B_1^2 - \gamma \kappa_2 B_1 - (\kappa_2 - \gamma \xi_2 B_1)B_2 + \frac{1}{2} \xi_2^2 B_2^2
\]

Equation for \( B_1 \) is the same form as the usual Heston type of model which can be solved explicitly. However, the equation for \( B_2 \) cannot be solved analytically due to the presence of the term \( \gamma \kappa_2 B_1 \). Thus both \( B_1 \) and \( A \) need to be solved numerically using 4th order Runge-Kutta method.

References


Notes: The top panel plots the daily effective spread of S&P500 index computed as the equally-weighted average of individual firms’ effective spread on each day. The middle panel plots the realized variance constructed from intra-day overlapping 5-min grid of returns on the S&P500 index. The last panel plots the implied variance computed from CBOE’s VIX index. The samples goes from January 2, 2004 to Dec 31, 2012. The gray-shaded region corresponds to the NBER recession dates.
Figure 2: Cross-correlation plot between ES and RV, IV, and VRP

Graph A: Cross-correlation between ES and RV
Graph B: Cross-correlation between ES and IV
Graph C: Cross-correlation between ES and VRP

Notes: Sample cross-correlation between effective spreads and variance measures are plotted from using lag orders from -10 to 10. The horizontal line denotes the 2-standard error Bartlett confidence interval around 0. The sample goes from January 2, 2004 to Dec 31, 2012.
Figure 3: Cross-correlation plot between ES and RV, IV, and VRP ARMA(1,1) residuals

Notes: Sample cross-correlation between ARMA(1,1) residuals of the effective spreads and variance measures are plotted using lag orders from -10 to 10. The horizontal line denotes the 2-standard error Bartlett confidence interval around 0. The sample goes from January 2, 2004 to Dec 31, 2012.
Figure 4: Impulse Response function plot from VAR(1) fit on ES, RV, and VRP

Notes: Using the VAR(1) estimates from Table 1, a one standard deviation shock is applied to each series and its impulse response onto other series are plotted. The horizontal axis denotes the number of days and the vertical axis denotes the magnitude of the impact.
Figure 5: Term-structure of conditional expected variance from SV-RVL model with different parametrization

Notes: I plot the term structure of expected future variance under the SV-RVL model. The spot variance is fixed at the long-run mean of the variance $\theta_1$ which is set at 0.4. The top panel uses $\kappa_1 = 5$ and $\kappa_2 = 5.01$ to demonstrate the case when the two speeds of mean reversion are close to each other. The second panel uses $\kappa_1 = 1$ and $\kappa_2 = 10$. The last panel uses $\kappa_1 = 10$ and $\kappa_2 = 1$. Horizontal axis denotes the number of days.
Figure 6: Filtered Volatility plot from different models

Notes: I plot the filtered volatility from each model using parameters values from Table 2. The top two panels plot one-factor models verses each of two-factor models. The last panel plots the 2 two-factor models only. The Kalman Filter is applied on all models. The plots go from January 2, 2004 to December 31, 2013. The gray-shaded region corresponds to the NBER recession dates.
Figure 7: QQ-plot of filtered observation equation residuals

Notes: QQ-plots of measurement equation residual quantiles against standard normal quantiles are plotted for each model. The top two panels plot residuals from univariate models using RV and effective spreads in the observation equation. The bottom two panels plot residuals from two-factor models using only RV in the observation equation.
Figure 8: Daily percentage of variance explained by liquidity component in SV-RVL2F Model

Notes: I plot the percentage of variance explained by the liquidity component in the SV-RVL2F model. All parameters are from Table 2 and the decomposition is done as described in Section 3.2. The gray-shaded region corresponds to the NBER recession dates.
Figure 9: Term-structure of conditional expected variance from 2-factor models

Notes: Term structures of conditional expected future variance are plotted for the two-factor models. All parameters are taken from Table 2. The spot variance is set to 0.04 in all cases. The top panel plots 4 different decompositions of the initial spot variance, varying from 100% latent variance factor to 100% liquidity factor. The bottom panel fixes spot variance to 0.04 and varies the spot liquidity from 0.01 to 0.04.
Figure 10: IV Bias by Moneyness for each model

Notes: IV Bias defined as model IV minus market IV is plotted across moneyness using the numbers from Table 7. The horizontal line indicates a bias of zero.
<table>
<thead>
<tr>
<th>Effective Spread</th>
<th>Realized Variance</th>
<th>Implied Variance</th>
<th>Variance Premium</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ES</strong></td>
<td>0.983 (81.11)</td>
<td>-0.040 (-6.72)***</td>
<td></td>
</tr>
<tr>
<td><strong>RV</strong></td>
<td>0.834 (17.30)***</td>
<td>0.313 (13.29)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effective Spread</th>
<th>Implied Variance</th>
<th>variance Premium</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ES</strong></td>
<td>0.717 (45.25)</td>
<td>0.198 (14.74)***</td>
</tr>
<tr>
<td><strong>IV</strong></td>
<td>0.076 (6.11)***</td>
<td>0.913 (86.37)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effective Spread</th>
<th>Realized Variance</th>
<th>Variance Premium</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ES</strong></td>
<td>0.962 (115.58)***</td>
<td>0.079 (13.40)***</td>
</tr>
<tr>
<td><strong>VP</strong></td>
<td>-0.199 (-6.49)***</td>
<td>0.200 (9.23)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effective Spread</th>
<th>Realized Variance</th>
<th>Variance Premium</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ES</strong></td>
<td>0.777 (47.81)***</td>
<td>0.173 (13.04)***</td>
</tr>
<tr>
<td><strong>RV</strong></td>
<td>0.001 (0.02)</td>
<td>1.176 (22.33)***</td>
</tr>
<tr>
<td><strong>VP</strong></td>
<td>0.087 (1.41)</td>
<td>-0.268 (-5.33)***</td>
</tr>
</tbody>
</table>

Table 1: **VAR(1) Estimation of ES, RV and VP time series**

Notes: Using effective spread (ES), realized variance (RV), implied variance (IV), and variance risk premium (VP) as defined in section 2.1, I run bivariate and trivariate VAR(1) regression on ES, RV, and VRP. T-statistics are reported in parenthesis with *** indicating statistical significance of coefficients at the 1% level.
Table 2: Maximum Likelihood Estimation on Daily Returns, RV, and Effective Spread. 2004∼2012

<table>
<thead>
<tr>
<th>Parameters</th>
<th>SV-RV</th>
<th>SV-L</th>
<th>SV-RVL</th>
<th>SV-RVL2F</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa_1$</td>
<td>Estimate (5.2E-03)</td>
<td>Estimate (9.60E-07)</td>
<td>Estimate (3.18E-08)</td>
<td></td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>0.0406 (2.12E-02)</td>
<td>0.0363 (1.80E-06)</td>
<td>0.0317 (6.04E-07)</td>
<td></td>
</tr>
<tr>
<td>$\xi_1$</td>
<td>0.4412 (2.68E-01)</td>
<td>0.4352 (1.93E-06)</td>
<td>0.6109 (4.72E-07)</td>
<td></td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>-0.6341 (1.90E-01)</td>
<td>-0.4884 (1.40E-06)</td>
<td>-0.5087 (7.92E-07)</td>
<td></td>
</tr>
<tr>
<td>$\kappa_2$</td>
<td>2.2024 (3.18E-06)</td>
<td>3.2315 (1.82E-06)</td>
<td>2.8717 (1.93E-06)</td>
<td></td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>0.0260 (2.15E-06)</td>
<td>0.0422 (2.11E-06)</td>
<td>0.0416 (2.08E-06)</td>
<td></td>
</tr>
<tr>
<td>$\xi_2$</td>
<td>0.1057 (2.24E-05)</td>
<td>0.1403 (6.48E-06)</td>
<td>0.1373 (9.09E-07)</td>
<td></td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>-0.6201 (1.73E-05)</td>
<td>-0.7250 (1.69E-06)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma$</td>
<td>1.1495 (2.39E-06)</td>
<td>0.3588 (1.74E-06)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Model Properties

| Average Physical Volatility | 15.44 | 15.59 | 15.67 | 15.74 |
| Log Likelihoods |
| Returns | 17,215 | 17,704 | 17,610 | 17,350 |

Notes: Using daily returns, RV, and effective spread measures I estimate the four models described in section 3 using Maximum Likelihood Estimation. Average physical volatility denotes the average volatility of the daily latent spot volatility process from the Kalman Filter. The last row reports the log-likelihood of returns from each model for comparison. Each model is estimated to maximize fit on returns, RV, and effective spreads. The sample goes from January 2, 2004 to Dec 31, 2012. Standard errors are computed using the BHHH method and reported in parentheses.
<table>
<thead>
<tr>
<th>Statistics</th>
<th>SV-RV</th>
<th>SV-L</th>
<th>SV-RVL</th>
<th>SV-RVL2F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>15.44</td>
<td>15.59</td>
<td>15.67</td>
<td>15.74</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>8.805</td>
<td>2.213</td>
<td>8.639</td>
<td>8.663</td>
</tr>
<tr>
<td>Skewness</td>
<td>2.225</td>
<td>2.128</td>
<td>2.306</td>
<td>2.458</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>9.091</td>
<td>8.251</td>
<td>9.529</td>
<td>10.328</td>
</tr>
<tr>
<td>Minimum</td>
<td>2.14</td>
<td>12.53</td>
<td>3.55</td>
<td>7.00</td>
</tr>
<tr>
<td>Maximum</td>
<td>60.73</td>
<td>27.28</td>
<td>60.68</td>
<td>62.01</td>
</tr>
</tbody>
</table>

Table 3: **Summary Statistics of Filtered Volatility from different models**

Notes: Sample moments as well as the minimum and maximum from the filtered daily spot volatility series are reported. All volatility series are annualized by scaling the variance by 252. Parameters from Table 2 are used for filtration in all models.
Table 4: S&P500 Index Option Data by Moneyness, Maturity, and VIX Level. 2004-2012

<table>
<thead>
<tr>
<th>By Moneyness</th>
<th>Delta&lt;0.3</th>
<th>0.3&lt;Delta&lt;0.4</th>
<th>0.4&lt;Delta&lt;0.5</th>
<th>0.5&lt;Delta&lt;0.6</th>
<th>0.6&lt;Delta&lt;0.7</th>
<th>Delta&gt;0.7</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Contracts</td>
<td>2,087</td>
<td>928</td>
<td>2,890</td>
<td>1,149</td>
<td>996</td>
<td>3,684</td>
<td>11,734</td>
</tr>
<tr>
<td>Average Price</td>
<td>12.58</td>
<td>21.71</td>
<td>39.97</td>
<td>65.78</td>
<td>84.93</td>
<td>161.15</td>
<td>78.04</td>
</tr>
<tr>
<td>Average Implied Volatility</td>
<td>18.66</td>
<td>16.53</td>
<td>17.77</td>
<td>19.73</td>
<td>21.75</td>
<td>28.35</td>
<td>21.68</td>
</tr>
<tr>
<td>Average Bid-Ask Spread</td>
<td>1.574</td>
<td>1.652</td>
<td>1.978</td>
<td>1.756</td>
<td>1.679</td>
<td>1.537</td>
<td>1.695</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>By Maturity</th>
<th>DTM&lt;30</th>
<th>30&lt;DTM&lt;60</th>
<th>60&lt;DTM&lt;90</th>
<th>90&lt;DTM&lt;120</th>
<th>120&lt;DTM&lt;150</th>
<th>DTM&gt;150</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Contracts</td>
<td>1,412</td>
<td>3,408</td>
<td>2,717</td>
<td>1,916</td>
<td>1,123</td>
<td>1,158</td>
<td>11,734</td>
</tr>
<tr>
<td>Average Price</td>
<td>44.35</td>
<td>65.06</td>
<td>79.02</td>
<td>93.30</td>
<td>99.91</td>
<td>108.58</td>
<td>78.04</td>
</tr>
<tr>
<td>Average Bid-Ask Spread</td>
<td>0.919</td>
<td>1.375</td>
<td>1.801</td>
<td>2.111</td>
<td>2.089</td>
<td>2.261</td>
<td>1.695</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>By VIX Level</th>
<th>VIX&lt;15</th>
<th>15&lt;VIX&lt;20</th>
<th>20&lt;VIX&lt;25</th>
<th>25&lt;VIX&lt;30</th>
<th>30&lt;VIX&lt;35</th>
<th>VIX&gt;35</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Contracts</td>
<td>2,682</td>
<td>3,846</td>
<td>2,567</td>
<td>1,005</td>
<td>519</td>
<td>1,115</td>
<td>11,734</td>
</tr>
<tr>
<td>Average Price</td>
<td>55.67</td>
<td>74.09</td>
<td>89.68</td>
<td>97.30</td>
<td>101.32</td>
<td>101.32</td>
<td>78.04</td>
</tr>
<tr>
<td>Average Implied Volatility</td>
<td>13.26</td>
<td>20.18</td>
<td>23.34</td>
<td>26.67</td>
<td>31.12</td>
<td>31.12</td>
<td>21.68</td>
</tr>
<tr>
<td>Average Bid-Ask Spread</td>
<td>1.104</td>
<td>1.552</td>
<td>1.772</td>
<td>1.995</td>
<td>2.131</td>
<td>2.954</td>
<td>1.695</td>
</tr>
</tbody>
</table>

Notes: This table summarizes the characteristics of option data panel used in estimation. Standard filters described in the text are used. A total of 11,734 S&P500 index option contracts are obtained from the OptionMetrics database. The three panels sort the data by the Black-Scholes delta, time to maturity in days, and the VIX level on the day of the option quote, respectively.
Table 5: Sequential Estimation on Options. 2004–2012

<table>
<thead>
<tr>
<th>Parameters</th>
<th>SV-RV</th>
<th>SV-L</th>
<th>SV-RVL</th>
<th>SV-RVL2F</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa_1^*$</td>
<td>Estimate</td>
<td>Std Error</td>
<td>Estimate</td>
<td>Std Error</td>
</tr>
<tr>
<td>$\theta_1^*$</td>
<td>3.3029</td>
<td>4.4405</td>
<td>2.2331</td>
<td>0.0836</td>
</tr>
<tr>
<td>$\xi_1$</td>
<td>0.0492</td>
<td>0.0482</td>
<td>0.0836</td>
<td></td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>0.4412</td>
<td>0.4352</td>
<td>0.6109</td>
<td></td>
</tr>
<tr>
<td>$\kappa_2^*$</td>
<td>-0.6341</td>
<td>-0.4884</td>
<td>-0.5087</td>
<td></td>
</tr>
<tr>
<td>$\theta_2^*$</td>
<td>2.0650</td>
<td>3.2315</td>
<td>34.1172</td>
<td></td>
</tr>
<tr>
<td>$\xi_2$</td>
<td>0.0278</td>
<td>0.0422</td>
<td>0.0035</td>
<td></td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>0.1057</td>
<td>0.1403</td>
<td>0.1373</td>
<td></td>
</tr>
<tr>
<td>$\gamma$</td>
<td>-0.6201</td>
<td>-0.7250</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_\nu$</td>
<td>-1.5945</td>
<td>(5.80E-02)</td>
<td>-2.3741</td>
<td>(1.09E-01)</td>
</tr>
<tr>
<td>$\lambda_\ell$</td>
<td>-1.2995</td>
<td>(9.83E-01)</td>
<td>227.6111</td>
<td>(1.23E+01)</td>
</tr>
</tbody>
</table>

Model Properties

<table>
<thead>
<tr>
<th>Options</th>
<th>Average Model IV</th>
<th>20.50</th>
<th>15.92</th>
<th>21.38</th>
<th>20.94</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Likelihoods</td>
<td>19,287</td>
<td>13,967</td>
<td>19,155</td>
<td>20,596</td>
<td></td>
</tr>
<tr>
<td>Option Errors</td>
<td>IVRMSE</td>
<td>5.32</td>
<td>10.09</td>
<td>5.04</td>
<td>4.72</td>
</tr>
<tr>
<td></td>
<td>Ratio to SV-RV</td>
<td>1.000</td>
<td>1.896</td>
<td>0.947</td>
<td>0.887</td>
</tr>
<tr>
<td></td>
<td>VWRMSE</td>
<td>4.68</td>
<td>8.19</td>
<td>4.73</td>
<td>4.18</td>
</tr>
<tr>
<td></td>
<td>Ratio to SV-RV</td>
<td>1.000</td>
<td>1.751</td>
<td>1.011</td>
<td>0.894</td>
</tr>
</tbody>
</table>

Notes: Using the option data set summarized in Table 3, I fit the risk-neutral version of each model by minimizing Vega Weighted Root Mean Squared Error (VWRMSE). Average Model IV denotes the average Black-Scholes implied volatility across option contracts where the option prices are computed using calibrated model parameters. I also report the IVRMSE of each model in the last panel. Only the risk premium parameter $\lambda$ is estimated, whereas all other parameters values are set to the physical estimates from Table 2. Standard errors are computed using the BHHH method and are reported in parenthesis.
Table 6: IVRMSE Option Error by Moneyness, Maturity, and VIX Level 2004-2012

Panel A: IVRMSE by Moneyness

<table>
<thead>
<tr>
<th>Model</th>
<th>Delta&lt;0.3</th>
<th>0.3&lt;Delta&lt;0.4</th>
<th>0.4&lt;Delta&lt;0.5</th>
<th>0.5&lt;Delta&lt;0.6</th>
<th>0.6&lt;Delta&lt;0.7</th>
<th>Delta&gt;0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>SV-RVL</td>
<td>7.105</td>
<td>4.172</td>
<td>4.044</td>
<td>3.863</td>
<td>3.984</td>
<td>5.110</td>
</tr>
<tr>
<td>SV-RVL2F</td>
<td>7.289</td>
<td>3.963</td>
<td>3.506</td>
<td>3.351</td>
<td>3.528</td>
<td>4.505</td>
</tr>
</tbody>
</table>

Panel B: IVRMSE by Maturity

<table>
<thead>
<tr>
<th>Model</th>
<th>DTM&lt;30</th>
<th>30&lt;DTM&lt;60</th>
<th>60&lt;DTM&lt;90</th>
<th>90&lt;DTM&lt;120</th>
<th>120&lt;DTM&lt;150</th>
<th>DTM&gt;150</th>
</tr>
</thead>
<tbody>
<tr>
<td>SV-RV</td>
<td>6.359</td>
<td>5.653</td>
<td>5.014</td>
<td>4.797</td>
<td>4.619</td>
<td>5.064</td>
</tr>
<tr>
<td>SV-RVL</td>
<td>4.717</td>
<td>4.768</td>
<td>4.923</td>
<td>5.327</td>
<td>5.417</td>
<td>5.566</td>
</tr>
</tbody>
</table>

Panel C: IVRMSE by VIX Level

<table>
<thead>
<tr>
<th>Model</th>
<th>VIX&lt;15</th>
<th>15&lt;VIX&lt;20</th>
<th>20&lt;VIX&lt;25</th>
<th>25&lt;VIX&lt;30</th>
<th>30&lt;VIX&lt;35</th>
<th>VIX&gt;35</th>
</tr>
</thead>
<tbody>
<tr>
<td>SV-RV</td>
<td>3.772</td>
<td>3.989</td>
<td>5.750</td>
<td>5.836</td>
<td>7.802</td>
<td>8.576</td>
</tr>
<tr>
<td>SV-L</td>
<td>3.087</td>
<td>5.367</td>
<td>9.046</td>
<td>11.578</td>
<td>15.606</td>
<td>22.936</td>
</tr>
</tbody>
</table>

Notes: Using parameter estimates from Table 5, I report the decomposition of IVRMSE. The three panels sort the data by the Black-Scholes delta, time to maturity in days, and the VIX level on the day of the option quote, respectively.
Table 7: **IV Bias Option Error by Moneyness, Maturity, and VIX Level**

**Panel A: IV Bias by Moneyness**

<table>
<thead>
<tr>
<th>Model</th>
<th>Delta&lt;0.3</th>
<th>0.3&lt;Delta&lt;0.4</th>
<th>0.4&lt;Delta&lt;0.5</th>
<th>0.5&lt;Delta&lt;0.6</th>
<th>0.6&lt;Delta&lt;0.7</th>
<th>Delta&gt;0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>SV-RV</td>
<td>-1.839</td>
<td>-0.877</td>
<td>-0.137</td>
<td>0.520</td>
<td>1.321</td>
<td>4.613</td>
</tr>
<tr>
<td>SV-RVL</td>
<td>1.004</td>
<td>-1.170</td>
<td>-0.842</td>
<td>-0.197</td>
<td>0.587</td>
<td>1.244</td>
</tr>
<tr>
<td>SV-RVL2F</td>
<td>2.043</td>
<td>0.047</td>
<td>0.156</td>
<td>0.425</td>
<td>0.834</td>
<td>0.715</td>
</tr>
</tbody>
</table>

**Panel B: IV Bias by Maturity**

<table>
<thead>
<tr>
<th>Model</th>
<th>DTM&lt;30</th>
<th>30&lt;DTM&lt;60</th>
<th>60&lt;DTM&lt;90</th>
<th>90&lt;DTM&lt;120</th>
<th>120&lt;DTM&lt;150</th>
<th>DTM&gt;150</th>
</tr>
</thead>
<tbody>
<tr>
<td>SV-RV</td>
<td>3.025</td>
<td>2.047</td>
<td>0.963</td>
<td>1.159</td>
<td>-0.727</td>
<td>-1.215</td>
</tr>
<tr>
<td>SV-L</td>
<td>4.894</td>
<td>5.242</td>
<td>5.502</td>
<td>7.322</td>
<td>6.298</td>
<td>5.855</td>
</tr>
<tr>
<td>SV-RVL</td>
<td>1.575</td>
<td>0.537</td>
<td>-0.147</td>
<td>0.829</td>
<td>-0.555</td>
<td>-0.953</td>
</tr>
<tr>
<td>SV-RVL2F</td>
<td>2.240</td>
<td>1.242</td>
<td>0.469</td>
<td>0.993</td>
<td>-0.598</td>
<td>-1.026</td>
</tr>
</tbody>
</table>

**Panel C: IV Bias by VIX Level**

<table>
<thead>
<tr>
<th>Model</th>
<th>VIX&lt;15</th>
<th>15&lt;VIX&lt;20</th>
<th>20&lt;VIX&lt;25</th>
<th>25&lt;VIX&lt;30</th>
<th>30&lt;VIX&lt;35</th>
<th>VIX&gt;35</th>
</tr>
</thead>
<tbody>
<tr>
<td>SV-RV</td>
<td>-2.548</td>
<td>0.996</td>
<td>2.972</td>
<td>2.581</td>
<td>4.949</td>
<td>3.656</td>
</tr>
<tr>
<td>SV-RVL</td>
<td>-3.952</td>
<td>-1.106</td>
<td>1.734</td>
<td>2.877</td>
<td>5.405</td>
<td>7.373</td>
</tr>
<tr>
<td>SV-RVL2F</td>
<td>-2.619</td>
<td>0.081</td>
<td>2.172</td>
<td>2.623</td>
<td>4.669</td>
<td>4.296</td>
</tr>
</tbody>
</table>

Notes: Using parameter estimates from Table 5, I report the decomposition of IV Bias. The three panels sort the data by the Black-Scholes delta, time to maturity in days, and the VIX level on the day of the option quote, respectively.